

EDITORS

Jean-Pierre Franoise
Universit  P.-M. Curie, Paris VI
Paris, France

Gregory L. Naber
Drexel University
Philadelphia, PA, USA

Tsou Sheung Tsun
University of Oxford
Oxford, UK

EDITORIAL ADVISORY BOARD

Sergio Albeverio

Rheinische Friedrich-Wilhelms-Universität Bonn
Bonn, Germany

Huzihiro Araki

Kyoto University
Kyoto, Japan

Abhay Ashtekar

Pennsylvania State University
University Park, PA, USA

Andrea Braides

Università di Roma "Tor Vergata"
Roma, Italy

Francesco Calogero

Università di Roma "La Sapienza"
Roma, Italy

Cecile DeWitt-Morette

The University of Texas at Austin
Austin, TX, USA

Artur Ekert

University of Cambridge
Cambridge, UK

Giovanni Gallavotti

Università di Roma "La Sapienza"
Roma, Italy

Simon Gindikin

Rutgers University
Piscataway, NJ, USA

Gennadi Henkin

Université P.-M. Curie, Paris VI
Paris, France

Allen C. Hirshfeld

Universität Dortmund
Dortmund, Germany

Lisa Jeffrey

University of Toronto
Toronto, Canada

T.W.B. Kibble

Imperial College of Science, Technology and Medicine
London, UK

Antti Kupiainen

University of Helsinki
Helsinki, Finland

Shahn Majid

Queen Mary, University of London
London, UK

Barry M. McCoy

State University of New York Stony Brook
Stony Brook, NY, USA

Hiroshi Ooguri

California Institute of Technology
Pasadena, CA, USA

Roger Penrose

University of Oxford
Oxford, UK

Pierre Ramond

University of Florida
Gainesville, FL, USA

Tudor Ratiu

Ecole Polytechnique Federale de Lausanne
Lausanne, Switzerland

Rudolf Schmid

Emory University
Atlanta, GA, USA

Albert Schwarz

University of California
Davis, CA, USA

Yakov Sinai
Princeton University
Princeton, NJ, USA

Herbert Spohn
Technische Universität München
München, Germany

Stephen J. Summers
University of Florida
Gainesville, FL, USA

Roger Temam
Indiana University
Bloomington, IN, USA

Craig A. Tracy
University of California
Davis, CA, USA

Andrzej Trautman
Warsaw University
Warsaw, Poland

Vladimir Turaev
Institut de Recherche Mathématique Avancée,
Strasbourg, France

Gabriele Veneziano
CERN, Genève, Switzerland

Reinhard F. Werner
Technische Universität Braunschweig
Braunschweig, Germany

C.N. Yang
Tsinghua University
Beijing, China

Eberhard Zeidler
Max-Planck Institut für Mathematik in
den Naturwissenschaften
Leipzig, Germany

Steve Zelditch
Johns Hopkins University
Baltimore, MD, USA

FOREWORD

In bygone centuries, our physical world appeared to be filled to the brim with mysteries. Divine powers could provide for genuine miracles; water and sunlight could turn arid land into fertile pastures, but the same powers could lead to miseries and disasters. The force of life, the *vis vitalis*, was assumed to be the special agent responsible for all living things. The heavens, whatever they were for, contained stars and other heavenly bodies that were the exclusive domain of the Gods.

Mathematics did exist, of course. Indeed, there was one aspect of our physical world that was recognised to be controlled by precise, mathematical logic: the geometric structure of space, elaborated to become a genuine form of art by the ancient Greeks. From my perspective, the Greeks were the first practitioners of ‘mathematical physics’, when they discovered that all geometric features of space could be reduced to a small number of axioms. Today, these would be called ‘fundamental laws of physics’. The fact that the flow of *time* could be addressed with similar exactitude, and that it could be handled geometrically together with space, was only recognised much later. And, yes, there were a few crazy people who were interested in the magic of numbers, but the real world around us seemed to contain so much more that was way beyond our capacities of analysis.

Gradually, all this changed. The Moon and the planets appeared to follow geometrical laws. Galilei and Newton managed to identify their logical rules of motion, and by noting that the concept of mass could be applied to things in the sky just like apples and cannon balls on Earth, they made the sky a little bit more accessible to us. Electricity, magnetism, light and sound were also found to behave in complete accordance with mathematical equations.

Yet all of this was just a beginning. The real changes came with the twentieth century. A completely new way of thinking, by emphasizing mathematical, logical analysis rather than empirical evidence, was pioneered by Albert Einstein. Applying advanced mathematical concepts, only known to a few pure mathematicians, to notions as mundane as space and time, was new to the physicists of his time. Einstein himself had a hard time struggling through the logic of connections and curvatures, notions that were totally new to him, but are only too familiar to students of mathematical physics today. Indeed, there is no better testimony of Einstein’s deep insights at that time, than the fact that we now teach these things regularly in our university classrooms.

Special and general relativity are only small corners of the realm of modern physics that is presently being studied using advanced mathematical methods. We have notoriously complex subjects such as phase transitions in condensed matter physics, superconductivity, Bose–Einstein condensation, the quantum Hall effect, particularly the fractional quantum Hall effect, and numerous topics from elementary particle physics, ranging from fibre bundles and renormalization groups to supergravity, algebraic topology, superstring theory, Calabi–Yau spaces and what not, all of which require the utmost of our mental skills to comprehend them.

The most bewildering observation that we make today is that it seems that our *entire* physical world appears to be controlled by mathematical equations, and these are not just sloppy and debatable models, but precisely documented properties of materials, of systems, and of phenomena in all echelons of our universe.

Does this really apply to our entire world, or only to parts of it? Do features, notions, entities exist that are emphatically *not* mathematical? What about intuition, or dreams, and what about consciousness? What about religion? Here, most of us would say, one should not even try to apply mathematical analysis, although even here, some brave social scientists are making attempts at coordinating rational approaches.

No, there are clear and important differences between the physical world and the mathematical world. Where the physical world stands out is the fact that it refers to ‘reality’, whatever ‘reality’ is. Mathematics is the world of pure logic and pure reasoning. In physics, it is the experimental evidence that ultimately decides whether a theory is acceptable or not. Also, the methodology in physics is different.

A beautiful example is the serendipitous discovery of superconductivity. In 1911, the Dutch physicist Heike Kamerlingh Onnes was the first to achieve the liquefaction of helium, for which a temperature below 4.25 K had to be realized. Heike decided to measure the specific conductivity of mercury, a metal that is frozen solid at such low temperatures. But something appeared to go wrong during the measurements, since the volt meter did not show any voltage at all. All experienced physicists in the team assumed that they were dealing with a malfunction. It would not have been the first time for a short circuit to occur in the electrical equipment, but, this time, in spite of several efforts, they failed to locate it. One of the assistants was responsible for keeping the temperature of the sample well within that of liquid helium, a dull job, requiring nothing else than continuously watching some dials. During one of the many tests, however, he dozed off. The temperature rose, and suddenly the measurements showed the normal values again. It then occurred to the investigators that the effect and its temperature dependence were completely reproducible. Below 4.19 degrees Kelvin the conductivity of mercury appeared to be strictly infinite. Above that temperature, it is finite, and the transition is a very sudden one. Superconductivity was discovered (D. van Delft, “Heike Kamerlingh Onnes”, Uitgeverij Bert Bakker, Amsterdam, 2005 (in Dutch)).

This is not the way mathematical discoveries are made. Theorems are not produced by assistants falling asleep, even if examples do exist of incidents involving some miraculous fortune.

The hybrid science of mathematical physics is a very curious one. Some of the topics in this Encyclopedia are undoubtedly physical. High T_c superconductivity, breaking water waves, and magneto-hydrodynamics, are definitely topics of physics where experimental data are considered more decisive than any high-brow theory. Cohomology theory, Donaldson–Witten theory, and AdS/CFT correspondence, however, are examples of purely mathematical exercises, even if these subjects, like all of the others in this compilation, are strongly inspired by, and related to, questions posed in physics.

It is inevitable, in a compilation of a large number of short articles with many different authors, to see quite a bit of variation in style and level. In this Encyclopedia, theoretical physicists as well as mathematicians together made a huge effort to present in a concise and understandable manner their vision on numerous important issues in advanced mathematical physics. All include references for further reading. We hope and expect that these efforts will serve a good purpose.

Gerard 't Hooft,
Spinoza Institute,
Utrecht University,
The Netherlands.

PREFACE

Mathematical Physics as a distinct discipline is relatively new. The International Association of Mathematical Physics was founded only in 1976. The interaction between physics and mathematics has, of course, existed since ancient times, but the recent decades, perhaps partly because we are living through them, appear to have witnessed tremendous progress, yielding new results and insights at a dizzying pace, so much so that an encyclopedia seems now needed to collate the gathered knowledge.

Mathematical Physics brings together the two great disciplines of Mathematics and Physics to the benefit of both, the relationship between them being symbiotic. On the one hand, it uses mathematics as a tool to organize physical ideas of increasing precision and complexity, and on the other it draws on the questions that physicists pose as a source of inspiration to mathematicians. A classical example of this relationship exists in Einstein's theory of relativity, where differential geometry played an essential role in the formulation of the physical theory while the problems raised by the ensuing physics have in turn boosted the development of differential geometry. It is indeed a happy coincidence that we are writing now a preface to an encyclopedia of mathematical physics in the centenary of Einstein's *annus mirabilis*.

The project of putting together an encyclopedia of mathematical physics looked, and still looks, to us a formidable enterprise. We would never have had the courage to undertake such a task if we did not believe, first, that it is worthwhile and of benefit to the community, and second, that we would get the much-needed support from our colleagues. And this support we did get, in the form of advice, encouragement, and practical help too, from members of our Editorial Advisory Board, from our authors, and from others as well, who have given unstintingly so much of their time to help us shape this Encyclopedia.

Mathematical Physics being a relatively new subject, it is not yet clearly delineated and could mean different things to different people. In our choice of topics, we were guided in part by the programs of recent International Congresses on Mathematical Physics, but mainly by the advice from our Editorial Advisory Board and from our authors. The limitations of space and time, as well as our own limitations, necessitated the omission of certain topics, but we have tried to include all that we believe to be core subjects and to cover as much as possible the most active areas.

Our subject being interdisciplinary, we think it appropriate that the Encyclopedia should have certain special features. Applications of the same mathematical theory, for instance, to different problems in physics will have different emphasis and treatment. By the same token, the same problem in physics can draw upon resources from different mathematical fields. This is why we divide the Encyclopedia into two broad sections: physics subjects and related mathematical subjects. Articles in either section are deliberately allowed a fair amount of overlap with one another and many articles will appear under more than one heading, but all are linked together by elaborate cross referencing. We think this gives a better picture of the subject as a whole and will serve better a community of researchers from widely scattered yet related fields.

The Encyclopedia is intended primarily for experienced researchers but should be of use also to beginning graduate students. For the latter category of readers, we have included eight elementary introductory articles for easy reference, with those on mathematics aimed at physics graduates and those on physics aimed at mathematics graduates, so that these articles can serve as their first port of call to enable them to embark on any of the main articles without the need to consult other material beforehand. In fact, we think these articles may even form the

foundation of advanced undergraduate courses, as we know that some authors have already made such use of them.

In addition to the printed version, an on-line version of the Encyclopedia is planned, which will allow both the contents and the articles themselves to be updated if and when the occasion arises. This is probably a necessary provision in such a rapidly advancing field.

This project was some four years in the making. Our foremost thanks at its completion go to the members of our Editorial Advisory Board, who have advised, helped and encouraged us all along, and to all our authors who have so generously devoted so much of their time to writing these articles and given us much useful advice as well. We ourselves have learnt a lot from these colleagues, and made some wonderful contacts with some among them. Special thanks are due also to Arthur Greenspoon whose technical expertise was indispensable.

The project was started with Academic Press, which was later taken over by Elsevier. We thank warmly members of their staff who have made this transition admirably seamless and gone on to assist us greatly in our task: both Carey Chapman and Anne Guillaume, who were in charge of the whole project and have been with us since the beginning, and Edward Taylor responsible for the copy-editing. And Martin Ruck, who manages to keep an overwhelming amount of details constantly at his fingertips, and who is never known to have lost a single email, deserves a very special mention.

As a postscript, we would like to express our gratitude to the very large number of authors who generously agreed to donate their honorariums to support the Committee for Developing Countries of the European Mathematical Society in their work to help our less fortunate colleagues in the developing world.

Jean-Pierre Franoise
Gregory L. Naber
Tsou Sheung Tsun

PERMISSION ACKNOWLEDGMENTS

The following material is reproduced with kind permission of Nature Publishing Group
Figures 11 and 12 of “Point-vortex Dynamics”

<http://www.nature.com/nature>

The following material is reproduced with kind permission of Oxford University Press
Figure 1 of “Random Walks in Random Environments”

<http://www.oup.co.uk>

Introductory Articles

Introductory Article: Classical Mechanics

G Gallavotti, Università di Roma “La Sapienza,”
Rome, Italy

© 2006 G Gallavotti. Published by Elsevier Ltd.
All rights reserved.

General Principles

Classical mechanics is a theory of motions of point particles. If $\mathbf{X} = (x_1, \dots, x_n)$ are the particle positions in a Cartesian inertial system of coordinates, the equations of motion are determined by their masses (m_1, \dots, m_n) , $m_j > 0$, and by the potential energy of interaction, $V(x_1, \dots, x_n)$, as

$$m_i \ddot{x}_i = -\partial_{x_i} V(x_1, \dots, x_n), \quad i = 1, \dots, n \quad [1]$$

here $x_i = (x_{i1}, \dots, x_{id})$ are coordinates of the i th particle and ∂_{x_i} is the gradient $(\partial_{x_{i1}}, \dots, \partial_{x_{id}})$; d is the space dimension (i.e., $d = 3$, usually). The potential energy function will be supposed “smooth,” that is, analytic except, possibly, when two positions coincide. The latter exception is necessary to include the important cases of gravitational attraction or, when dealing with electrically charged particles, of Coulomb interaction. A basic result is that if V is bounded below, eqn [1] admits, given initial data $\mathbf{X}_0 = \mathbf{X}(0)$, $\dot{\mathbf{X}}_0 = \dot{\mathbf{X}}(0)$, a unique global solution $t \rightarrow \mathbf{X}(t)$, $t \in (-\infty, \infty)$; otherwise a solution can fail to be global if and only if, in a finite time, it reaches infinity or a singularity point (i.e., a configuration in which two or more particles occupy the same point: an event called a collision).

In eqn [1], $-\partial_{x_i} V(x_1, \dots, x_n)$ is the force acting on the points. More general forces are often admitted. For instance, velocity-dependent friction forces: they are not considered here because of their phenomenological nature as models for microscopic phenomena which should also, in principle, be explained in terms of conservative forces (furthermore, even from a macroscopic viewpoint, they are rather incomplete models, as they should be considered together with the important heat generation phenomena that accompany them). Another interesting example of

forces not corresponding to a potential are certain velocity-dependent forces like the Coriolis force (which, however, appears only in noninertial frames of reference) and the closely related Lorentz force (in electromagnetism): they could be easily accommodated in the Hamiltonian formulation of mechanics; see Appendix 2.

The action principle states that an equivalent formulation of the eqns [1] is that a motion $t \rightarrow \mathbf{X}_0(t)$ satisfying [1] during a time interval $[t_1, t_2]$ and leading from $\mathbf{X}^1 = \mathbf{X}_0(t_1)$ to $\mathbf{X}^2 = \mathbf{X}_0(t_2)$, renders stationary the action

$$\mathcal{A}(\{\mathbf{X}\}) = \int_{t_1}^{t_2} \left(\sum_{i=1}^n \frac{1}{2} m_i \dot{\mathbf{X}}_i(t)^2 - V(\mathbf{X}(t)) \right) dt \quad [2]$$

within the class $\mathcal{M}_{t_1, t_2}(\mathbf{X}^1, \mathbf{X}^2)$ of smooth (i.e., analytic) “motions” $t \rightarrow \mathbf{X}(t)$ defined for $t \in [t_1, t_2]$ and leading from \mathbf{X}^1 to \mathbf{X}^2 .

The function

$$\mathcal{L}(\mathbf{Y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^n m_i y_i^2 - V(\mathbf{X}) \stackrel{\text{def}}{=} K(\mathbf{Y}) - V(\mathbf{X}),$$
$$\mathbf{Y} = (y_1, \dots, y_n)$$

is called the Lagrangian function and the action can be written as

$$\int_{t_1}^{t_2} \mathcal{L}(\dot{\mathbf{X}}(t), \mathbf{X}(t)) dt$$

The quantity $K(\dot{\mathbf{X}}(t))$ is called kinetic energy and motions satisfying [1] conserve energy as time t varies, that is,

$$K(\dot{\mathbf{X}}(t)) + V(\mathbf{X}(t)) = E = \text{const.} \quad [3]$$

Hence the action principle can be intuitively thought of as saying that motions proceed by keeping constant the energy, sum of the kinetic and potential energies, while trying to share as evenly as possible their (average over time) contribution to the energy.

In the special case in which V is translation invariant, motions conserve linear momentum $\mathbf{Q} \stackrel{\text{def}}{=} \sum_i m_i \dot{\mathbf{x}}_i$; if V

is rotation invariant around the origin O , motions conserve angular momentum $M \stackrel{\text{def}}{=} \sum_i m_i \mathbf{x}_i \wedge \dot{\mathbf{x}}_i$, where \wedge denotes the vector product in \mathbb{R}^d , that is, it is the tensor $(\mathbf{a} \wedge \mathbf{b})_{ij} = a_i b_j - b_j a_i$, $i, j = 1, \dots, d$: if the dimension $d = 3$ the $\mathbf{a} \wedge \mathbf{b}$ will be naturally regarded as a vector. More generally, to any continuous symmetry group of the Lagrangian correspond conserved quantities: this is formalized in the *Noether theorem*.

It is convenient to think that the scalar product in \mathbb{R}^{dn} is defined in terms of the ordinary scalar product in \mathbb{R}^d , $\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^d a_j b_j$, by $(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^n m_i \mathbf{v}_i \cdot \mathbf{w}_i$: so that kinetic energy and line element ds can be written as $K(\dot{\mathbf{X}}) = \frac{1}{2}(\dot{\mathbf{X}}, \dot{\mathbf{X}})$ and $ds^2 = \sum_{i=1}^n m_i dx_i^2$, respectively. Therefore, the metric generated by the latter scalar product can be called *kinetic energy metric*.

The interest of the kinetic metric appears from the *Maupertuis' principle* (equivalent to [1]): the principle allows us to identify the trajectory traced in \mathbb{R}^d by a motion that leads from X^1 to X^2 moving with energy E . Parametrizing such trajectories as $\tau \rightarrow X(\tau)$ by a parameter τ varying in $[0, 1]$ so that the line element is $ds^2 = (\partial_\tau X, \partial_\tau X) d\tau^2$, the principle states that the trajectory of a motion with energy E which leads from X^1 to X^2 makes stationary, among the analytic curves $\xi \in \mathcal{M}_{0,1}(X^1, X^2)$, the function

$$L(\xi) = \int_\xi \sqrt{E - V(\xi(s))} ds \quad [4]$$

so that the possible trajectories traced by the solutions of [1] in \mathbb{R}^{nd} and with energy E can be identified with the geodesics of the metric $dm^2 \stackrel{\text{def}}{=} (E - V(X)) \cdot ds^2$.

For more details, the reader is referred to [Landau and Lifshitz \(1976\)](#) and [Gallavotti \(1983\)](#).

Constraints

Often particles are subject to constraints which force the motion to take place on a surface $M \subset \mathbb{R}^{nd}$, i.e., $X(t)$ is forced to be a point on the manifold M . A typical example is provided by rigid systems in which motions are subject to forces which keep the mutual distances of the particles constant: $|\mathbf{x}_i - \mathbf{x}_j| = \rho_{ij}$, with ρ_{ij} time-independent positive quantities. In essentially all cases, the forces that imply constraints, called constraint reactions, are velocity dependent and, therefore, are not in the class of conservative forces considered here, cf. [1]. Hence, from a fundamental viewpoint admitting only conservative forces, constrained systems should be regarded as idealizations of systems subject to conservative forces which approximately imply the constraints.

In general, the ℓ -dimensional manifold M will not admit a global system of coordinates: however, it will be possible to describe points in the vicinity of any $X^0 \in M$ by using $N = nd$ coordinates $\mathbf{q} = (q_1, \dots, q_\ell, q_{\ell+1}, \dots, q_N)$ varying in an open ball $B_{X^0}: X = X(q_1, \dots, q_\ell, q_{\ell+1}, \dots, q_N)$.

The q -coordinates can be chosen well adapted to the surface M and to the kinetic metric, i.e., so that the points of M are identified by $q_{\ell+1} = \dots = q_N = 0$ (which is the meaning of “adapted”); furthermore, infinitesimal displacements $(0, \dots, 0, d\varepsilon_{\ell+1}, \dots, d\varepsilon_N)$ out of a point $X^0 \in M$ are orthogonal to M (in the kinetic metric) and have a length independent of the position of X^0 on M (which is the meaning of “well adapted” to the kinetic metric).

Motions constrained on M arise when the potential V has the form

$$V(X) = V_a(X) + \lambda W(X) \quad [5]$$

where W is a smooth function which reaches its minimum value, say equal to 0, precisely on the manifold M while V_a is another smooth potential. The factor $\lambda > 0$ is a parameter called the rigidity of the constraint.

A particularly interesting case arises when the level surfaces of W also have the geometric property of being “parallel” to the surface M : in the precise sense that the matrix $\partial_{q_i q_j}^2 W(X)$, $i, j > \ell$ is positive definite and X -independent, for all $X \in M$, in a system of coordinates well adapted to the kinetic metric.

A potential W with the latter properties can be called an approximately ideal constraint reaction. In fact, it can be proved that, given an initial datum $X^0 \in M$ with velocity \dot{X}^0 tangent to M , i.e., given an initial datum whose coordinates in a local system of coordinates are $(q_0, 0)$ and $(\dot{q}_0, 0)$ with $q_0 = (q_{01}, \dots, q_{0\ell})$ and $\dot{q}_0 = (\dot{q}_{01}, \dots, \dot{q}_{0\ell})$, the motion generated by [1] with V given by [5] is a motion $t \rightarrow X_\lambda(t)$ which

1. as $\lambda \rightarrow \infty$ tends to a motion $t \rightarrow X_\infty(t)$;
2. as long as $X_\infty(t)$ stays in the vicinity of the initial data, say for $0 \leq t \leq t_1$, so that it can be described in the above local adapted coordinates, its coordinates have the form $t \rightarrow (q(t), 0) = (q_1(t), \dots, q_\ell(t), 0, \dots, 0)$: that is, it is a motion developing on the constraint surface M ; and
3. the curve $t \rightarrow X_\infty(t)$, $t \in [0, t_1]$, as an element of the space $\mathcal{M}_{0,t_1}(X^0, X_\infty(t_1))$ of analytic curves on M connecting X^0 to $X_\infty(t_1)$, renders the action

$$A(X) = \int_0^{t_1} (K(\dot{X}(t)) - V_a(X(t))) dt \quad [6]$$

stationary.

The latter property can be formulated “intrinsically,” that is, referring only to M as a surface, via the restriction of the metric ds^2 to line elements $ds = (dq_1, \dots, dq_\ell, 0, \dots, 0)$ tangent to M at the point $\mathbf{X} = (q_0, 0, \dots, 0) \in M$; we write $ds^2 = \sum_{i,j}^{1,\ell} g_{ij}(\mathbf{q}) \times dq_i dq_j$. The $\ell \times \ell$ symmetric positive-definite matrix g can be called the *metric* on M induced by the kinetic energy. Then the action in [6] can be written as

$$\mathcal{A}(\mathbf{q}) = \int_0^{t_1} \left(\frac{1}{2} \sum_{i,j}^{1,\ell} g_{ij}(\mathbf{q}(t)) \dot{q}_i(t) \dot{q}_j(t) - \bar{V}_a(\mathbf{q}(t)) \right) dt \quad [7]$$

where $\bar{V}_a(\mathbf{q}) \stackrel{\text{def}}{=} V_a(\mathbf{X}(q_1, \dots, q_\ell, 0, \dots, 0))$: the function

$$\begin{aligned} \mathcal{L}(\boldsymbol{\eta}, \mathbf{q}) &\stackrel{\text{def}}{=} \frac{1}{2} \sum_{i,j}^{1,\ell} g_{ij}(\mathbf{q}) \eta_i \eta_j - \bar{V}_a(\mathbf{q}) \\ &\equiv \frac{1}{2} g(\mathbf{q}) \boldsymbol{\eta} \cdot \boldsymbol{\eta} - \bar{V}_a(\mathbf{q}) \end{aligned} \quad [8]$$

is called the constrained Lagrangian of the system.

An important property is that the constrained motions conserve the energy defined as $E = \frac{1}{2} (g(\mathbf{q}) \dot{\mathbf{q}}, \dot{\mathbf{q}}) + \bar{V}_a(\mathbf{q})$; see next section.

The constrained motion $\mathbf{X}_\infty(t)$ of energy E satisfies the Maupertuis’ principle in the sense that the curve on M on which the motion develops renders

$$L(\boldsymbol{\xi}) = \int_{\boldsymbol{\xi}} \sqrt{E - V_a(\boldsymbol{\xi}(s))} ds \quad [9]$$

stationary among the (smooth) curves that develop on M connecting two fixed values \mathbf{X}_1 and \mathbf{X}_2 . In the particular case in which $\ell = n$ this is again Maupertuis’ principle for unconstrained motions under the potential $V(\mathbf{X})$. In general, ℓ is called the number of degrees of freedom because a complete description of the initial data requires 2ℓ coordinates $q(0), \dot{q}(0)$.

If W is minimal on M but the condition on W of having level surfaces parallel to M is not satisfied, i.e., if W is not an approximate ideal constraint reaction, it still remains true that the limit motion $\mathbf{X}_\infty(t)$ takes place on M . However, in general, it will not satisfy the above variational principles. For this reason, motions arising as limits (as $\lambda \rightarrow \infty$) of motions developing under the potential [5] with W having minimum on M and level curves parallel (in the above sense) to M are called ideally constrained motions or motions subject by ideal constraints to the surface M .

As an example, suppose that W has the form $W(\mathbf{X}) = \sum_{i,j \in \mathcal{P}} w_{ij}(|\mathbf{x}_i - \mathbf{x}_j|)$ with $w_{ij}(|\boldsymbol{\xi}|) \geq 0$ an analytic function vanishing only when $|\boldsymbol{\xi}| = \rho_{ij}$ for i, j in some set of pairs \mathcal{P} and for some given distances ρ_{ij} (e.g., $w_{ij}(\boldsymbol{\xi}) = (\boldsymbol{\xi}^2 - \rho_{ij}^2)^2 \gamma$, $\gamma > 0$). Then W can be shown to

satisfy the mentioned conditions and therefore, the so constrained motions $\mathbf{X}_\infty(t)$ of the body satisfy the variational principles mentioned in connection with [7] and [9]: in other words, the above natural way of realizing a rather general rigidity constraint is ideal.

The modern viewpoint on the physical meaning of the constraint reactions is as follows: looking at motions in an inertial Cartesian system, it will appear that the system is subject to the applied forces with potential $V_a(\mathbf{X})$ and to constraint forces which are defined as the differences $\mathbf{R}_i = m_i \ddot{\mathbf{x}}_i + \partial_{\mathbf{x}_i} V_a(\mathbf{X})$. The latter reflect the action of the forces with potential $\lambda W(\mathbf{X})$ in the limit of infinite rigidity ($\lambda \rightarrow \infty$).

In applications, sometimes the action of a constraint can be regarded as ideal: the motion will then verify the variational principles mentioned and \mathbf{R} can be computed as the differences between the $m_i \ddot{\mathbf{x}}_i$ and the active forces $-\partial_{\mathbf{x}_i} V_a(\mathbf{X})$. In dynamics problems it is, however, a very difficult and important matter, particularly in engineering, to judge whether a system of particles can be considered as subject to ideal constraints: this leads to important decisions in the construction of machines. It simplifies the calculations of the reactions and fatigue of the materials but a misjudgment can have serious consequences about stability and safety. For statics problems, the difficulty is of lower order: usually assuming that the constraint reaction is ideal leads to an overestimate of the requirements for stability of equilibria. Hence, employing the action principle to statics problems, where it constitutes the principle of *virtual work*, generally leads to economic problems rather than to safety issues. Its discovery even predates Newtonian mechanics.

We refer the reader to Arnol’d (1989) and Gallavotti (1983) for more details.

Lagrange and Hamilton Forms of the Equations of Motion

The stationarity condition for the action $\mathcal{A}(\mathbf{q})$, cf. [7], [8], is formulated in terms of the Lagrangian $\mathcal{L}(\boldsymbol{\eta}, \boldsymbol{\xi})$, see [8], by

$$\begin{aligned} \frac{d}{dt} \partial_{\boldsymbol{\eta}} \mathcal{L}(\dot{\mathbf{q}}(t), \mathbf{q}(t)) \\ = \partial_{\boldsymbol{\xi}} \mathcal{L}(\dot{\mathbf{q}}(t), \mathbf{q}(t)), \quad i = 1, \dots, \ell \end{aligned} \quad [10]$$

which is a second-order differential equation called the Lagrangian equation of motion. It can be cast in “normal form”: for this purpose, adopting the convention of “summation over repeated indices,” introduce the “generalized momenta”

$$p_i \stackrel{\text{def}}{=} g(\mathbf{q})_{ij} \dot{q}_j, \quad i = 1, \dots, \ell \quad [11]$$

Since $g(q) > 0$, the motions $t \rightarrow q(t)$ and the corresponding velocities $t \rightarrow \dot{q}(t)$ can be described equivalently by $t \rightarrow (q(t), \dot{p}(t))$: and the equations of motion [10] become the first-order equations

$$\dot{q}_i = \partial_{p_i} \mathcal{H}(\mathbf{p}, \mathbf{q}), \quad \dot{p}_i = -\partial_{q_i} \mathcal{H}(\mathbf{p}, \mathbf{q}) \quad [12]$$

where the function \mathcal{H} , called the Hamiltonian of the system, is defined by

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) \stackrel{\text{def}}{=} \frac{1}{2}(g(\mathbf{q})^{-1} \mathbf{p}, \mathbf{p}) + \bar{V}_a(\mathbf{q}) \quad [13]$$

Equations [12], regarded as equations of motion for phase space points (\mathbf{p}, \mathbf{q}) , are called Hamilton equations. In general, \mathbf{q} are local coordinates on M and motions are specified by giving $\mathbf{q}, \dot{\mathbf{q}}$ or \mathbf{p}, \mathbf{q} .

Looking for a coordinate-free representation of motions consider the pairs \mathbf{X}, \mathbf{Y} with $\mathbf{X} \in M$ and \mathbf{Y} a vector $\mathbf{Y} \in T_{\mathbf{X}}$ tangent to M at the point \mathbf{X} . The collection of pairs (\mathbf{Y}, \mathbf{X}) is denoted $T(M) = \cup_{\mathbf{X} \in M} (T_{\mathbf{X}} \times \{\mathbf{X}\})$ and a motion $t \rightarrow (\dot{\mathbf{X}}(t), \mathbf{X}(t)) \in T(M)$ in local coordinates is represented by $(\dot{q}(t), q(t))$. The space $T(M)$ can be called the space of initial data for Lagrange's equations of motion: it has 2ℓ dimensions (also known as the "tangent bundle" of M).

Likewise, the space of initial data for the Hamilton equations will be denoted $T^*(M)$ and it consists of pairs \mathbf{X}, \mathbf{P} with $\mathbf{X} \in M$ and $\mathbf{P} = g(\mathbf{X})\mathbf{Y}$ with \mathbf{Y} a vector tangent to M at \mathbf{X} . The space $T^*(M)$ is called the phase space of the system: it has 2ℓ dimensions (and it is occasionally called the "cotangent bundle" of M).

Immediate consequence of [12] is

$$\frac{d}{dt} \mathcal{H}(\mathbf{p}(t), \mathbf{q}(t)) \equiv 0$$

and it means that $\mathcal{H}(\mathbf{p}(t), \mathbf{q}(t))$ is constant along the solutions of [12]. Noting that $\mathcal{H}(\mathbf{p}, \mathbf{q}) = (1/2)(g(\mathbf{q}) \dot{\mathbf{q}}, \dot{\mathbf{q}}) + \bar{V}_a(\mathbf{q})$ is the sum of the kinetic and potential energies, it follows that the conservation of \mathcal{H} along solutions means energy conservation in presence of ideal constraints.

Let S_t be the flow generated on the phase space variables (\mathbf{p}, \mathbf{q}) by the solutions of the equations of motion [12], that is, let $t \rightarrow S_t(\mathbf{p}, \mathbf{q}) \equiv (\mathbf{p}(t), \mathbf{q}(t))$ denote a solution of [12] with initial data (\mathbf{p}, \mathbf{q}) . Then a (measurable) set Δ in phase space evolves in time t into a new set $S_t \Delta$ with the same volume: this is obvious because the Hamilton equations [12] have manifestly zero divergence ("Liouville's theorem").

The Hamilton equations also satisfy a variational principle, called the Hamilton action principle: that is, if $\mathcal{M}_{t_1, t_2}((\mathbf{p}_1, \mathbf{q}_1), (\mathbf{p}_2, \mathbf{q}_2); M)$ denotes the space of the analytic functions $\varphi: t \rightarrow (\boldsymbol{\pi}(t), \boldsymbol{\kappa}(t))$ which in the time interval $[t_1, t_2]$ lead from $(\mathbf{p}_1, \mathbf{q}_1)$ to $(\mathbf{p}_2, \mathbf{q}_2)$, then the condition that $\varphi_0(t) = (\mathbf{p}(t), \mathbf{q}(t))$ satisfies

[12] can be equivalently formulated by requiring that the function

$$\mathcal{A}_{\mathcal{H}}(\varphi) \stackrel{\text{def}}{=} \int_{t_1}^{t_2} (\boldsymbol{\pi}(t) \cdot \dot{\boldsymbol{\kappa}}(t) - \mathcal{H}(\boldsymbol{\pi}(t), \boldsymbol{\kappa}(t))) dt \quad [14]$$

be stationary for $\varphi = \varphi_0$: in fact, eqns [12] are the stationarity conditions for the Hamilton action [14] on $\mathcal{M}_{t_0, t_1}((\mathbf{p}_1, \mathbf{q}_1), (\mathbf{p}_2, \mathbf{q}_2); M)$. And, since the derivatives of $\boldsymbol{\pi}(t)$ do not appear in [14], stationarity is even achieved in the larger space $\mathcal{M}_{t_1, t_2}(\mathbf{q}_1, \mathbf{q}_2; M)$ of the motions $\varphi: t \rightarrow (\boldsymbol{\pi}(t), \boldsymbol{\kappa}(t))$ leading from \mathbf{q}_1 to \mathbf{q}_2 without any restriction on the initial and final momenta $\mathbf{p}_1, \mathbf{p}_2$ (which, therefore, cannot be prescribed *a priori* independently of $\mathbf{q}_1, \mathbf{q}_2$). If the prescribed data $\mathbf{p}_1, \mathbf{q}_1, \mathbf{p}_2, \mathbf{q}_2$ are not compatible with the equations of motion (e.g., $H(\mathbf{p}_1, \mathbf{q}_2) \neq H(\mathbf{p}_2, \mathbf{q}_2)$), then the action functional has no stationary trajectory in $\mathcal{M}_{t_1, t_2}((\mathbf{p}_1, \mathbf{q}_1), (\mathbf{p}_2, \mathbf{q}_2); M)$.

For more details, the reader is referred to Landau and Lifshitz (1976), Arnol'd (1989), and Gallavotti (1983).

Canonical Transformations of Phase Space Coordinates

The Hamiltonian form, [13], of the equations of motion turns out to be quite useful in several problems. It is, therefore, important to remark that it is invariant under a special class of transformations of coordinates, called canonical transformations.

Consider a local change of coordinates on phase space, i.e., a smooth, smoothly invertible map $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$ between an open set U in the phase space of a Hamiltonian system with ℓ degrees of freedom, into an open set U' in a 2ℓ -dimensional space. The change of coordinates is said to be *canonical* if for any solution $t \rightarrow (\boldsymbol{\pi}(t), \boldsymbol{\kappa}(t))$ of equations like [12], for any Hamiltonian $\mathcal{H}(\boldsymbol{\pi}, \boldsymbol{\kappa})$ defined on U , the \mathcal{C} -image $t \rightarrow (\boldsymbol{\pi}'(t), \boldsymbol{\kappa}'(t)) = \mathcal{C}(\boldsymbol{\pi}(t), \boldsymbol{\kappa}(t))$ is a solution of [12] with the "same" Hamiltonian, that is, with Hamiltonian $\mathcal{H}'(\boldsymbol{\pi}', \boldsymbol{\kappa}') \stackrel{\text{def}}{=} \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{\pi}', \boldsymbol{\kappa}'))$.

The condition that a transformation of coordinates is canonical is obtained by using the arbitrariness of the function \mathcal{H} and is simply expressed as a necessary and sufficient property of the Jacobian L ,

$$L = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad [15]$$

$$A_{ij} = \partial_{\pi_i} \pi'_j, \quad B_{ij} = \partial_{\kappa_i} \pi'_j,$$

$$C_{ij} = \partial_{\pi_i} \kappa'_j, \quad D_{ij} = \partial_{\kappa_i} \kappa'_j$$

where $i, j = 1, \dots, \ell$. Let

$$E = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

denote the $2\ell \times 2\ell$ matrix formed by four $\ell \times \ell$ blocks, equal to the 0 matrix or, as indicated, to the \pm (identity matrix); then, if a superscript T denotes matrix transposition, the condition that the map be canonical is that

$$L^{-1} = EL^T E^T \text{ or } L^{-1} = \begin{pmatrix} D^T & -B^T \\ -C^T & A^T \end{pmatrix} \quad [16]$$

which immediately implies that $\det L = \pm 1$. In fact, it is possible to show that [16] implies $\det L = 1$. Equation [16] is equivalent to the four relations $AD^T - BC^T = 1$, $-AB^T + BA^T = 0$, $CD^T - DC^T = 0$, and $-CB^T + DA^T = 1$. More explicitly, since the first and the fourth relations coincide, these can be expressed as

$$\{\pi'_i, \kappa'_j\} = \delta_{ij}, \quad \{\pi'_i, \pi'_j\} = 0, \quad \{\kappa'_i, \kappa'_j\} = 0 \quad [17]$$

where, for any two functions $F(\boldsymbol{\pi}, \boldsymbol{\kappa})$, $G(\boldsymbol{\pi}, \boldsymbol{\kappa})$, the Poisson bracket is

$$\{F, G\}(\boldsymbol{\pi}, \boldsymbol{\kappa}) \stackrel{\text{def}}{=} \sum_{k=1}^{\ell} (\partial_{\pi_k} F(\boldsymbol{\pi}, \boldsymbol{\kappa}) \partial_{\kappa_k} G(\boldsymbol{\pi}, \boldsymbol{\kappa}) - \partial_{\kappa_k} F(\boldsymbol{\pi}, \boldsymbol{\kappa}) \partial_{\pi_k} G(\boldsymbol{\pi}, \boldsymbol{\kappa})) \quad [18]$$

The latter satisfies *Jacobi's identity*: $\{\{F, G\}, Q\} + \{\{G, Q\}, F\} + \{\{Q, F\}, G\} = 0$, for any three functions F, G, Q on the phase space. It is quite useful to remark that if $t \rightarrow (\boldsymbol{p}(t), \boldsymbol{q}(t)) = S_t(\boldsymbol{p}, \boldsymbol{q})$ is a solution to Hamilton equations with Hamiltonian \mathcal{H} then, given any observable $F(\boldsymbol{p}, \boldsymbol{q})$, it “evolves” as $F(t) \stackrel{\text{def}}{=} F(\boldsymbol{p}(t), \boldsymbol{q}(t))$ satisfying

$$\partial_t F(\boldsymbol{p}(t), \boldsymbol{q}(t)) = \{\mathcal{H}, F\}(\boldsymbol{p}(t), \boldsymbol{q}(t))$$

Requiring the latter identity to hold for all observables F is equivalent to requiring that the $t \rightarrow (\boldsymbol{p}(t), \boldsymbol{q}(t))$ be a solution of Hamilton's equations for \mathcal{H} .

Let $\mathcal{C}: U \longleftrightarrow U'$ be a smooth, smoothly invertible transformation between two open 2ℓ -dimensional sets: $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$. Suppose that there is a function $\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}')$ defined on a suitable domain W such that

$$\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}') \Rightarrow \begin{cases} \boldsymbol{\pi} = \partial_{\boldsymbol{\kappa}'} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}') \\ \boldsymbol{\kappa}' = \partial_{\boldsymbol{\pi}'} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}') \end{cases} \quad [19]$$

then \mathcal{C} is canonical. This is because [19] implies that if $\boldsymbol{\kappa}, \boldsymbol{\pi}'$ are varied and if $\boldsymbol{\pi}, \boldsymbol{\kappa}', \boldsymbol{\pi}', \boldsymbol{\kappa}$ are related by $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$, then $\boldsymbol{\pi} \cdot d\boldsymbol{\kappa} + \boldsymbol{\kappa}' \cdot d\boldsymbol{\pi}' = d\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}')$, which implies that

$$\boldsymbol{\pi} \cdot d\boldsymbol{\kappa} - \mathcal{H}(\boldsymbol{\pi}, \boldsymbol{\kappa}) dt \equiv \boldsymbol{\pi}' \cdot d\boldsymbol{\kappa}' - \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{\pi}', \boldsymbol{\kappa}')) dt + d\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}') - d(\boldsymbol{\pi}' \cdot \boldsymbol{\kappa}') \quad [20]$$

It means that the Hamiltonians $\mathcal{H}(\boldsymbol{p}, \boldsymbol{q})$ and $\mathcal{H}'(\boldsymbol{p}', \boldsymbol{q}')$ $\stackrel{\text{def}}{=} \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{p}', \boldsymbol{q}'))$ have Hamilton actions $\mathcal{A}_{\mathcal{H}}$ and $\mathcal{A}_{\mathcal{H}'}$ differing by a constant, if evaluated on corresponding motions $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ and $(\boldsymbol{p}'(t), \boldsymbol{q}'(t)) = \mathcal{C}(\boldsymbol{p}(t), \boldsymbol{q}(t))$.

The constant depends only on the initial and final values $(\boldsymbol{p}(t_1), \boldsymbol{q}(t_1))$ and $(\boldsymbol{p}(t_2), \boldsymbol{q}(t_2))$ and, respectively, $(\boldsymbol{p}'(t_1), \boldsymbol{q}'(t_1))$ and $(\boldsymbol{p}'(t_2), \boldsymbol{q}'(t_2))$ so that if $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ makes $\mathcal{A}_{\mathcal{H}}$ extreme, then $(\boldsymbol{p}'(t), \boldsymbol{q}'(t)) = \mathcal{C}(\boldsymbol{p}(t), \boldsymbol{q}(t))$ also makes $\mathcal{A}_{\mathcal{H}'}$ extreme.

Hence, if $t \rightarrow (\boldsymbol{p}(t), \boldsymbol{q}(t))$ solves the Hamilton equations with Hamiltonian $\mathcal{H}(\boldsymbol{p}, \boldsymbol{q})$ then the motion $t \rightarrow (\boldsymbol{p}'(t), \boldsymbol{q}'(t)) = \mathcal{C}(\boldsymbol{p}(t), \boldsymbol{q}(t))$ solves the Hamilton equations with Hamiltonian $\mathcal{H}'(\boldsymbol{p}', \boldsymbol{q}') = \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{p}', \boldsymbol{q}'))$ no matter which it is: therefore, the transformation is canonical. The function Φ is called its generating function.

Equation [19] provides a way to construct canonical maps. Suppose that a function $\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}')$ is given and defined on some domain W ; then setting

$$\begin{cases} \boldsymbol{\pi} = \partial_{\boldsymbol{\kappa}'} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}') \\ \boldsymbol{\kappa}' = \partial_{\boldsymbol{\pi}'} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}') \end{cases}$$

and inverting the first equation in the form $\boldsymbol{\pi}' = \boldsymbol{\Xi}(\boldsymbol{\pi}, \boldsymbol{\kappa})$ and substituting the value for $\boldsymbol{\pi}'$ thus obtained, in the second equation, a map $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$ is defined on some domain (where the mentioned operations can be performed) and if such domain is open and not empty then \mathcal{C} is a canonical map.

For similar reasons, if $\Gamma(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$ is a function defined on some domain then setting $\boldsymbol{\pi} = \partial_{\boldsymbol{\kappa}'} \Gamma(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$, $\boldsymbol{\pi}' = -\partial_{\boldsymbol{\kappa}} \Gamma(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$ and solving the first relation to express $\boldsymbol{\kappa}' = \boldsymbol{\Delta}(\boldsymbol{\pi}, \boldsymbol{\kappa})$ and substituting in the second relation a map $(\boldsymbol{\pi}', \boldsymbol{\kappa}') = \mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa})$ is defined on some domain (where the mentioned operations can be performed) and if such domain is open and not empty then \mathcal{C} is a canonical map.

Likewise, canonical transformations can be constructed starting from *a priori* given functions $F(\boldsymbol{\pi}, \boldsymbol{\kappa}')$ or $G(\boldsymbol{\pi}, \boldsymbol{\pi}')$. And the most general canonical map can be generated locally (i.e., near a given point in phase space) by a single one of the above four ways, possibly composed with a few “trivial” canonical maps in which one pair of coordinates (π_i, κ_i) is transformed into $(-\kappa_i, \pi_i)$. The necessity of also including the trivial maps can be traced to the existence of homogeneous canonical maps, that is, maps such that $\boldsymbol{\pi} \cdot d\boldsymbol{\kappa} = \boldsymbol{\pi}' \cdot d\boldsymbol{\kappa}'$ (e.g., the identity map, see below or [49] for nontrivial examples) which are action preserving hence canonical, but which evidently cannot be generated by a function $\Phi(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$ although they can be generated by a function depending on $\boldsymbol{\pi}', \boldsymbol{\kappa}$.

Simple examples of homogeneous canonical maps are maps in which the coordinates q are changed into $q' = R(q)$ and, correspondingly, the p 's are transformed as $p' = (\partial_q R(q))^{-1T} p$, linearly: indeed, this map is generated by the function $F(p', q) \stackrel{\text{def}}{=} p' \cdot R(q)$.

For instance, consider the map ‘‘Cartesian–polar’’ coordinates $(q_1, q_2) \longleftrightarrow (\rho, \theta)$ with (ρ, θ) the polar coordinates of q (namely $\rho = \sqrt{q_1^2 + q_2^2}, \theta = \arctan(q_2/q_1)$) and let $n \stackrel{\text{def}}{=} q/|q| = (n_1, n_2)$ and $t = (-n_2, n_1)$. Setting $p_\rho \stackrel{\text{def}}{=} p \cdot n, p_\theta \stackrel{\text{def}}{=} \rho p \cdot t$, the map $(p_1, p_2, q_1, q_2) \longleftrightarrow (p_\rho, p_\theta, \rho, \theta)$ is homogeneous canonical (because $p \cdot dq = p \cdot n d\rho + p \cdot t \rho d\theta = p_\rho d\rho + p_\theta d\theta$).

As a further example, any area-preserving map $(p, q) \longleftrightarrow (p', q')$ defined on an open region of the plane \mathbb{R}^2 is canonical: because in this case the matrices A, B, C, D are just numbers, which satisfy $AD - BC = 1$ and, therefore, [16] holds.

For more details, the reader is referred to Landau and Lifshitz (1976) and Gallavotti (1983).

Quadratures

The simplest mechanical systems are integrable by quadratures. For instance, the Hamiltonian on \mathbb{R}^2 ,

$$\mathcal{H}(p, q) = \frac{1}{2m} p^2 + V(q) \quad [21]$$

generates a motion $t \rightarrow q(t)$ with initial data q_0, \dot{q}_0 such that $\mathcal{H}(p_0, q_0) = E$, i.e., $\frac{1}{2} m \dot{q}_0^2 + V(q_0) = E$, satisfying

$$\dot{q}(t) = \pm \sqrt{\frac{2}{m} (E - V(q(t)))}$$

If the equation $E = V(q)$ has only two solutions $q_-(E) < q_+(E)$ and $|\partial_q V(q_\pm(E))| > 0$, the motion is periodic with period

$$T(E) = 2 \int_{q_-(E)}^{q_+(E)} \frac{dx}{\sqrt{(2/m)(E - V(x))}} \quad [22]$$

The special solution with initial data $q_0 = q_-(E), \dot{q}_0 = 0$ will be denoted $Q(t)$, and it is an analytic function (by the general regularity theorem on ordinary differential equations). For $0 \leq t \leq T/2$ or for $T/2 \leq t \leq T$ it is given, respectively, by

$$t = \int_{q_-(E)}^{Q(t)} \frac{dx}{\sqrt{(2/m)(E - V(x))}} \quad [23a]$$

or

$$t = \frac{T}{2} - \int_{Q(t)}^{q_+(E)} \frac{dx}{\sqrt{(2/m)(E - V(x))}} \quad [23b]$$

The most general solution with energy E has the form $q(t) = Q(t_0 + t)$, where t_0 is defined by $q_0 = Q(t_0), \dot{q}_0 = \dot{Q}(t_0)$, i.e., it is the time needed for the ‘‘standard solution’’ $Q(t)$ to reach the initial data for the new motion.

If the derivative of V vanishes in one of the extremes or if at least one of the two solutions $q_\pm(E)$ does not exist, the motion is not periodic and it may be unbounded: nevertheless, it is still expressible via integrals of the type [22]. If the potential V is periodic in q and the variable q is considered to be varying on a circle then essentially all solutions are periodic: exceptions can occur if the energy E has a value such that $V(q) = E$ admits a solution where V has zero derivative.

Typical examples are the harmonic oscillator, the pendulum, and the Kepler oscillator: whose Hamiltonians, if m, ω, g, h, G, k are positive constants, are, respectively,

$$\begin{aligned} & \frac{p^2}{2m} + \frac{1}{2} m \omega^2 q^2 \\ & \frac{p^2}{2m} + mg \left(1 - \cos \frac{q}{h} \right) \\ & \frac{p^2}{2m} - mk \frac{1}{|q|} + m \frac{G^2}{2q^2} \end{aligned} \quad [24]$$

the Kepler oscillator Hamiltonian has a potential which is singular at $q = 0$ but if $G \neq 0$ the energy conservation forbids too close an approach to $q = 0$ and the singularity becomes irrelevant.

The integral in [23] is called a *quadrature* and the systems in [21] are therefore *integrable by quadratures*. Such systems, at least when the motion is periodic, are best described in new coordinates in which periodicity is more manifest. Namely when $V(q) = E$ has only two roots $q_\pm(E)$ and $\mp V'(q_\pm(E)) > 0$ the *energy–time coordinates* can be used by replacing q, \dot{q} or p, q by E, τ , where τ is the time needed for the standard solution $t \rightarrow Q(t)$ to reach the given data, that is, $Q(\tau) = q, \dot{Q}(\tau) = \dot{q}$. In such coordinates, the motion is simply $(E, \tau) \rightarrow (E, \tau + t)$ and, of course, the variable τ has to be regarded as varying on a circle of radius $T/2\pi$. The E, τ variables are a kind of polar coordinates, as can be checked by drawing the curves of constant E , ‘‘energy levels,’’ in the plane p, q in the cases in [24]; see Figure 1.

In the harmonic oscillator case, all trajectories are periodic. In the pendulum case, all motions are periodic except the ones which separate the oscillatory motions (the closed curves in the second drawing) from the rotatory motions (the apparently open curves) which, in fact, are on closed curves as well if the q coordinate, that is, the vertical

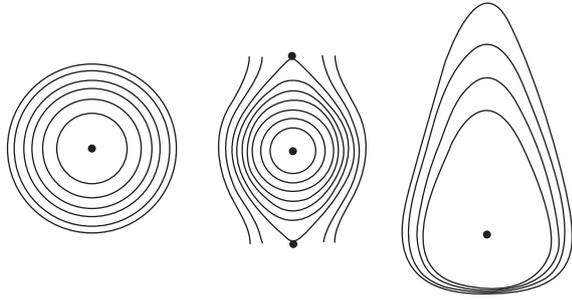


Figure 1 The energy levels of the harmonic oscillator, the pendulum, and the Kepler motion.

coordinate in **Figure 1**, is regarded as “periodic” with period $2\pi h$. In the Kepler case, only the negative-energy trajectories are periodic and a few of them are drawn in **Figure 1**. The single dots represent the equilibrium points in phase space.

The region of phase space where motions are periodic is a set of points (p, q) with the topological structure of $\cup_{u \in U} (\{u\} \times C_u)$, where u is a coordinate varying in an open interval U (e.g., the set of values of the energy), and C_u is a closed curve whose points (p, q) are identified by a coordinate (e.g., by the time necessary for an arbitrarily fixed datum with the same energy to evolve into (p, q)).

In the above cases, [24], if the “radial” coordinate is chosen to be the energy the set U is the interval $(0, +\infty)$ for the harmonic oscillator, $(0, 2mg)$ or $(2mg, +\infty)$ for the pendulum, and $(-\frac{1}{2}mk^2/G^2, 0)$ in the Kepler case. The fixed datum for the reference motion can be taken, in all cases, to be of the form $(0, q_0)$ with the time coordinate t_0 given by [23].

It is remarkable that the energy–time coordinates are canonical coordinates: for instance, in the vicinity of (p_0, q_0) and if $p_0 > 0$, this can be seen by setting

$$S(q, E) = \int_{q_0}^q \sqrt{2m(E - V(x))} dx \quad [25]$$

and checking that $p = \partial_q S(q, E)$, $t = \partial_E S(q, E)$ are identities if (p, q) and (E, t) are coordinates for the same point so that the criterion expressed by [20] applies.

It is convenient to standardize the coordinates by replacing the time variable by an angle $\alpha = (2\pi/T(E))t$; and instead of the energy any invertible function of it can be used.

It is natural to look for a coordinate $A = A(E)$ such that the map $(p, q) \longleftrightarrow (A, \alpha)$ is a canonical map: this is easily done as the function

$$\hat{S}(q, A) = \int_{q_0}^q \sqrt{2m(E(A) - V(x))} dx \quad [26]$$

generates (locally) the correspondence between $p = \sqrt{2m(E(A) - V(q))}$ and

$$\alpha = E'(A) \int_0^q \frac{dx}{\sqrt{2m^{-1}(E(A) - V(x))}}$$

Therefore, by the criterion [20], if

$$E'(A) = \frac{2\pi}{T(E(A))}$$

i.e., if $A'(E) = T(E)/2\pi$, the coordinates (A, α) will be canonical coordinates. Hence, by [22], $A(E)$ can be taken as

$$\begin{aligned} A &= \frac{1}{2\pi} 2 \int_{q_-(E)}^{q_+(E)} \sqrt{2m(E - V(q))} dq \\ &\equiv \frac{1}{2\pi} \oint p dq \end{aligned} \quad [27]$$

where the last integral is extended to the closed curve of energy E ; see **Figure 1**. The *action–angle coordinates* (A, α) are defined in open regions of phase space covered by periodic motions: in action–angle coordinates such regions have the form $W = J \times \mathbb{T}$ of a product of an open interval J and a one-dimensional “torus” $\mathbb{T} = [0, 2\pi]$ (i.e., a unit circle).

For details, the reader is again referred to [Landau and Lifshitz \(1976\)](#), [Arnol’d \(1989\)](#), and [Gallavotti \(1983\)](#).

Quasiperiodicity and Integrability

A Hamiltonian is called *integrable* in an open region $W \subset T^*(M)$ of phase space if

1. there is an analytic and nonsingular (i.e., with nonzero Jacobian) change of coordinates $(p, q) \longleftrightarrow (I, \varphi)$ mapping W into a set of the form $\mathcal{I} \times \mathbb{T}^\ell$ with $\mathcal{I} \subset \mathbb{R}^\ell$ (open); and furthermore
2. the flow $t \rightarrow S_t(p, q)$ on phase space is transformed into $(I, \varphi) \rightarrow (I, \varphi + \omega(I)t)$ where $\omega(I)$ is a smooth function on \mathcal{I} .

This means that, in suitable coordinates, which can be called “integrating coordinates,” the system appears as a set of ℓ points with coordinates $\varphi = (\varphi_1, \dots, \varphi_\ell)$ moving on a unit circle at angular velocities $\omega(I) = (\omega_1(I), \dots, \omega_\ell(I))$ depending on the actions of the initial data.

A system integrable in a region W which, in integrating coordinates I, φ , has the form $\mathcal{I} \times \mathbb{T}^\ell$ is said to be *anisochronous* if $\det \partial_I \omega(I) \neq 0$. It is said to be *isochronous* if $\omega(I) \equiv \omega$ is independent of I . The motions of integrable systems are called *quasiperiodic* with frequency spectrum $\omega(I)$, or with *frequencies* $\omega(I)/2\pi$, in the coordinates (I, φ) .

Clearly, an integrable system admits ℓ independent constants of motion, the $I = (I_1, \dots, I_\ell)$, and, for each

choice of \mathbf{I} , the other coordinates vary on a “standard” ℓ -dimensional torus \mathbb{T}^ℓ : hence, it is possible to say that a phase space region of integrability is *foliated* into ℓ -dimensional invariant tori $\mathcal{T}(\mathbf{I})$ parametrized by the values of the constants of motion $\mathbf{I} \in \mathcal{I}$.

If an integrable system is anisochronous then it is *canonically integrable*: that is, it is possible to define on W a canonical change of coordinates $(\mathbf{p}, \mathbf{q}) = \mathcal{C}(\mathbf{A}, \boldsymbol{\alpha})$ mapping W onto $J \times \mathbb{T}^\ell$ and such that $\mathcal{H}(\mathcal{C}(\mathbf{A}, \boldsymbol{\alpha})) = h(\mathbf{A})$ for a suitable h . Then, if $\boldsymbol{\omega}(\mathbf{A}) \stackrel{\text{def}}{=} \partial_{\mathbf{A}} h(\mathbf{A})$, the equations of motion become

$$\dot{\mathbf{A}} = \mathbf{0}, \quad \dot{\boldsymbol{\alpha}} = \boldsymbol{\omega}(\mathbf{A}) \quad [28]$$

Given a system $(\mathbf{I}, \boldsymbol{\varphi})$ of coordinates integrating an anisochronous system the construction of action–angle coordinates can be performed, in principle, via a classical procedure (under a few extra assumptions).

Let $\gamma_1, \dots, \gamma_\ell$ be ℓ topologically independent circles on \mathbb{T}^ℓ , for definiteness let $\gamma_i(\mathbf{I}) = \{\boldsymbol{\varphi} \mid \varphi_1 = \varphi_2 = \dots = \varphi_{i-1} = \varphi_{i+1} = \dots = 0, \varphi_i \in [0, 2\pi]\}$, and set

$$A_i(\mathbf{I}) = \frac{1}{2\pi} \oint_{\gamma_i(\mathbf{I})} \mathbf{p} \cdot d\mathbf{q} \quad [29]$$

If the map $\mathbf{I} \longleftrightarrow \mathbf{A}(\mathbf{I})$ is analytically invertible as $\mathbf{I} = \mathbf{I}(\mathbf{A})$, the function

$$S(\mathbf{A}, \boldsymbol{\varphi}) = (\lambda) \int_0^\varphi \mathbf{p} \cdot d\mathbf{q} \quad [30]$$

is well defined if the integral is over any path λ joining the points $(\mathbf{p}(\mathbf{I}(\mathbf{A}), \mathbf{0}), \mathbf{q}(\mathbf{I}(\mathbf{A}), \mathbf{0}))$ and $(\mathbf{p}(\mathbf{I}(\mathbf{A}), \boldsymbol{\varphi}), \mathbf{q}(\mathbf{I}(\mathbf{A}), \boldsymbol{\varphi}))$ and lying on the torus parametrized by $\mathbf{I}(\mathbf{A})$.

The key remark in the proof that [30] really defines a function of the only variables $\mathbf{A}, \boldsymbol{\varphi}$ is that anisochrony implies the vanishing of the Poisson brackets (cf. [18]): $\{I_i, I_j\} = 0$ (hence also $\{A_i, A_j\} \equiv \sum_{b,k} \partial_{I_k} A_i \partial_{I_b} A_j \{I_k, I_b\} = 0$). And the property $\{I_i, I_j\} = 0$ can be checked to be precisely the integrability condition for the differential form $\mathbf{p} \cdot d\mathbf{q}$ restricted to the surface obtained by varying \mathbf{q} while \mathbf{p} is constrained so that (\mathbf{p}, \mathbf{q}) stays on the surface $\mathbf{I} = \text{constant}$, i.e., on the invariant torus of the points with fixed \mathbf{I} .

The latter property is necessary and sufficient in order that the function $S(\mathbf{A}, \boldsymbol{\varphi})$ be well defined (i.e., be independent on the integration path λ) up to an additive quantity of the form $\sum_i 2\pi n_i A_i$ with $\mathbf{n} = (n_1, \dots, n_\ell)$ integers.

Then the action–angle variables are defined by the canonical change of coordinates with $S(\mathbf{A}, \boldsymbol{\varphi})$ as generating function, i.e., by setting

$$\alpha_i = \partial_{A_i} S(\mathbf{A}, \boldsymbol{\varphi}), \quad I_i = \partial_{\varphi_i} S(\mathbf{A}, \boldsymbol{\varphi}) \quad [31]$$

and, since the computation of $S(\mathbf{A}, \boldsymbol{\varphi})$ is “reduced to integrations” which can be regarded as a natural extension of the quadratures discussed in the one-dimensional cases, such systems are also called *integrable by quadratures*. The just-described construction is a version of the more general *Arnol’d–Liouville theorem*.

In practice, however, the actual evaluation of the integrals in [29], [30] can be difficult: its analysis in various cases (even as “elementary” as the pendulum) has in fact led to key progress in various domains, for example, in the theory of special functions and in group theory.

In general, any surface on phase space on which the restriction of the differential form $\mathbf{p} \cdot d\mathbf{q}$ is locally integrable is called a *Lagrangian manifold*: hence the invariant tori of an anisochronous integrable system are Lagrangian manifolds.

If an integrable system is anisochronous, it cannot admit more than ℓ independent constants of motion; furthermore, it does not admit invariant tori of dimension $> \ell$. Hence ℓ -dimensional invariant tori are called maximal.

Of course, invariant tori of dimension $< \ell$ can also exist: this happens when the variables \mathbf{I} are such that the frequencies $\boldsymbol{\omega}(\mathbf{I})$ admit nontrivial rational relations; i.e., there is an integer components vector $\mathbf{v} \in \mathbb{Z}^\ell$, $\mathbf{v} = (v_1, \dots, v_\ell) \neq \mathbf{0}$ such that

$$\boldsymbol{\omega}(\mathbf{I}) \cdot \mathbf{v} = \sum_i \omega_i(\mathbf{I}) v_i = 0 \quad [32]$$

in this case, the invariant torus $\mathcal{T}(\mathbf{I})$ is called *resonant*. If the system is anisochronous then $\det \partial_{\mathbf{I}} \boldsymbol{\omega}(\mathbf{I}) \neq 0$ and, therefore, the resonant tori are associated with values of the constants of motion \mathbf{I} which form a set of measure zero in the space \mathcal{I} but which is not empty and dense.

Examples of isochronous systems are the systems of harmonic oscillators, i.e., systems with Hamiltonian

$$\sum_{i=1}^{\ell} \frac{1}{2m_i} p_i^2 + \frac{1}{2} \sum_{i,j}^{1,\ell} c_{ij} q_i q_j$$

where the matrix ν is a positive-definite matrix. This is an isochronous system with frequencies $\boldsymbol{\omega} = (\omega_1, \dots, \omega_\ell)$ whose squares are the eigenvalues of the matrix $m_i^{-1/2} c_{ij} m_j^{-1/2}$. It is integrable in the region W of the data $\mathbf{x} = (\mathbf{p}, \mathbf{q}) \in \mathbb{R}^{2\ell}$ such that, setting

$$A_\beta = \frac{1}{2\omega_\beta} \left(\left(\sum_{i=1}^{\ell} \frac{v_{\beta,i} p_i}{\sqrt{m_i}} \right)^2 + \omega_\beta^2 \left(\sum_{i=1}^{\ell} \frac{v_{\beta,i} q_i}{\sqrt{m_i^{-1}}} \right)^2 \right)$$

for all eigenvectors \mathbf{v}_β , $\beta = 1, \dots, \ell$, of the above matrix, the vectors \mathbf{A} have all components > 0 .

Even though this system is isochronous, it nevertheless admits a system of canonical action–angle coordinates in which the Hamiltonian takes the simplest form

$$h(\mathbf{A}) = \sum_{\beta=1}^{\ell} \omega_{\beta} A_{\beta} \equiv \boldsymbol{\omega} \cdot \mathbf{A} \quad [33]$$

with

$$\alpha_{\beta} = -\arctan \left(\frac{\sum_{i=1}^{\ell} \frac{v_{\beta,i} p_i}{\sqrt{m_i}}}{\sum_{i=1}^{\ell} \sqrt{m_i} \omega_{\beta} v_{\beta,i} q_i} \right)$$

as conjugate angles.

An example of anisochronous system is the *free rotators* or *free wheels*: i.e., ℓ noninteracting points on a circle of radius R or ℓ noninteracting homogeneous coaxial wheels of radius R . If $J_i = m_i R^2$ or, respectively, $J_i = (1/2)m_i R^2$ are the inertia moments and if the positions are determined by ℓ angles $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{\ell})$, the angular velocities are constants related to the angular momenta $\mathbf{A} = (A_1, \dots, A_{\ell})$ by $\omega_i = A_i/J_i$. The Hamiltonian and the spectrum are

$$h(\mathbf{A}) = \sum_{i=1}^{\ell} \frac{1}{2J_i} A_i^2, \quad \boldsymbol{\omega}(\mathbf{A}) = \left(\frac{1}{J_i} A_i \right)_{i=1, \dots, \ell} \quad [34]$$

For further details see Landau and Lifshitz (1976), Gallavotti (1983), Arnol'd (1989), and Fassò (1998).

Multidimensional Quadratures: Central Motion

Several important mechanical systems with more than one degree of freedom are integrable by canonical quadratures in vast regions of phase space. This is checked by showing that there is a foliation into invariant tori $\mathcal{T}(I)$ of dimension equal to the number of degrees of freedom (ℓ) parametrized by ℓ constants of motion I in involution, i.e., such that $\{I_i, I_j\} = 0$. One then performs, if possible, the construction of the action–angle variables by the quadratures discussed in the previous section.

The above procedure is well illustrated by the theory of the planar motion of a unit mass attracted by a coplanar center of force: the Lagrangian is, in polar coordinates (ρ, θ) ,

$$\mathcal{L} = \frac{m}{2} (\dot{\rho}^2 + \rho^2 \dot{\theta}^2) - V(\rho)$$

The planarity of the motion is not a strong restriction as central motion always takes place on a plane.

Hence, the equations of motion are

$$\frac{d}{dt} m \rho^2 \dot{\theta} = 0$$

i.e., $m \rho^2 \dot{\theta} = G$ is a constant of motion (it is the angular momentum), and

$$\begin{aligned} \ddot{\rho} &= -\partial_{\rho} V(\rho) + \partial_{\rho} \frac{m}{2} \rho^2 \dot{\theta}^2 \\ &= -\partial_{\rho} V(\rho) + \frac{G^2}{m \rho^3} \\ &\stackrel{\text{def}}{=} -\partial_{\rho} V_G(\rho) \end{aligned}$$

Then the energy conservation yields a second constant of motion E ,

$$\begin{aligned} \frac{m}{2} \dot{\rho}^2 + \frac{1}{2} \frac{G^2}{m \rho^2} + V(\rho) &= E \\ &= \frac{1}{2m} p_{\rho}^2 + \frac{1}{2m} \frac{p_{\theta}^2}{\rho^2} + V(\rho) \end{aligned} \quad [35]$$

The right-hand side (rhs) is the Hamiltonian for the system, derived from \mathcal{L} , if p_{ρ} , p_{θ} denote conjugate momenta of ρ , θ : $p_{\rho} = m \dot{\rho}$ and $p_{\theta} = m \rho^2 \dot{\theta}$ (note that $p_{\theta} = G$).

Suppose $\rho^2 V(\rho) \xrightarrow{\rho \rightarrow 0} 0$: then the singularity at the origin cannot be reached by any motion starting with $\rho > 0$ if $G > 0$. Assume also that the function

$$V_G(\rho) \stackrel{\text{def}}{=} \frac{1}{2} \frac{G^2}{m \rho^2} + V(\rho)$$

has only one minimum $E_0(G)$, no maximum and no horizontal inflection, and tends to a limit $E_{\infty}(G) \leq \infty$ when $\rho \rightarrow \infty$. Then the system is integrable in the domain $W = \{(p, q) \mid E_0(G) < E < E_{\infty}(G), G \neq 0\}$.

This is checked by introducing a “standard” periodic solution $t \rightarrow R(t)$ of $m \ddot{\rho} = -\partial_{\rho} V_G(\rho)$ with energy $E_0(G) < E < E_{\infty}(G)$ and initial data $\rho = \rho_{E_{-}}(G)$, $\dot{\rho} = 0$ at time $t = 0$, where $\rho_{E, \pm}(G)$ are the two solutions of $V_G(\rho) = E$, see the section “Quadratures”: this is a periodic analytic function of t with period

$$T(E, G) = 2 \int_{\rho_{E_{-}}(G)}^{\rho_{E_{+}}(G)} \frac{dx}{\sqrt{(2/m)(E - V_G(x))}}$$

The function $R(t)$ is given, for $0 \leq t \leq \frac{1}{2} T(E, G)$ or for $\frac{1}{2} T(E, G) \leq t \leq T(E, G)$, by the quadratures

$$t = \int_{\rho_{E_{-}}(G)}^{R(t)} \frac{dx}{\sqrt{(2/m)(E - V_G(x))}} \quad [36a]$$

or

$$t = \frac{T(E, G)}{2} - \int_{R(t)}^{\rho_{E_{+}}(G)} \frac{dx}{\sqrt{(2/m)(E - V_G(x))}} \quad [36b]$$

respectively. The analytic regularity of $R(t)$ follows from the general existence, uniqueness, and regularity theorems applied to the differential equation for $\ddot{\rho}$.

Given an initial datum $\dot{\rho}_0, \rho_0, \dot{\theta}_0, \theta_0$ with energy E and angular momentum G , define t_0 to be the time such that $R(t_0) = \rho_0, \dot{R}(t_0) = \dot{\rho}_0$: then $\rho(t) \equiv R(t + t_0)$ and $\theta(t)$ can be computed as

$$\theta(t) = \theta_0 + \int_0^t \frac{G}{mR(t' + t_0)^2} dt'$$

a second quadrature. Therefore, we can use as coordinates for the motion E, G, t_0 , which determine $\dot{\rho}_0, \rho_0, \dot{\theta}_0$ and a fourth coordinate that determines θ_0 which could be θ_0 itself but which is conveniently determined, via the second quadrature, as follows.

The function $Gm^{-1}R(t)^{-2}$ is periodic with period $T(E, G)$; hence it can be expressed in a Fourier series

$$\chi_0(E, G) + \sum_{k \neq 0} \chi_k(E, G) \exp\left(\frac{2\pi}{T(E, G)} itk\right)$$

the quadrature for $\theta(t)$ can be performed by integrating the series terms. Setting

$$\bar{\theta}(t_0) \stackrel{\text{def}}{=} \frac{T(E, G)}{2\pi} \sum_{k \neq 0} \frac{\chi_k(E, G)}{k} \exp\left(\frac{2\pi}{T(E, G)} it_0 k\right)$$

and $\varphi_1(0) = \theta_0 - \bar{\theta}(t_0)$, the expression

$$\theta(t) = \theta_0 + \int_0^t \frac{G}{mR(t' + t_0)^2} dt'$$

becomes

$$\varphi_1(t) = \varphi_1(0) + \chi_0(E, G) t \quad [37]$$

Hence the system is integrable and the spectrum is $\boldsymbol{\omega}(E, G) = (\omega_0(E, G), \omega_1(E, G)) \equiv (\omega_0, \omega_1)$ with

$$\omega_0 \stackrel{\text{def}}{=} \frac{2\pi}{T(E, G)} \quad \text{and} \quad \omega_1 \stackrel{\text{def}}{=} \chi_0(E, G)$$

while $\boldsymbol{I} = (E, G)$ are constants of motion and the angles $\boldsymbol{\varphi} = (\varphi_0, \varphi_1)$ can be taken as

$$\varphi_0 \stackrel{\text{def}}{=} \omega_0 t_0, \quad \varphi_1 \stackrel{\text{def}}{=} \theta_0 - \bar{\theta}(t_0)$$

At E, G fixed, the motion takes place on a two-dimensional torus $\mathcal{T}(E, G)$ with φ_0, φ_1 as angles.

In the anisochronous cases, i.e., when $\det \partial_{E, G} \boldsymbol{\omega}(E, G) \neq 0$, canonical action–angle variables conjugated to $(p_\rho, \rho, p_\theta, \theta)$ can be constructed via [29], [30] by using two cycles γ_1, γ_2 on the torus $\mathcal{T}(E, G)$. It is convenient to choose

1. γ_1 as the cycle consisting of the points $\rho = x, \theta = 0$ whose first half (where $p_\rho \geq 0$) consists in the set $\rho_{E, -(G)} \leq x \leq \rho_{E, +(G)}, p_\rho = \sqrt{2m(E - V_G(x))}$ and $d\theta = 0$; and

2. γ_2 as the cycle $\rho = \text{const}, \theta \in [0, 2\pi]$ on which $d\rho = 0$ and $p_\theta = G$ obtaining

$$A_1 = \frac{2}{2\pi} \int_{\rho_{E, -(G)}}^{\rho_{E, +(G)}} \sqrt{2m(E - V_G(x))} dx, \quad [38]$$

$$A_2 = G$$

According to the general theory (cf. the previous section) a generating function for the canonical change of coordinates from $(p_\rho, \rho, p_\theta, \theta)$ to action–angle variables is (if, to fix ideas, $p_\rho > 0$)

$$S(A_1, A_2, \rho, \theta) = G\theta + \int_{\rho_{E, -}}^{\rho} \sqrt{2m(E - V_G(x))} dx \quad [39]$$

In terms of the above ω_0, χ_0 the Jacobian matrix $\partial(E, G)/\partial(A_1, A_2)$ is computed from [38], [39] to be $\begin{pmatrix} \omega_0 & \chi_0 \\ 0 & 1 \end{pmatrix}$. It follows that $\partial_E S = t, \partial_G S = \theta - \bar{\theta}(t) - \chi_0 t$ so that, see [31],

$$\alpha_1 \stackrel{\text{def}}{=} \partial_{A_1} S = \omega_0 t, \quad \alpha_2 \stackrel{\text{def}}{=} \partial_{A_2} S = \theta - \bar{\theta}(t) \quad [40]$$

and $(A_1, \alpha_1), (A_2, \alpha_2)$ are the action–angle pairs.

For more details, see Landau and Lifshitz (1976) and Gallavotti (1983).

Newtonian Potential and Kepler's Laws

The anisochrony property, that is, $\det \partial(\omega_0, \chi_0)/\partial(A_1, A_2) \neq 0$ or, equivalently, $\det \partial(\omega_0, \chi_0)/\partial(E, G) \neq 0$, is not satisfied in the important cases of the harmonic potential and the Newtonian potential. Anisochrony being only a sufficient condition for canonical integrability it is still possible (and true) that, nevertheless, in both cases the canonical transformation generated by [39] integrates the system. This is expected since the two potentials are limiting cases of anisochronous ones (e.g., $|q|^{2+\varepsilon}$ and $|q|^{-1-\varepsilon}$ with $\varepsilon \rightarrow 0$).

The Newtonian potential

$$\mathcal{H}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2m} \boldsymbol{p}^2 - \frac{km}{|q|}$$

is integrable in the region $G \neq 0, E_0(G) = -k^2 m^3 / 2G^2 < E < 0, |G| < \sqrt{k^2 m^3 / (-2E)}$. Proceeding as in the last section, one finds integrating coordinates and that the integrable motions develop on ellipses with one focus on the center of attraction S so that motions are periodic, hence not anisochronous: nevertheless, the construction of the canonical coordinates via [29]–[31] (hence [39]) works and leads to canonical coordinates $(L', \lambda', G', \gamma')$. To obtain action–angle variables with a simple

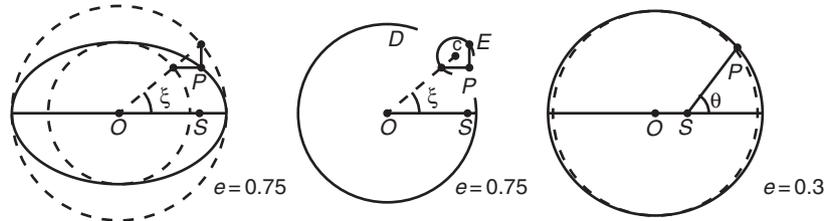


Figure 2 Eccentric and true anomalies of P , which moves on a small circle E centered at a point c moving on the circle D located half-way between the two concentric circles containing the Keplerian ellipse: the anomaly of c with respect to the axis OS is ξ . The circle D is *eccentric* with respect to S and therefore ξ is, even today, called *eccentric anomaly*, whereas the circle D is, in ancient terminology, the *deferent* circle (eccentric circles were introduced in astronomy by Ptolemy). The small circle E on which the point P moves is, in ancient terminology, an *epicycle*. The deferent and the epicyclical motions are synchronous (i.e., they have the same period); Kepler discovered that his key *a priori* hypothesis of inverse proportionality between angular velocity on the deferent and distance between P and S (i.e., $\rho\dot{\xi} = \text{constant}$) implied both synchrony and elliptical shape of the orbit, with focus in S . The latter law is equivalent to $\rho^2\dot{\theta} = \text{constant}$ (because of the identity $a\xi = \rho\theta$). Small eccentricity ellipses can hardly be distinguished from circles.

interpretation, it is convenient to perform on the variables $(L', \lambda', G', \gamma')$ (constructed by following the procedure just indicated) a further trivial canonical transformation by setting $L = L' + G'$, $G = G'$, $\lambda = \lambda'$, $\gamma = \gamma' - \lambda'$; then

1. λ (average anomaly) is the time necessary for the point P to move from the pericenter to its actual position, in units of the period, times 2π ;
2. L (action) is essentially the energy $E = -k^2 m^3 / 2L^2$;
3. G (angular momentum);
4. γ (axis longitude), is the angle between a fixed axis and the major axis of the ellipse oriented from the center of the ellipse O to the center of attraction S .

The eccentricity of the ellipse is e such that $G = \pm L\sqrt{1 - e^2}$. The ellipse equation is $\rho = a(1 - e \cos \xi)$, where ξ is the eccentric anomaly (see [Figure 2](#)), $a = L^2 / km^2$ is the major semiaxis, and ρ is the distance to the center of attraction S .

Finally, the relations between eccentric anomaly ξ , average anomaly λ , true anomaly θ (the latter is the polar angle), and SP distance ρ are given by the Kepler equations

$$\begin{aligned} \lambda &= \xi - e \sin \xi \\ (1 - e \cos \xi)(1 + e \cos \theta) &= 1 - e^2 \\ \lambda &= (1 - e^2)^{3/2} \int_0^\theta \frac{d\theta'}{(1 + e \cos \theta')^2} \\ \frac{\rho}{a} &= \frac{1 - e^2}{1 + e \cos \theta} \end{aligned} \quad [41]$$

and the relation between true anomaly and average anomaly can be inverted in the form

$$\begin{aligned} \xi &= \lambda + g_\lambda \\ \theta &= \lambda + f_\lambda \Rightarrow \frac{\rho}{a} = \frac{1 - e^2}{1 + e \cos(\lambda + f_\lambda)} \end{aligned} \quad [42]$$

where $g_\lambda = g(e \sin \lambda, e \cos \lambda)$, $f_\lambda = f(e \sin \lambda, e \cos \lambda)$, and $g(x, y), f(x, y)$ are suitable functions analytic for $|x|, |y| < 1$. Furthermore, $g(x, y) = x(1 + y + \dots)$, $f(x, y) = 2x(1 + \frac{5}{4}y + \dots)$ and the ellipses denote terms of degree 2 or higher in x, y , containing only even powers of x .

For more details, the reader is referred to [Landau and Lifshitz \(1976\)](#) and [Gallavotti \(1983\)](#).

Rigid Body

Another fundamental integrable system is the rigid body in the absence of gravity and with a fixed point O . It can be naturally described in terms of the Euler angles $\theta_0, \varphi_0, \psi_0$ (see [Figure 3](#)) and their derivatives $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$.

Let I_1, I_2, I_3 be the three *principal inertia moments* of the body along the three principal axes with unit vectors $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$. The inertia moments and the principal axes are the eigenvalues and the associated unit eigenvectors of the 3×3 inertia matrix \mathcal{I} , which is defined by $\mathcal{I}_{bk} = \sum_{i=1}^n m_i(\mathbf{x}_i)_b(\mathbf{x}_i)_k$, where $b, k = 1, 2, 3$ and \mathbf{x}_i is the position of the i th particle in a reference frame with origin at O and in which

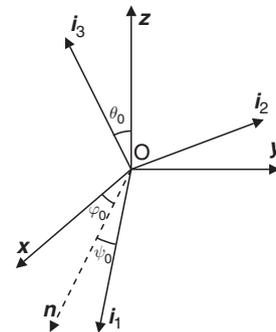


Figure 3 The *Euler angles* of the comoving frame $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ with respect to a fixed frame $\mathbf{x}, \mathbf{y}, \mathbf{z}$. The direction \mathbf{n} is the “node line”, intersection between the planes \mathbf{x}, \mathbf{y} and $\mathbf{i}_1, \mathbf{i}_2$.

all particles are at rest: this *comoving frame* exists as a consequence of the rigidity constraint. The principal axes form a coordinate system which is comoving as well: that is, in the frame $(O; \mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3)$ as well, the particles are at rest.

The Lagrangian is simply the kinetic energy: we imagine the rigidity constraint to be ideal (e.g., as realized by internal central forces in the limit of infinite rigidity, as mentioned in the section “**Lagrange and Hamilton forms of equations of motion**”). The *angular velocity* of the rigid motion is defined by

$$\boldsymbol{\omega} = \dot{\theta}_0 \mathbf{n} + \dot{\varphi}_0 \mathbf{z} + \dot{\psi}_0 \mathbf{i}_3 \quad [43]$$

expressing that a generic infinitesimal motion must consist of a variation of the three Euler angles and, therefore, it has to be a rotation of speeds $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$ around the axes $\mathbf{n}, \mathbf{z}, \mathbf{i}_3$ as shown in **Figure 3**.

Let $(\omega_1, \omega_2, \omega_3)$ be the components of $\boldsymbol{\omega}$ along the principal axes $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$: for brevity, the latter axes will often be called **1, 2, 3**. Then the angular momentum \mathbf{M} , with respect to the pivot point O , and the kinetic energy K can be checked to be

$$\begin{aligned} \mathbf{M} &= I_1 \omega_1 \mathbf{i}_1 + I_2 \omega_2 \mathbf{i}_2 + I_3 \omega_3 \mathbf{i}_3 \\ K &= \frac{1}{2} (I_1 \omega_1^2 + I_2 \omega_2^2 + I_3 \omega_3^2) \end{aligned} \quad [44]$$

and are constants of motion. From **Figure 3** it follows that $\omega_1 = \dot{\theta}_0 \cos \psi_0 + \dot{\varphi}_0 \sin \theta_0 \sin \psi_0$, $\omega_2 = -\dot{\theta}_0 \sin \psi_0 + \dot{\varphi}_0 \sin \theta_0 \cos \psi_0$ and $\omega_3 = \dot{\varphi}_0 \cos \theta_0 + \dot{\psi}_0$, so that the Lagrangian, uninspiring at first, is

$$\begin{aligned} \mathcal{L} &\stackrel{\text{def}}{=} \frac{1}{2} I_1 (\dot{\theta}_0 \cos \psi_0 + \dot{\varphi}_0 \sin \theta_0 \sin \psi_0)^2 \\ &+ \frac{1}{2} I_2 (-\dot{\theta}_0 \sin \psi_0 + \dot{\varphi}_0 \sin \theta_0 \cos \psi_0)^2 \\ &+ \frac{1}{2} I_3 (\dot{\varphi}_0 \cos \theta_0 + \dot{\psi}_0)^2 \end{aligned} \quad [45]$$

Angular momentum conservation does not imply that the components ω_j are constants because $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ also change with time according to

$$\frac{d}{dt} \mathbf{i}_j = \boldsymbol{\omega} \wedge \mathbf{i}_j, \quad j = 1, 2, 3$$

Hence, $\dot{\mathbf{M}} = 0$ becomes, by the first of [44] and denoting $\mathbf{I}\boldsymbol{\omega} = (I_1 \omega_1, I_2 \omega_2, I_3 \omega_3)$, the *Euler equations* $\mathbf{I}\dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \wedge \mathbf{I}\boldsymbol{\omega} = 0$, or

$$\begin{aligned} I_1 \dot{\omega}_1 &= (I_2 - I_3) \omega_2 \omega_3 \\ I_2 \dot{\omega}_2 &= (I_3 - I_1) \omega_3 \omega_1 \\ I_3 \dot{\omega}_3 &= (I_1 - I_2) \omega_1 \omega_2 \end{aligned} \quad [46]$$

which can be considered together with the conserved quantities [44].

Since angular momentum is conserved, it is convenient to introduce the *laboratory frame* $(O; \mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0)$ with fixed axes $\mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0$ and (see **Figure 4**):

1. $(O; \mathbf{x}, \mathbf{y}, \mathbf{z})$, the *momentum frame* with fixed axes, but with \mathbf{z} -axis oriented as \mathbf{M} , and \mathbf{x} -axis coinciding with the node (i.e., the intersection) of the $\mathbf{x}_0\text{-}\mathbf{y}_0$ plane and the $\mathbf{x}\text{-}\mathbf{y}$ plane (orthogonal to \mathbf{M}). Therefore, $\mathbf{x}, \mathbf{y}, \mathbf{z}$ is determined by the two Euler angles ζ, γ of $(O; \mathbf{x}, \mathbf{y}, \mathbf{z})$ in $(O; \mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0)$;
2. $(O; \mathbf{1}, \mathbf{2}, \mathbf{3})$, the *comoving frame*, that is, the frame fixed with the body, and with unit vectors $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ parallel to the principal axes of the body. The frame is determined by three Euler angles $\theta_0, \varphi_0, \psi_0$;
3. the Euler angles of $(O; \mathbf{1}, \mathbf{2}, \mathbf{3})$ with respect to $(O; \mathbf{x}, \mathbf{y}, \mathbf{z})$, which are denoted θ, φ, ψ ;
4. G , the total angular momentum: $G^2 = \sum_i I_i^2 \omega_i^2$;
5. M_3 , the angular momentum along the \mathbf{z}_0 axis; $M_3 = G \cos \zeta$; and
6. L , the projection of \mathbf{M} on the axis **3**, $L = G \cos \theta$.

The quantities $G, M_3, L, \varphi, \gamma, \psi$ determine $\theta_0, \varphi_0, \psi_0$ and $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$, or the $p_{\theta_0}, p_{\varphi_0}, p_{\psi_0}$ variables conjugated to $\theta_0, \varphi_0, \psi_0$ as shown by the following comment.

Considering **Figure 4**, the angles ζ, γ determine location, in the fixed frame $(O; \mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0)$ of the direction of \mathbf{M} and the node line \mathbf{m} , which are, respectively, the \mathbf{z} -axis and the \mathbf{x} -axis of the fixed frame associated with the angular momentum; the angles θ, φ, ψ then determine the position of the comoving frame with respect to the fixed frame $(O; \mathbf{x}, \mathbf{y}, \mathbf{z})$, hence its position with respect to $(O; \mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0)$, that is, $(\theta_0, \varphi_0, \psi_0)$. From this and G , it is possible to determine $\boldsymbol{\omega}$ because

$$\begin{aligned} \cos \theta &= \frac{I_3 \omega_3}{G}, \quad \tan \psi = \frac{I_2 \omega_2}{I_1 \omega_1} \\ \omega_2^2 &= I_2^{-2} (G^2 - I_1^2 \omega_1^2 - I_3^2 \omega_3^2) \end{aligned} \quad [47]$$

and, from [43], $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$ are determined.

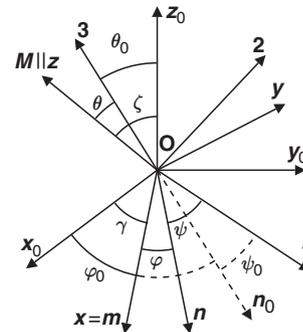


Figure 4 The laboratory frame, the angular momentum frame, and the comoving frame (and the Deprit angles).

The Lagrangian [45] gives immediately (after expressing $\boldsymbol{\omega}$, i.e., $\mathbf{n}, \mathbf{z}, \mathbf{i}_3$, in terms of the Euler angles $\theta_0, \varphi_0, \psi_0$) an expression for the variables $p_{\theta_0}, p_{\varphi_0}, p_{\psi_0}$ conjugated to $\theta_0, \varphi_0, \psi_0$:

$$p_{\theta_0} = \mathbf{M} \cdot \mathbf{n}_0, \quad p_{\varphi_0} = \mathbf{M} \cdot \mathbf{z}_0, \quad p_{\psi_0} = \mathbf{M} \cdot \mathbf{i}_3 \quad [48]$$

and, in principle, we could proceed to compute the Hamiltonian.

However, the computation can be avoided because of the very remarkable property (DEPRIT), which can be checked with some patience, making use of [48] and of elementary spherical trigonometry identities,

$$M_3 d\gamma + G d\varphi + L d\psi \\ = p_{\varphi_0} d\varphi_0 + p_{\psi_0} d\psi_0 + p_{\theta_0} d\theta_0 \quad [49]$$

which means that the map $((M_3, \gamma), (L, \psi), (G, \varphi)) \longleftrightarrow ((p_{\theta_0}, \theta_0), (p_{\varphi_0}, \varphi_0), (p_{\psi_0}, \psi_0))$ is a canonical map. And in the new coordinates, the kinetic energy, hence the Hamiltonian, takes the form

$$K = \frac{1}{2} \left[\frac{L^2}{I_3} + (G^2 - L^2) \left(\frac{\sin^2 \psi}{I_1} + \frac{\cos^2 \psi}{I_2} \right) \right] \quad [50]$$

This again shows that G, M_3 are constants of motion, and the L, ψ variables are determined by a quadrature, because the Hamilton equation for ψ combined with the energy conservation yields

$$\dot{\psi} = \pm \left(\frac{1}{I_3} - \frac{\sin^2 \psi}{I_1} - \frac{\cos^2 \psi}{I_2} \right) \\ \times \sqrt{\frac{2E - G^2 \left(\frac{\sin^2 \psi}{I_1} + \frac{\cos^2 \psi}{I_2} \right)}{\frac{1}{I_3} - \frac{\sin^2 \psi}{I_1} - \frac{\cos^2 \psi}{I_2}}} \quad [51]$$

In the integrability region, this motion is periodic with some period $T_L(E, G)$. Once $\psi(t)$ is determined, the Hamilton equation for φ leads to the further quadrature

$$\dot{\varphi} = \left(\frac{\sin^2 \psi(t)}{I_1} + \frac{\cos^2 \psi(t)}{I_2} \right) G \quad [52]$$

which determines a second periodic motion with period $T_G(E, G)$. The γ, M_3 are constants and, therefore, the motion takes place on three-dimensional invariant tori \mathcal{T}_{E, G, M_3} in phase space, each of which is “always” foliated into two-dimensional invariant tori parametrized by the angle γ which is constant (by [50], because K is M_3 -independent): the latter are in turn foliated by one-dimensional invariant tori, that is, by periodic orbits, with E, G such that the value of $T_L(E, G)/T_G(E, G)$ is rational.

Note that if $I_1 = I_2 = I$, the above analysis is extremely simplified. Furthermore, if gravity g acts on the system the Hamiltonian will simply change by the addition of a potential $-mgz$ if z is the height of the center of mass. Then (see Figure 4), if the center of mass of the body is on the axis \mathbf{i}_3 and $z = h \cos \theta_0$, and h is the distance of the center of mass from O , since $\cos \theta_0 = \cos \theta \cos \zeta - \sin \theta \sin \zeta \cos \varphi$, the Hamiltonian will become $\mathcal{H} = K - mgh \cos \theta_0$ or

$$\mathcal{H} = \frac{G^2}{2I_3} + \frac{G^2 - L^2}{2I} - mgh \left(\frac{M_3 L}{G^2} - \left(1 - \frac{M_3^2}{G^2} \right)^{1/2} \right) \\ \times \left(1 - \frac{L^2}{G^2} \right)^{1/2} \cos \varphi \quad [53]$$

so that, again, the system is integrable by quadratures (with the roles of ψ and φ “interchanged” with respect to the previous case) in suitable regions of phase space. This is called the *Lagrange’s gyroscope*.

A less elementary integrable case is when the inertia moments are related as $I_1 = I_2 = 2I_3$ and the center of mass is in the \mathbf{i}_1 – \mathbf{i}_2 plane (rather than on the \mathbf{i}_3 -axis) and only gravity acts, besides the constraint force on the pivot point O ; this is called *Kowalevskaia’s gyroscope*.

For more details, see Gallavotti (1983).

Other Quadratures

An interesting classical integrable motion is that of a point mass attracted by two equal-mass centers of gravitational attraction, or a point ideally constrained to move on the surface of a general ellipsoid.

New integrable systems have been discovered quite recently and have generated a wealth of new developments ranging from group theory (as integrable systems are closely related to symmetries) to partial differential equations.

It is convenient to extend the notion of integrability by stating that a system is integrable in a region W of phase space if

1. there is a change of coordinates $(\mathbf{p}, \mathbf{q}) \in W \longleftrightarrow \{\mathbf{A}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{y}\} \in (U \times \mathbb{T}^\ell) \times (V \times \mathbb{R}^m)$ where $U \subset \mathbb{R}^\ell, V \subset \mathbb{R}^m$, with $\ell + m \geq 1$, are open sets; and
2. the \mathbf{A}, \mathbf{Y} are constants of motion while the other coordinates vary “linearly”:

$$(\boldsymbol{\alpha}, \mathbf{y}) \rightarrow (\boldsymbol{\alpha} + \boldsymbol{\omega}(\mathbf{A}, \mathbf{Y})t, \mathbf{y} + \boldsymbol{\nu}(\mathbf{A}, \mathbf{Y})t) \quad [54]$$

where $\boldsymbol{\omega}(\mathbf{A}, \mathbf{Y}), \boldsymbol{\nu}(\mathbf{A}, \mathbf{Y})$ are smooth functions.

In the new sense, the systems studied in the previous sections are integrable in much wider regions (essentially on the entire phase space with the exception of a set of data which lie on lower-dimensional surfaces

forming sets of zero volume). The notion is convenient also because it allows us to say that even the systems of free particles are integrable.

Two very remarkable systems integrable in the new sense are the Hamiltonian systems, respectively called *Toda lattice* (KRUSKAL, ZABUSKY), and *Calogero lattice* (CALOGERO, MOSER); if $(p_i, q_i) \in \mathbb{R}^2$, they are

$$\begin{aligned}\mathcal{H}_T(\mathbf{p}, \mathbf{q}) &= \frac{1}{2m} \sum_{i=1}^n p_i^2 + \sum_{i=1}^{n-1} g e^{-\kappa(q_{i+1}-q_i)} \\ \mathcal{H}_C(\mathbf{p}, \mathbf{q}) &= \frac{1}{2m} \sum_{i=1}^n p_i^2 + \sum_{i < j}^n \frac{g}{(q_i - q_j)^2} \\ &\quad + \frac{1}{2} \sum_{i=1}^n m\omega^2 q_i^2\end{aligned}\quad [55]$$

where $m > 0$ and $\kappa, \omega, g \geq 0$. They describe the motion of n interacting particles on a line.

The integration method for the above systems is again to find first the constants of motion and later to look for quadratures, when appropriate. The constants of motion can be found with the method of the *Lax pairs*. One shows that there is a pair of self-adjoint $n \times n$ matrices $M(\mathbf{p}, \mathbf{q}), N(\mathbf{p}, \mathbf{q})$ such that the equations of motion become

$$\frac{d}{dt} M(\mathbf{p}, \mathbf{q}) = i[M(\mathbf{p}, \mathbf{q}), N(\mathbf{p}, \mathbf{q})], \quad i = \sqrt{-1} \quad [56]$$

which imply that $M(t) = U(t)M(0)U(t)^{-1}$, with $U(t)$ a unitary matrix. When the equations can be written in the above form, it is clear that the n eigenvalues of the matrix $M(0) = M(\mathbf{p}_0, \mathbf{q}_0)$ are constants of motion. When appropriate (e.g., in the Calogero lattice case with $\omega > 0$), it is possible to proceed to find canonical action-angle coordinates: a task that is quite difficult due to the arbitrariness of n , but which is possible.

The Lax pairs for the Calogero lattice (with $\omega = 0, g = m = 1$) are

$$\begin{aligned}M_{bb} &= p_b, & N_{bb} &= 0 \\ M_{bk} &= \frac{i}{(q_b - q_k)}, & N_{bk} &= \frac{1}{(q_b - q_k)^2} b \neq k\end{aligned}\quad [57]$$

while for the Toda lattice (with $m = g = \frac{1}{2}\kappa = 1$) the nonzero matrix elements of M, N are

$$\begin{aligned}M_{bb} &= p_b, & M_{b, b+1} &= M_{b+1, b} = e^{-(q_b - q_{b+1})} \\ N_{b, b+1} &= -N_{b+1, b} = i e^{-(q_b - q_{b+1})}\end{aligned}\quad [58]$$

which are checked by first trying the case $n = 2$.

Another integrable system (SUTHERLAND) is

$$\mathcal{H}_S(\mathbf{p}, \mathbf{q}) = \frac{1}{2m} \sum_{i=k}^n p_k^2 + \sum_{b < k}^n \frac{g}{\sinh^2(q_b - q_k)} \quad [59]$$

whose Lax pair is related to that of the Calogero lattice.

By taking suitable limits as $n \rightarrow \infty$ and as the other parameters tend to 0 or ∞ at suitable rates, integrability of a few differential equations, among which the *Korteweg-deVries equation* or the non-linear *Schrödinger equation*, can be derived.

As mentioned in the introductory section, symmetry properties under continuous groups imply existence of constants of motion. Hence, it is natural to think that integrability of a mechanical system reflects enough symmetry to imply the existence of as many constants of motion, independent and in involution, as the number of degrees of freedom, n .

This is in fact always true, and in some respects it is a tautological statement in the anisochronous cases. Integrability in a region W implies existence of canonical action-angle coordinates (A, α) (see the section “[Quasiperiodicity and integrability](#)”) and the Hamiltonian depends solely on the A ’s: therefore, its restriction to W is invariant with respect to the action of the continuous commutative group T^n of the translations of the angle variables. The actions can be seen as constants of motion whose existence follows from Noether’s theorem, at least in the anisochronous cases in which the Hamiltonian formulation is equivalent to a Lagrangian one.

What is nontrivial is to recognize, prior to realizing integrability, that a system admits this kind of symmetry: in most of the interesting cases, the systems either do not exhibit obvious symmetries or they exhibit symmetries apparently unrelated to the group T^n , which nevertheless imply existence of sufficiently many independent constants of motion as required for integrability. Hence, nontrivial integrable systems possess a “hidden” symmetry under T^n : the rigid body is an example.

However, very often the symmetries of a Hamiltonian H which imply integrability also imply partial isochrony, that is, they imply that the number of independent frequencies is smaller than n (see the section “[Quasiperiodicity and integrability](#)”). Even in such cases, often a map exists from the original coordinates (\mathbf{p}, \mathbf{q}) to the integrating variables (A, α) in which A are constants of motion and the α are uniformly rotating angles (some of which are also constant) with spectrum $\omega(A)$, which is the gradient $\partial_A h(A)$ for some function $h(A)$ depending only on a few of the A coordinates. However, the map might fail to be canonical. The system is then said to be bi-Hamiltonian: in the sense that one can represent motions in two systems of canonical coordinates, not related by a canonical transformation, and by two Hamiltonian functions H and $H' \equiv h$ which generate the same motions in the respective

coordinates (the latter changes of variables are sometimes called “canonical with respect to the pair H, H' ” while the transformations considered in the section “Canonical transformations of phase space coordination” are called completely canonical).

For more details, we refer the reader to [Calogero and Degasperis \(1982\)](#).

Generic Nonintegrability

It is natural to try to prove that a system “close” to an integrable one has motions with properties very close to quasiperiodic. This is indeed the case, but in a rather subtle way. That there is a problem is easily seen in the case of a perturbation of an anisochronous integrable system.

Assume that a system is integrable in a region W of phase space which, in the integrating action–angle variables (A, α) , has the standard form $U \times \mathbb{T}^\ell$ with a Hamiltonian $b(A)$ with gradient $\omega(A) = \partial_A b(A)$. If the forces are perturbed by a potential which is smooth then the new system will be described, in the same coordinates, by a Hamiltonian like

$$\mathcal{H}_\varepsilon(A, \alpha) = b(A) + \varepsilon f(A, \alpha) \quad [60]$$

with b, f analytic in the variables A, α .

If the system really behaved like the unperturbed one, it ought to have ℓ constants of motion of the form $F_\varepsilon(A, \alpha)$ analytic in ε near $\varepsilon=0$ and uniform, that is, single valued (which is the same as periodic) in the variables α . However, the following theorem (POINCARÉ) shows that this is a somewhat unlikely possibility.

Theorem 1 *If the matrix $\partial_{AA}^2 b(A)$ has rank ≥ 2 , the Hamiltonian [60] “generically” (an intuitive notion precised below) cannot be integrated by a canonical transformation $C_\varepsilon(A, \alpha)$ which*

- (i) *reduces to the identity as $\varepsilon \rightarrow 0$; and*
- (ii) *is analytic in ε near $\varepsilon=0$ and in $(A, \alpha) \in U' \times \mathbb{T}^\ell$, with $U' \subset U$ open.*

Furthermore, no uniform constants of motion $F_\varepsilon(A, \alpha)$, defined for ε near 0 and (A, α) in an open domain $U' \times \mathbb{T}^\ell$, exist other than the functions of \mathcal{H}_ε itself.

Integrability in the sense (i), (ii) can be called analytic integrability and it is the strongest (and most naive) sense that can be given to the attribute.

The first part of the theorem, that is, (i), (ii), holds simply because, if integrability was assumed, a generating function of the integrating map would have the form $A' \cdot \alpha + \Phi_\varepsilon(A', \alpha)$ with Φ admitting a

power series expansion in ε as $\Phi_\varepsilon = \varepsilon \Phi^1 + \varepsilon^2 \Phi^2 + \dots$. Hence, Φ^1 would have to satisfy

$$\omega(A') \cdot \partial_\alpha \Phi^1(A', \alpha) + f(A', \alpha) = \bar{f}(A') \quad [61]$$

where $\bar{f}(A')$ depends only on A' (hence integrating both sides with respect to α , it appears that $\bar{f}(A')$ must coincide with the average of $f(A', \alpha)$ over α).

This implies that the Fourier transform $f_\nu(A)$, $\nu \in \mathbb{Z}^\ell$, should satisfy

$$f_\nu(A') = 0 \quad \text{if } \omega(A') \cdot \nu = 0, \quad \nu \neq 0 \quad [62]$$

which is equivalent to the existence of $\tilde{f}_\nu(A')$ such that $f_\nu(A) = \omega(A') \cdot \nu \tilde{f}_\nu(A)$ for $\nu \neq 0$. But since there is no relation between $\omega(A)$ and $f(A, \alpha)$, this property “generically” will not hold in the sense that as close as wished to an f which satisfies the property [62] there will be another f which does not satisfy it essentially no matter how “closeness” is defined, (e.g., with respect to the metric $\|f - g\| = \sum_\nu |f_\nu(A) - g_\nu(A)|$). This is so because the rank of $\partial_{AA}^2 b(A)$ is higher than 1 and $\omega(A)$ varies at least on a two-dimensional surface, so that $\omega \cdot \nu = 0$ becomes certainly possible for some $\nu \neq 0$ while $f_\nu(A)$ in general will not vanish, so that Φ^1 , hence Φ_ε , does not exist.

This means that close to a function f there is a function f' which violates [62] for some ν . Of course, this depends on what is meant by “close”: however, here essentially any topology introduced on the space of the functions f will make the statement correct. For instance, if the distance between two functions is defined by $\sum_\nu \sup_{A \in U} |f_\nu(A) - g_\nu(A)|$ or by $\sup_{A, \alpha} |f(A, \alpha) - g(A, \alpha)|$.

The idea behind the last statement of the theorem is in essence the same: consider, for simplicity, the anisochronous case in which the matrix $\partial_{AA}^2 b(A)$ has maximal rank ℓ , that is, the determinant $\det \partial_{AA}^2 b(A)$ does not vanish. Anisochrony implies that $\omega(A) \cdot \nu \neq 0$ for all $\nu \neq 0$ and A on a dense set, and this property will be used repeatedly in the following analysis.

Let $B(\varepsilon, A, \alpha)$ be a “uniform” constant of motion, meaning that it is single valued and analytic in the non-simply-connected region $U \times \mathbb{T}^\ell$ and, for ε small,

$$B(\varepsilon, A, \alpha) = B_0(A, \alpha) + \varepsilon B_1(A, \alpha) + \varepsilon^2 B_2(A, \alpha) + \dots \quad [63]$$

The condition that B is a constant of motion can be written order by order in its expansion in ε : the first two orders are

$$\begin{aligned} \omega(A) \cdot \partial_\alpha B_0(A, \alpha) &= 0 \\ \partial_A f(A, \alpha) \cdot \partial_\alpha B_0(A, \alpha) - \partial_\alpha f(A, \alpha) \cdot \partial_A B_0(A, \alpha) &+ \omega(A) \cdot \partial_\alpha B_1(A, \alpha) = 0 \end{aligned} \quad [64]$$

Then the above two relations and anisochrony imply (1) that B_0 must be a function of A only and (2) that $\boldsymbol{\omega}(A) \cdot \mathbf{v}$ and $\partial_A B_0(A) \cdot \mathbf{v}$ vanish simultaneously for all \mathbf{v} . Hence, the gradient of B_0 must be proportional to $\boldsymbol{\omega}(A)$, that is, to the gradient of $h(A) : \partial_A B_0(A) = \lambda(A) \partial_A h(A)$. Therefore, generically (because of the anisochrony) it must be that B_0 depends on A through $h(A) : B_0(A) = F(h(A))$ for some F .

Looking again, with the new information, at the second of [64] it follows that at fixed A the $\boldsymbol{\alpha}$ -derivative in the direction $\boldsymbol{\omega}(A)$ of B_1 equals $F'(h(A))$ times the $\boldsymbol{\alpha}$ -derivative of f , that is, $B_1(A, \boldsymbol{\alpha}) = f(A, \boldsymbol{\alpha}) F'(h(A)) + C_1(A)$.

Summarizing: the constant of motion B has been written as $B(A, \boldsymbol{\alpha}) = F(h(A)) + \varepsilon F'(h(A)) f(A, \boldsymbol{\alpha}) + \varepsilon C_1(A) + \varepsilon^2 B_2 + \dots$ which is equivalent to $B(A, \boldsymbol{\alpha}) = F(\mathcal{H}_\varepsilon) + \varepsilon(B'_0 + \varepsilon B'_1 + \dots)$ and therefore $B'_0 + \varepsilon B'_1 + \dots$ is another analytic constant of motion. Repeating the argument also $B'_0 + \varepsilon B'_1 + \dots$ must have the form $F_1(\mathcal{H}_\varepsilon) + \varepsilon(B''_0 + \varepsilon B''_1 + \dots)$; conclusion

$$B = F(\mathcal{H}_\varepsilon) + \varepsilon F_1(\mathcal{H}_\varepsilon) + \varepsilon^2 F_2(\mathcal{H}_\varepsilon) + \dots + \varepsilon^n F_n(\mathcal{H}_\varepsilon) + O(\varepsilon^{n+1}) \quad [65]$$

By analyticity, $B = F_\varepsilon(\mathcal{H}_\varepsilon(A, \boldsymbol{\alpha}))$ for some F_ε : hence generically all constants of motion are trivial.

Therefore, a system close to integrable cannot behave as it would naively be expected. The problem, however, was not manifest until POINCARÉ's proof of the above results: because in most applications the function f has only finitely many Fourier components, or at least is replaced by an approximation with this property, so that at least [62] and even a few of the higher-order constraints like [64] become possible in open regions of action space. In fact, it may happen that the values of A of interest are restricted so that $\boldsymbol{\omega}(A) \cdot \mathbf{v} = 0$ only for "large" values of \mathbf{v} for which $f_{\mathbf{v}} = 0$. Nevertheless, the property that $f_{\mathbf{v}}(A) = (\boldsymbol{\omega}(A) \cdot \mathbf{v}) \tilde{f}_{\mathbf{v}}(A)$ (or the analogous higher-order conditions, e.g., [64]), which we have seen to be necessary for analytic integrability of the perturbed system, can be checked to fail in important problems, if no approximation is made on f . Hence a conceptual problem arises.

For more details see Poincaré (1987).

Perturbing Functions

To check, in a given problem, the nonexistence of nontrivial constants of motion along the lines indicated in the previous section, it is necessary to express the potential, usually given in Cartesian

coordinates as $\varepsilon V(\mathbf{x})$, in terms of the action–angle variables of the unperturbed, integrable, system.

In particular, the problem arises when trying to check nonexistence of nontrivial constants of motion when the anisochrony assumption (cf. the previous section) is not satisfied. Usually it becomes satisfied "to second order" (or higher): but to show this, a more detailed information on the structure of the perturbing function expressed in action–angle variables is needed. For instance, this is often necessary even when the perturbation is approximated by a trigonometric polynomial, as it is essentially always the case in celestial mechanics.

Finding explicit expressions for the action–angle variables is in itself a rather nontrivial task which leads to many problems of intrinsic interest even in seemingly simple cases. For instance, in the case of the planar gravitational central motion, the Kepler equation $\lambda = \xi - \varepsilon \sin \xi$ (see the first of [41]) must be solved expressing ξ in terms of λ (see the first of [42]). It is obvious that for small ε , the variable ξ can be expressed as an analytic function of ε : nevertheless, the actual construction of this expression leads to several problems. For small ε , an interesting algorithm is the following.

Let $h(\lambda) = \xi - \lambda$, so that the equation to solve (i.e., the first of [41]) is

$$h(\lambda) = \varepsilon \sin(\lambda + h(\lambda)) \equiv -\varepsilon \frac{\partial c}{\partial \lambda}(\lambda + h(\lambda)) \quad [66]$$

where $c(\lambda) = \cos \lambda$; the function $\lambda \rightarrow h(\lambda)$ should be periodic in λ , with period 2π , and analytic in ε, λ for ε small and λ real. If $h(\lambda) = \varepsilon h^{(1)} + \varepsilon^2 h^{(2)} + \dots$, the Fourier transform of $h^{(k)}(\lambda)$ satisfies the recursion relation

$$h^{(k)}_{\nu} = - \sum_{p=1}^{\infty} \frac{1}{p!} \sum_{\substack{k_1 + \dots + k_p = k-1 \\ \nu_0 + \nu_1 + \dots + \nu_p = \nu}} (i\nu_0) c_{\nu_0} (i\nu_0)^p \times \prod h^{(k_j)}_{\nu_j}, \quad k > 1 \quad [67]$$

with c_{ν} the Fourier transform of the cosine ($c_{\pm 1} = \frac{1}{2}$, $c_{\nu} = 0$ if $\nu \neq \pm 1$), and (of course) $h^{(1)}_{\nu} = -i\nu c_{\nu}$. Equation [67] is obtained by expanding the RHS of [66] in powers of h and then taking the Fourier transform of both sides retaining only terms of order k in ε .

Iterating the above relation, imagine drawing all trees θ with k "branches," or "lines," distinguished by a label taking k values, and k nodes and attach to each node ν a harmonic label $\nu_{\nu} = \pm 1$ as in Figure 5. The trees will be assumed to start with a root line νr linking a point r and the "first node" ν (see Figure 5)

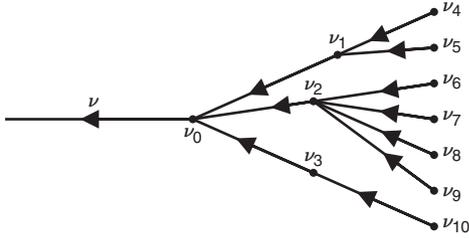


Figure 5 An example of a tree graph and its labels. It contains only one simple node (3). Harmonics are indicated next to their nodes. Labels distinguishing lines are not marked.

and then bifurcate arbitrarily (such trees are sometimes called “rooted trees”).

Imagine the tree oriented from the endpoints towards the root r (not to be considered a node) and given a node ν call ν' the node immediately following it. If ν is the first node before the root r , let $\nu' = r$ and $\nu_{\nu'} = 1$. For each such decorated tree define its numerical value

$$\text{Val}(\theta) = \frac{-i}{k!} \prod_{\text{lines } l=\nu'v} (\nu_{\nu'} \nu_{\nu}) \prod_{\text{nodes}} c_{\nu_{\nu}} \quad [68]$$

and define a current $\nu(l)$ on a line $l = \nu'v$ to be the sum of the harmonics of the nodes preceding ν' : $\nu(l) = \sum_{w \leq \nu} \nu_w$. Call $\nu(\theta)$ the current flowing in the root branch and call order of θ the number of nodes (or branches). Then

$$h_{\nu}^{(k)} = \sum_{\substack{\theta, \nu(\theta)=\nu \\ \text{order}(\theta)=k}} \text{Val}(\theta) \quad [69]$$

provided trees are considered identical if they can be overlapped (labels included) after suitably scaling the lengths of their branches and pivoting them around the nodes out of which they emerge (the root is always imagined to be fixed at the origin).

If the trees are stripped of the harmonic labels, their number is finite and it can be estimated to be $\leq k!4^k$ (because the labels which distinguish the lines can be attached to an unlabeled tree in many ways). The harmonic labels (i.e., $\nu_{\nu} = \pm 1$) can be laid down in 2^k ways, and the value of each tree can be bounded by $(1/k!)2^{-k}$ (because $c_{\pm 1} = \frac{1}{2}$).

Hence $\sum_{\nu} |h_{\nu}^{(k)}| \leq 4^k$, which gives a (rough) estimate of the radius of convergence of the expansion of h in powers of ε : namely 0.25 (easily improvable to 0.3678 if $4^k k!$ is replaced by k^{k-1} using Cayley’s formula for the enumeration of rooted trees). A simple expression for $h^{(k)}(\psi)$ (LAGRANGE) is

$$h^{(k)}(\psi) = \frac{1}{k!} \partial_{\psi}^{k-1} \sin^k \psi$$

(also readable from the tree representation): the actual radius of convergence, first determined by Laplace, of the series for h can also be determined from the latter expression for h (ROUCHÉ) or directly from the tree representation: it is ~ 0.6627 .

One can find better estimates or at least more efficient methods for evaluating the sums in [69]: in fact, in performing the sum in [69] important cancellations occur. For instance, the harmonic labels can be subject to the further strong constraint that no line carries zero current because the sum of the values of the trees of fixed order and with at least one line carrying zero current vanishes.

The above expansion can also be simplified by *partial resummations*. For the purpose of an example, let the nodes with one entering and one exiting line (see Figure 5) be called as “simple” nodes. Then all tree graphs which, on any line between two nonsimple nodes, contain any number of simple nodes can be eliminated. This is done by replacing, in evaluating the (remaining) tree values, the factors $\nu_{\nu'} \nu_{\nu}$ in [68] by $\nu_{\nu'} \nu_{\nu} / (1 - \varepsilon \cos \psi)$: then the value of θ (denoted $\text{Val}(\theta)_{\psi}$) for a tree becomes a function of ψ and ε and [69] is replaced by

$$h(\psi) = \sum_{k=1}^{\infty} \sum_{\substack{\theta, \nu(\theta)=\nu \\ \text{order}(\theta)=k}}^* \varepsilon^k e^{i\nu\psi} \text{Val}(\theta)_{\psi} \quad [70]$$

where the $*$ means that the trees are subject to the further restriction of not containing any simple node. It should be noted that the above graphical representation of the solution of the Kepler equation is strongly reminiscent of the representations of quantities in terms of graphs that occur often in quantum field theory. Here the trees correspond to *Feynman graphs*, the factors associated with the nodes are the *couplings*, the factors associated with the lines are the *propagators*, and the resummations are analogous to the *self-energy resummations*, while the cancellations mentioned above can be related to the class of identities called *Ward identities*. Not only the analogy can be shown not to be superficial, but it also turns out to be very helpful in key mechanical problems: see Appendix 1.

The existence of a vast number of identities relating the tree values is shown already by the *simple* form of the Lagrange series and by the even more remarkable resummation (LEVI-CIVITA) leading to

$$h(\psi) = \sum_{k=1}^{\infty} \frac{(\varepsilon \sin \psi)^k}{k!} \left(\frac{1}{1 - \varepsilon \cos \psi} \partial_{\psi} \right)^k \psi \quad [71]$$

It is even possible to further collect the series terms to express it as a series with much better convergence properties; for instance, its terms can be reorganized and collected (resummed) so that h is expressed as a power series in the parameter

$$\eta = \frac{\varepsilon e^{\sqrt{1-\varepsilon^2}}}{1 + \sqrt{1-\varepsilon^2}} \quad [72]$$

with radius of convergence 1, which corresponds to $\varepsilon = 1$ (via a simple argument by Levi-Civita). The analyticity domain for the Lagrange series is $|\eta| < 1$. This also determines the value of Laplace radius, which is the point closest to the origin of the complex curve $|\eta(\varepsilon)| = 1$: it is imaginary so that it is the root of the equation

$$\varepsilon e^{\sqrt{1+\varepsilon^2}} / (1 + \sqrt{1+\varepsilon^2}) = 1$$

The analysis provides an example, in a simple case of great interest in applications, of the kind of computations actually necessary to represent the perturbing function in terms of action–angle variables. The property that the function $c(\lambda)$ in [66] is the cosine has been used only to limit the range of the label ν to be ± 1 ; hence the same method, with similar results, can be applied to study the inversion of the relation between the average anomaly λ and the true anomaly θ and to efficiently obtain, for instance, the properties of f, g in [42].

For more details, the reader is referred to [Levi-Civita \(1956\)](#).

Lindstedt and Birkhoff Series: Divergences

Nonexistence of constants of motion, rather than being the end of the attempts to study motions close to integrable ones by perturbation methods, marks the beginning of renewed efforts to understand their nature.

Let $(A, \alpha) \in U \times \mathbb{T}^\ell$ be action–angle variables defined in the integrability region for an analytic Hamiltonian and let $h(A)$ be its value in the action–angle coordinates. Suppose that $h(A_0)$ is anisochronous and let $f(A, \alpha)$ be an analytic perturbing function. Consider, for ε small, the Hamiltonian $\mathcal{H}_\varepsilon(A, \alpha) = \mathcal{H}_0(A) + \varepsilon f(A, \alpha)$.

Let $\omega_0 = \omega(A_0) \equiv \partial_A \mathcal{H}_0(A)$ be the frequency spectrum (see the section “[Quasiperiodicity and integrability](#)”) of one of the invariant tori of the unperturbed system corresponding to an action A_0 . Short of integrability, the question to ask at this point is whether the perturbed system admits an

analytic invariant torus on which the motion is quasiperiodic and

1. has the same spectrum ω_0 ,
2. depends analytically on ε at least for ε small,
3. reduces to the “unperturbed torus” $\{A_0\} \times \mathbb{T}^\ell$ as $\varepsilon \rightarrow 0$.

More concretely, the question is:

Are there functions $H_\varepsilon(\psi), h_\varepsilon(\psi)$ analytic in $\psi \in \mathbb{T}^\ell$ and in ε near 0, vanishing as $\varepsilon \rightarrow 0$ and such that the torus with parametric equations

$$A = A_0 + H_\varepsilon(\psi), \quad \alpha = \psi + h_\varepsilon(\psi), \quad \psi \in \mathbb{T}^\ell \quad [73]$$

is invariant and, if $\omega_0 \stackrel{\text{def}}{=} \omega(A_0)$, the motion on it is simply $\psi \rightarrow \psi + \omega_0 t$, i.e., it is quasiperiodic with spectrum ω_0 ?

In this context, Poincaré’s theorem (in the section “[Generic nonintegrability](#)”) had followed another key result, earlier developed in particular cases and completed by him, which provides a partial answer to the question.

Suppose that $\omega_0 = \omega(A_0) \in \mathbb{R}^\ell$ satisfies a Diophantine property, namely suppose that there exist constants $C, \tau > 0$ such that

$$|\omega_0 \cdot \nu| \geq \frac{1}{C|\nu|^\tau}, \quad \text{for all } 0 \neq \nu \in \mathbb{Z}^\ell \quad [74]$$

which, for each $\tau > \ell - 1$ fixed, is a property enjoyed by all $\omega \in \mathbb{R}^\ell$ but for a set of zero measure. Then the motions on the unperturbed torus run over trajectories that fill the torus densely because of the “irrationality” of ω_0 implied by [74]. Writing Hamilton’s equations,

$$\dot{\alpha} = \partial_A \mathcal{H}_0(A) + \varepsilon \partial_A f(A, \alpha), \quad \dot{A} = -\varepsilon \partial_\alpha f(A, \alpha)$$

with A, α given by [73] with ψ replaced by $\psi + \omega t$, and using the density of the unperturbed trajectories implied by [74], the condition that [73] are equations for an invariant torus on which the motion is $\psi \rightarrow \psi + \omega_0 t$ are

$$\begin{aligned} \omega_0 + (\omega_0 \cdot \partial_\psi) h_\varepsilon(\psi) &= \partial_A \mathcal{H}_0(A_0 + H_\varepsilon(\psi)) \\ &+ \varepsilon \partial_A f(A_0 + H_\varepsilon(\psi), \psi + h_\varepsilon(\psi)) (\omega_0 \cdot \partial_\psi) H_\varepsilon(\psi) \\ &= -\varepsilon \partial_\alpha f(A_0 + H_\varepsilon(\psi), \psi + h_\varepsilon(\psi)) \end{aligned} \quad [75]$$

The theorem referred to above (POINCARÉ) is that

Theorem 2 *If the unperturbed system is anisochronous and $\omega_0 = \omega(A_0)$ satisfies [74] for some $C, \tau > 0$ there exist two well defined power series $h_\varepsilon(\psi) = \sum_{k=1}^{\infty} \varepsilon^k h^{(k)}(\psi)$ and $H_\varepsilon(\psi) = \sum_{k=1}^{\infty} \varepsilon^k H^{(k)}(\psi)$ which*

solve [75] to all orders in ε . The series for H_ε is uniquely determined, and such is also the series for h_ε up to the addition of an arbitrary constant at each order, so that it is unique if h_ε is required, as henceforth done with no loss of generality, to have zero average over ψ .

The algorithm for the construction is illustrated in a simple case in the next section (see eqns [83], [84]). Convergence of the above series, called *Lindstedt series*, even for small ε has been a problem for rather a long time. Poincaré proved the existence of the formal solution; but his other result, discussed in the section “**Generic nonintegrability**,” casts doubts on convergence although it does not exclude it, as was immediately stressed by several authors (including Poincaré himself). The result in that section shows the impossibility of solving [75] for all ω_0 's near a given spectrum, analytically and uniformly, but it does not exclude the possibility of solving it for a single ω_0 .

The theorem admits several extensions or analogs: an interesting one is to the case of isochronous unperturbed systems:

Given the Hamiltonian $\mathcal{H}_\varepsilon(A, \alpha) = \omega_0 \cdot A + \varepsilon f(A, \alpha)$, with ω_0 satisfying [74] and f analytic, there exist power series $C_\varepsilon(A', \alpha')$, $u_\varepsilon(A')$ such that $\mathcal{H}_\varepsilon(C_\varepsilon(A', \alpha')) = \omega_0 \cdot A' + u_\varepsilon(A')$ holds as an equality between formal power series (i.e., order by order in ε) and at the same time the C_ε , regarded as a map, satisfies order by order the condition (i.e., (4.3)) that it is a canonical map.

This means that there is a generating function $A' \cdot \alpha + \Phi_\varepsilon(A', \alpha)$ also defined by a formal power series $\Phi_\varepsilon(A', \alpha) = \sum_{k=1}^{\infty} \varepsilon^k \Phi^{(k)}(A', \alpha)$, that is, such that if $C_\varepsilon(A', \alpha') = (A, \alpha)$ then it is true, order by order in powers of ε , that $A = A' + \partial_\alpha \Phi_\varepsilon(A', \alpha)$ and $\alpha' = \alpha + \partial_{A'} \Phi_\varepsilon(A', \alpha)$. The series for $\Phi_\varepsilon, u_\varepsilon$ are called Birkhoff series.

In this isochronous case, if Birkhoff series were convergent for small ε and (A', α) in a region of the form $U \times \mathbb{T}^\ell$, with $U \subset \mathbb{R}^\ell$ open and bounded, it would follow that, for small ε , \mathcal{H}_ε would be integrable in a large region of phase space (i.e., where the generating function can be used to build a canonical map: this would essentially be $U \times \mathbb{T}^\ell$ deprived of a small layer of points near the boundary of U). However, convergence for small ε is false (in general), as shown by the simple two-dimensional example

$$\begin{aligned} \mathcal{H}_\varepsilon(A, \alpha) &= \omega_0 \cdot A + \varepsilon (A_2 + f(\alpha)) \\ (A, \alpha) &\in \mathbb{R}^2 \times \mathbb{T}^2 \end{aligned} \quad [76]$$

with $f(\alpha)$ an arbitrary analytic function with all Fourier coefficients f_ν positive for $\nu \neq 0$ and $f_0 = 0$. In the latter case, the solution is

$$\begin{aligned} u_\varepsilon(A') &= \varepsilon A_2 \\ \Phi_\varepsilon(A', \alpha) &= \sum_{k=1}^{\infty} \varepsilon^k \sum_{0 \neq \nu \in \mathbb{Z}^2} f_\nu e^{i\alpha \cdot \nu} \frac{(i\nu_2)^k}{(i(\omega_{01}\nu_1 + \omega_{02}\nu_2))^{k+1}} \end{aligned} \quad [77]$$

The series does not converge: in fact, its convergence would imply integrability and, consequently, bounded trajectories in phase space: however, the equations of motion for [76] can be easily solved explicitly and in any open region near given initial data there are other data which have unbounded trajectories if $\omega_{01}/(\omega_{02} + \varepsilon)$ is rational.

Nevertheless, even in this elementary case a formal sum of the series yields

$$\begin{aligned} u(A') &= \varepsilon A'_2 \\ \Phi_\varepsilon(A', \alpha) &= \varepsilon \sum_{0 \neq \nu \in \mathbb{Z}^2} \frac{f_\nu e^{i\alpha \cdot \nu}}{i(\omega_{01}\nu_1 + (\omega_{02} + \varepsilon)\nu_2)} \end{aligned} \quad [78]$$

and the series in [78] (no longer a power series in ε) is really convergent if $\omega = (\omega_{01}, \omega_{02} + \varepsilon)$ is a Diophantine vector (by [74], because analyticity implies exponential decay of $|f_\nu|$). Remarkably, for such values of ε the Hamiltonian \mathcal{H}_ε is integrable and it is integrated by the canonical map generated by [78], in spite of the fact that [78] is obtained, from [77], via the nonrigorous sum rule

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z} \quad \text{for } z \neq 1 \quad [79]$$

(applied to cases with $|z| \geq 1$, which are certainly realized for a dense set of ε 's even if ω is Diophantine because the z 's have values $z = \nu_2/\omega_0 \cdot \nu$). In other words, the integration of the equations is elementary and once performed it becomes apparent that, if ω is diophantine, the solutions can be rigorously found from [78]. Note that, for instance, this means that relations like $\sum_{k=0}^{\infty} 2^k = -1$ are really used to obtain [78] from [77].

Another extension of Lindstedt series arises in a perturbation of an anisochronous system when asking the question as to what happens to the unperturbed invariant tori \mathcal{T}_{ω_0} on which the spectrum is resonant, that is, $\omega_0 \cdot \nu = 0$ for some $\nu \neq 0$, $\nu \in \mathbb{Z}^\ell$. The result is that even in such a case there is a formal power series solution showing that at least a few of the (infinitely many) invariant tori into which \mathcal{T}_{ω_0} is in turn foliated in the unperturbed case can be formally continued at $\varepsilon \neq 0$ (see the section “**Resonances and their stability**”).

For more details, we refer the reader to Poincaré (1987).

Quasiperiodicity and KAM Stability

To discuss more advanced results, it is convenient to restrict attention to a special (nontrivial) paradigmatic case

$$\mathcal{H}_\varepsilon(A, \alpha) = \frac{1}{2}A^2 + \varepsilon f(\alpha) \quad [80]$$

In this simple case (called *Thirring model*: representing ℓ particles on a circle interacting via a potential $\varepsilon f(\alpha)$) the equations for the maximal tori [75] reduce to equations for the only functions h_ε :

$$(\omega \cdot \partial_\psi)^2 h_\varepsilon(\psi) = -\varepsilon \partial_\alpha f(\psi + h_\varepsilon(\psi)), \quad \psi \in \mathbb{T}^\ell \quad [81]$$

as the second of [75] simply becomes the definition of H_ε because the RHS does not involve H_ε .

The real problem is therefore whether the formal series considered in the last section converge at least for small ε : and the example [76] on the Birkhoff series shows that sometimes sum rules might be needed in order to give a meaning to the series. In fact, whenever a problem (of physical interest) admits a formal power series solution which is not convergent, or which is such that it is not known whether it is convergent, then one should look for sum rules for it.

The modern theory of perturbations starts with the proof of the convergence for ε small enough of the Lindstedt series (KOLMOGOROV). The general “KAM” result is:

Theorem 3 (KAM) *Consider the Hamiltonian $\mathcal{H}_\varepsilon(A, \alpha) = h(A) + \varepsilon f(A, \alpha)$, defined in $U = V \times \mathbb{T}^\ell$ with $V \subset \mathbb{R}^\ell$ open and bounded and with $f(A, \alpha)$, $h(A)$ analytic in the closure $\bar{V} \times \mathbb{T}^\ell$ where $h(A)$ is also anisochronous; let $\omega_0 \stackrel{\text{def}}{=} \omega(A_0) = \partial_A h(A_0)$ and assume that ω_0 satisfies [74]. Then*

- (i) *there is $\varepsilon_{C,\tau} > 0$ such that the Lindstedt series converges for $|\varepsilon| < \varepsilon_{C,\tau}$;*
- (ii) *its sum yields two function $H_\varepsilon(\psi), h_\varepsilon(\psi)$ on \mathbb{T}^ℓ which parametrize an invariant torus $\mathcal{T}_{C,\tau}(A_0, \varepsilon)$;*
- (iii) *on $\mathcal{T}_{C,\tau}(A_0, \varepsilon)$ the motion is $\psi \rightarrow \psi + \omega_0 t$, see [73]; and*
- (iv) *the set of data in U which belong to invariant tori $\mathcal{T}_{C,\tau}(A_0, \varepsilon)$ with $\omega(A_0)$ satisfying [74] with prefixed C, τ has complement with volume $< \text{const } C^{-a}$ for a suitable $a > 0$ and with area also $< \text{const } C^{-a}$ on each nontrivial surface of constant energy $\mathcal{H}_\varepsilon = E$.*

In other words, for small ε the spectra of most unperturbed quasiperiodic motions can still be found as spectra of perturbed quasiperiodic motions developing on tori which are close to the corresponding unperturbed ones (i.e., with the same spectrum).

This is a stability result: for instance, in systems with two degrees of freedom the invariant tori of dimension two which lie on a given three-dimensional energy surface, will separate the points on the energy surface into the set which is “inside” the torus and the set which is “outside.” Hence, an initial datum starting (say) inside cannot reach the outside. Likewise, a point starting between two tori has to stay in between forever. Further, if the two tori are close, this means that motion will stay very localized in action space, with a trajectory accessing only points close to the tori and coming close to all such points, within a distance of the order of the distance between the confining tori. The case of three or more degrees of freedom is quite different (see sections “Diffusion in phase space” and “The three-body problem”).

In the simple case of the rotators system [80] the equations for the parametric representation of the tori are given by [81]. The latter bear some analogy with the easier problem in [66]: but [81] are ℓ equations instead of one and they are differential equations rather than ordinary equations. Furthermore, the function $f(\alpha)$ which plays here the role of $c(\lambda)$ in [66] has Fourier coefficient f_ν with no restrictions on ν , while the Fourier coefficients c_ν for c in [66] do not vanish only for $\nu = \pm 1$.

The above differences are, to some extent, “minor” and the power series solution to [81] can be constructed by the same algorithm as used in the case of [66]: namely one forms trees as in Figure 5 with the harmonic labels $\nu_\nu \in \mathbb{Z}$ replaced by $\mathbf{v}_\nu \in \mathbb{Z}^\ell$ (still to be thought of as possible harmonic indices in the Fourier expansion of the perturbing function f). All other labels affixed to the trees in the section “Generic nonintegrability” will be the same. In particular, the current flowing on a branch $l = \nu' \nu$ will be defined as the sum of the harmonics of the nodes $w \leq \nu$ preceding ν :

$$\mathbf{v}(l) \stackrel{\text{def}}{=} \sum_{w \leq \nu} \mathbf{v}_w \quad [82]$$

and we call $\mathbf{v}(\theta)$ the current flowing in the root branch.

Here the value $\text{Val}(\theta)$ of a tree has to be defined differently because the equation to be solved ([81]) contains the differential operator $(\omega_0 \cdot \partial_\psi)^2$ which, when Fourier transformed, becomes multiplication of the Fourier component with harmonic \mathbf{v} by $(i\omega \cdot \mathbf{v})^2$.

The variation due to the presence of the operator $(\omega_0 \cdot \partial_\psi)^2$ and the necessity of its inversion in the evaluation of $\mathbf{u} \cdot h_\nu^{(k)}$, that is, of the component of $h_\nu^{(k)}$ along an arbitrary unit vector \mathbf{u} , is nevertheless quite simple: the value of a tree graph θ of order k

(i.e., with k nodes and k branches) has to be defined by (cf. [68])

$$\text{Val}(\theta) \stackrel{\text{def}}{=} \frac{-i(-1)^k}{k!} \left(\prod_{\text{lines } l=r\nu} \frac{\mathbf{v}_{l'} \cdot \mathbf{v}_\nu}{(\boldsymbol{\omega}_0 \cdot \mathbf{v}(l))^2} \right) \times \left(\prod_{\text{nodes } \nu} f_{\nu} \right) \quad [83]$$

where the $\mathbf{v}_{l'}$ appearing in the factor relative to the root line $r\nu$ from the first node ν to the root r (see Figure 5) is interpreted as a unit vector \mathbf{u} (it was interpreted as 1 in the one-dimensional case [66]). Equation [83] makes sense only for trees in which no line carries zero current. Then the component along \mathbf{u} (the harmonic label attached to the root of a tree) of $\mathbf{h}^{(k)}$ is given (see also [69]) by

$$\mathbf{u} \cdot \mathbf{h}_\nu^{(k)} = \sum_{\substack{\theta, \nu(\theta)=\nu \\ \text{order}(\theta)=k}}^* \text{Val}(\theta) \quad [84]$$

where the $*$ means that the sum is only over trees in which a nonzero current $\mathbf{v}(l)$ flows on the lines $l \in \theta$. The quantity $\mathbf{u} \cdot \mathbf{h}_0^{(k)}$ will be defined to be 0 (see the previous section).

In the case of [66] zero-current lines could appear: but the contributions from tree graphs containing at least one zero current line would cancel. In the present case, the statement that the above algorithm actually gives $\mathbf{h}_\nu^{(k)}$ by simply ignoring trees with lines with zero current is nontrivial. It was Poincaré's contribution to the theory of Lindstedt series to show that even in the general case (cf. [75]) the equations for the invariant tori can be solved by a formal power series. Equation [84] is proved by induction on k after checking it for the first few orders.

The algorithm just described leading to [83] can be extended to the case of the general Hamiltonian considered in the KAM theorem.

The convergence proof is more delicate than the (elementary) one for eqn [66]. In fact, the values of trees of order k can give large contributions to $\mathbf{h}_\nu^{(k)}$: because the “new” factors $(\boldsymbol{\omega}_0 \cdot \mathbf{v}(l))^2$, although not zero, can be quite small and their small size can overwhelm the smallness of the factors f_ν and ε . In fact, even if f is a trigonometric polynomial (so that f_ν vanishes identically for $|\nu|$ large enough) the currents flowing in the branches can be very large, of the order of the number k of nodes in the tree; see [82].

This is called the *small-divisors problem*. The key to its solution goes back to a related work (SIEGEL) which shows that

Theorem 4 *Consider the contribution to the sum in [82] from graphs θ in which no pairs of lines*

which lie on the same path to the root carry the same current and, furthermore, the node harmonics are bounded by $|\nu| \leq N$ for some N . Then the number of lines ℓ in θ with divisor $\boldsymbol{\omega}_0 \cdot \mathbf{v}_\ell$ satisfying $2^{-n} < C|\boldsymbol{\omega}_0 \cdot \mathbf{v}_\ell| \leq 2^{-n+1}$ does not exceed $4Nk2^{-n/\tau}$.

Hence, setting

$$F \stackrel{\text{def}}{=} C^2 \max_{|\nu| \leq N} |f_\nu|$$

the corresponding $\text{Val}(\theta)$ can be bounded by

$$\frac{1}{k!} F^k N^{2k} \prod_{n=0}^{\infty} 2^{2n(4Nk2^{-n/\tau})} \stackrel{\text{def}}{=} \frac{1}{k!} B^k \quad [85]$$

$$B = FN^2 2 \sum_n 8n2^{-n/\tau}$$

since the product is convergent. In the case in which f is a trigonometric polynomial of degree N , the above restricted contributions to $\mathbf{u} \cdot \mathbf{h}_\nu^{(k)}$ would generate a convergent series for ε small enough. In fact, the number of trees is bounded (as in the section “*Perturbing functions*”) by $k!4^k(2N+1)^{\ell k}$ so that the series $\sum_\nu |\varepsilon|^k |\mathbf{u} \cdot \mathbf{h}_\nu^{(k)}|$ would converge for small ε (i.e., $|\varepsilon| < (B \cdot 4(2N+1)^\ell)^{-1}$).

Given this comment, the analysis of the “remaining contributions” becomes the real problem, and it requires new ideas because among the excluded trees there are some simple k th order trees whose value alone, if considered separately from the other contributions, would generate a factorially divergent power series in ε .

However, the contributions of all large-valued trees of order k can be shown to cancel: although not exactly (unlike the case of the elementary problem in the section “*Perturbing functions*,” where the cancellation is not necessary for the proof, in spite of its exact occurrence), but enough so that in spite of the existence of exceedingly large values of individual tree graphs their total sum can still be bounded by a constant to the power k so that the power series actually converges for ε small enough. The idea is discussed in Appendix 1.

For more details, the reader is referred to Poincaré (1987), Kolmogorov (1954), Moser (1962), and Arnol'd (1989).

Resonances and their Stability

A quasiperiodic motion with r rationally independent frequencies is called resonant if r is strictly less than the number of degrees of freedom, ℓ . The difference $s = \ell - r$ is the degree of the resonance.

Of particular interest are the cases of a perturbation of an integrable system in which resonant motions take place.

A typical example is the n -body problem which studies the mutual perturbations of the motions of $n - 1$ particles gravitating around a more massive particle. If the particle masses can be considered to be negligible, the system will consist of $n - 1$ central Keplerian motions: it will therefore have $\ell = 3(n - 1)$ degrees of freedom. In general, only one frequency per body occurs in the absence of the perturbations (the period of the Keplerian orbit). Hence, $r \leq n - 1$ and $s \geq 2(n - 1)$ (or in the planar case $s \geq (n - 1)$) with equality holding when the periods are rationally independent.

Another example is the rigid body with a fixed point perturbed by a conservative force: in this case, the unperturbed system has three degrees of freedom but, in general, only two frequencies (see the discussion following [52]).

Furthermore, in the above examples there is the possibility that the independent frequencies assume, for special initial data, values which are rationally related, giving rise to resonances of even higher order (i.e., with smaller values of r).

In an integrable anisochronous system, resonant motions will be dense in phase space because the frequencies $\boldsymbol{\omega}(\mathbf{A})$ will vary as much as the actions and therefore resonances of any order (i.e., any $r < \ell$) will be dense in phase space: in particular, the periodic motions (i.e., the highest-order resonances) will be dense.

Resonances, in integrable systems, can arise in *a priori* stable integrable systems and in *a priori* unstable systems: the former are systems whose Hamiltonian admits canonical action–angle coordinates $(\mathbf{A}, \boldsymbol{\alpha}) \in U \times \mathbb{T}^\ell$ with $U \subset \mathbb{R}^\ell$ open, while the latter are systems whose Hamiltonian has, in suitable local canonical coordinates, the form

$$\mathcal{H}_0(\mathbf{A}) + \sum_{i=1}^{s_1} \frac{1}{2} (p_i^2 - \lambda_i^2 q_i^2) + \sum_{j=1}^{s_2} \frac{1}{2} (\pi_j^2 + \mu_j^2 \kappa_j^2), \quad [86]$$

$\lambda_i, \mu_j > 0$

where $(\mathbf{A}, \boldsymbol{\alpha}) \in U \times \mathbb{T}^\ell$, $U \in \mathbb{R}^r$, $(\mathbf{p}, \mathbf{q}) \in V \subset \mathbb{R}^{2s_1}$, $(\boldsymbol{\pi}, \boldsymbol{\kappa}) \in V' \subset \mathbb{R}^{2s_2}$ with V, V' neighborhoods of the origin and $\ell = r + s_1 + s_2$, $s_i \geq 0$, $s_1 + s_2 > 0$ and $\pm\sqrt{\lambda_j}$, $\pm\sqrt{\mu_j}$ are called Lyapunov coefficients of the resonance. The perturbations considered are supposed to have the form $\varepsilon f(\mathbf{A}, \boldsymbol{\alpha}, \mathbf{p}, \mathbf{q}, \boldsymbol{\pi}, \boldsymbol{\kappa})$. The denomination of *a priori* stable or unstable refers to the properties of the “*a priori* given unperturbed Hamiltonian.” The label “*a priori* unstable” is certainly appropriate if $s_1 > 0$: here also $s_1 = 0$ is allowed for notational convenience implying that the Lyapunov coefficients in *a priori* unstable cases are all of order 1 (whether real λ_j or imaginary $i\sqrt{\mu_j}$). In

other words, the *a priori* stable case, $s_1 = s_2 = 0$ in [86], is the only excluded case. Of course, the stability properties of the motions when a perturbation acts will depend on the perturbation in both cases.

The *a priori* stable systems usually have a great variety of resonances (e.g., in the anisochronous case, resonances of any dimension are dense). The *a priori* unstable systems have (among possible other resonances) some very special r -dimensional resonances occurring when the unstable coordinates (\mathbf{p}, \mathbf{q}) and $(\boldsymbol{\pi}, \boldsymbol{\kappa})$ are zero and the frequencies of the r action–angle coordinates are rationally independent.

In the first case (*a priori* stable), the general question is whether the resonant motions, which form invariant tori of dimension r arranged into families that fill ℓ -dimensional invariant tori, continue to exist, in presence of small enough perturbations $\varepsilon f(\mathbf{A}, \boldsymbol{\alpha})$, on slightly deformed invariant tori. Similar questions can be asked in the *a priori* unstable cases. To examine the matter more closely consider the formulation of the simplest problems.

A priori stable resonances: more precisely, suppose $\mathcal{H}_0 = \frac{1}{2} \mathbf{A}^2$ and let $\{A_0\} \times \mathbb{T}^\ell$ be the unperturbed invariant torus \mathcal{T}_{A_0} with spectrum $\boldsymbol{\omega}_0 = \boldsymbol{\omega}(A_0) = \partial_A \mathcal{H}_0(A_0)$ with only r rationally independent components. For simplicity, suppose that $\boldsymbol{\omega}_0 = (\omega_1, \dots, \omega_r, 0, \dots, 0) \stackrel{\text{def}}{=} (\boldsymbol{\omega}, \mathbf{0})$ with $\boldsymbol{\omega} \in \mathbb{R}^r$. The more general case in which $\boldsymbol{\omega}$ has only r rationally independent components can be reduced to the special case above by a canonical linear change of coordinates at the price of changing the \mathcal{H}_0 to a new one, still quadratic in the actions but containing mixed products $A_i B_j$: the proofs of the results that are discussed here would not be really affected by such more general form of \mathcal{H} .

It is convenient to distinguish between the “fast” angles $\alpha_1, \dots, \alpha_r$ and the “resonant” angles $\alpha_{r+1}, \dots, \alpha_\ell$ (also called “slow” or “secular”) and call $\boldsymbol{\alpha} = (\boldsymbol{\alpha}', \boldsymbol{\beta})$ with $\boldsymbol{\alpha}' \in \mathbb{T}^r$ and $\boldsymbol{\beta} \in \mathbb{T}^s$. Likewise, we distinguish the fast actions $\mathbf{A}' = (A_1, \dots, A_r)$ and the resonant ones A_{r+1}, \dots, A_ℓ and set $\mathbf{A} = (\mathbf{A}', \mathbf{B})$ with $\mathbf{A}' \in \mathbb{R}^r$ and $\mathbf{B} \in \mathbb{R}^s$.

Therefore, the torus $\mathcal{T}_{A_0}, A_0 = (\mathbf{A}'_0, \mathbf{B}_0)$, is in turn a continuum of invariant tori $\mathcal{T}_{A_0, \boldsymbol{\beta}}$ with trivial parametric equations: $\boldsymbol{\beta}$ fixed, $\boldsymbol{\alpha}' = \boldsymbol{\psi}$, $\boldsymbol{\psi} \in \mathbb{T}^r$, and $\mathbf{A}' = \mathbf{A}'_0, \mathbf{B} = \mathbf{B}_0$. On each of them the motion is: $\mathbf{A}', \mathbf{B}, \boldsymbol{\beta}$ constant and $\boldsymbol{\alpha}' \rightarrow \boldsymbol{\alpha}' + \boldsymbol{\omega} t$, with rationally independent $\boldsymbol{\omega} \in \mathbb{R}^r$.

Then the natural question is whether there exist functions $\mathbf{h}_\varepsilon, \mathbf{k}_\varepsilon, H_\varepsilon, K_\varepsilon$ smooth in ε near $\varepsilon = 0$ and in $\boldsymbol{\psi} \in \mathbb{T}^r$, vanishing for $\varepsilon = 0$, and such that the torus $\mathcal{T}_{A_0, \boldsymbol{\beta}_0, \varepsilon}$ with parametric equations

$$\begin{aligned} \mathbf{A}' &= \mathbf{A}'_0 + H_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{\alpha}' &= \boldsymbol{\psi} + \mathbf{h}_\varepsilon(\boldsymbol{\psi}), \\ \mathbf{B} &= \mathbf{B}_0 + K_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{\beta} &= \boldsymbol{\beta}_0 + \mathbf{k}_\varepsilon(\boldsymbol{\psi}) \end{aligned} \quad \boldsymbol{\psi} \in \mathbb{T}^r \quad [87]$$

is invariant for the motions with Hamiltonian

$$\mathcal{H}_\varepsilon(A, \alpha) = \frac{1}{2}A^2 + \frac{1}{2}B^2 + \varepsilon f(\alpha', \beta)$$

and the motions on it are $\psi \rightarrow \psi + \omega t$. The above property, when satisfied, is summarized by saying that the unperturbed resonant motions $A = (A'_0, B_0)$, $\alpha = (\alpha'_0 + \omega' t, \beta_0)$ can be continued in presence of perturbation εf , for small ε , to quasiperiodic motions with the same spectrum and on a slightly deformed torus $\mathcal{T}_{A'_0, \beta_0, \varepsilon}$.

A priori unstable resonances: here the question is whether the special invariant tori continue to exist in presence of small enough perturbations, of course slightly deformed. This means asking whether, given A_0 such that $\omega(A_0) = \partial_A \mathcal{H}_0(A_0)$ has rationally independent components, there are functions $(H_\varepsilon(\psi), h_\varepsilon(\psi)), (P_\varepsilon(\psi), Q_\varepsilon(\psi))$ and $(\Pi_\varepsilon(\psi), K_\varepsilon(\psi))$ smooth in ε near $\varepsilon = 0$, vanishing for $\varepsilon = 0$, analytic in $\psi \in \mathbb{T}^r$ and such that the r -dimensional surface

$$\begin{aligned} A &= A_0 + H_\varepsilon(\psi), & \alpha &= \psi + h_\varepsilon(\psi) \\ p &= P_\varepsilon(\psi), & q &= Q_\varepsilon(\psi) & \psi &\in \mathbb{T}^r \quad [88] \\ \pi &= \Pi_\varepsilon(\psi), & \kappa &= K_\varepsilon(\psi) \end{aligned}$$

is an invariant torus $\mathcal{T}_{A_0, \varepsilon}$ on which the motion is $\psi \rightarrow \psi + \omega(A_0)t$. Again, the above property is summarized by saying that the unperturbed special resonant motions can be continued in presence of perturbation εf for small ε to quasiperiodic motions with the same spectrum and on a slightly deformed torus $\mathcal{T}_{A_0, \varepsilon}$.

Some answers to the above questions are presented in the following section. For more details, the reader is referred to [Gallavotti et al. \(2004\)](#).

Resonances and Lindstedt Series

We discuss eqns [87] in the paradigmatic case in which the Hamiltonian $\mathcal{H}_0(A)$ is $\frac{1}{2}A^2$ (cf. [80]). It will be $\omega(A') \equiv A'$ so that $A_0 = \omega, B_0 = 0$ and the perturbation $f(\alpha)$ can be considered as a function of $\alpha = (\alpha', \beta)$: let $\bar{f}(\beta)$ be defined as its average over α' . The determination of the invariant torus of dimension r which can be continued in the sense discussed in the last section is easily understood in this case.

A resonant invariant torus which, among the tori $\mathcal{T}_{A_0, \beta}$, has parametric equations that can be continued as a formal power series in ε is the torus $\mathcal{T}_{A_0, \beta_0}$ with β_0 a stationarity point for $\bar{f}(\beta)$, that is, an equilibrium point for the average perturbation: $\partial_\beta \bar{f}(\beta_0) = 0$. In fact, the following theorem holds:

Theorem 5 *If $\omega \in \mathbb{R}^r$ satisfies a Diophantine property and if β_0 is a nondegenerate stationarity point for the “fast angle average” $\bar{f}(\beta)$ (i.e., such that $\det \partial_{\beta\beta}^2 \bar{f}(\beta_0) \neq 0$), then the following equations for the functions $h_\varepsilon, k_\varepsilon$,*

$$\begin{aligned} (\omega \cdot \partial_\psi)^2 h_\varepsilon(\psi) &= -\varepsilon \partial_{\alpha'} f(\psi + h_\varepsilon(\psi), \beta_0 + k_\varepsilon(\psi)) \\ (\omega \cdot \partial_\psi)^2 k_\varepsilon(\psi) &= -\varepsilon \partial_\beta f(\psi + h_\varepsilon(\psi) + k_\varepsilon(\psi)) \end{aligned} \quad [89]$$

can be formally solved in powers of ε .

Given the simplicity of the Hamiltonian [80] that we are considering, it is not necessary to discuss the functions $H_\varepsilon, K_\varepsilon$ because the equations that they should obey reduce to their definitions as in the section “Quasiperiodicity and KAM stability,” and for the same reason.

In other words, also the resonant tori admit a Lindstedt series representation. It is however very unlikely that the series are, in general, convergent.

Physically, this new aspect is due to the fact that the linearization of the motion near the torus $\mathcal{T}_{A_0, \beta_0}$ introduces oscillatory motions around $\mathcal{T}_{A_0, \beta_0}$ with frequencies proportional to the square roots of the positive eigenvalues of the matrix $\varepsilon \partial_{\beta\beta}^2 \bar{f}(\beta_0)$: therefore, it is naively expected that it has to be necessary that a Diophantine property be required on the vector $(\omega, \sqrt{\varepsilon \mu_1}, \dots)$, where $\varepsilon \mu_j$ are the positive eigenvalues. Hence, some values of ε , namely those for which $(\omega, \sqrt{\varepsilon \mu_1}, \dots)$ is not a Diophantine vector or is too close to a non-Diophantine vector, should be excluded or at least should be expected to generate difficulties. Note that the problem arises irrespective of the assumptions about the nondegenerate matrix $\partial_{\beta\beta}^2 \bar{f}(\beta_0)$ (since ε can have either sign), and no matter how small $|\varepsilon|$ is supposed to be. But we can expect that if the matrix $\partial_{\beta\beta}^2 \bar{f}(\beta_0)$ is (say) positive definite (i.e., β_0 is a minimum point for $\bar{f}(\beta)$) then the problem should be easier for $\varepsilon < 0$ and vice versa, if β_0 is a maximum, it should be easier for $\varepsilon > 0$ (i.e., in the cases in which the eigenvalues of $\varepsilon \partial_{\beta\beta}^2 \bar{f}(\beta_0)$ are negative and their roots do not have the interpretation of frequencies).

Technically, the sums of the formal series can be given (so far) a meaning only via summation rules involving divergent series: typically, one has to identify in the formal expressions (denumerably many) geometric series which, although divergent, can be given a meaning by applying the rule [79]. Since the rule can only be applied if $z \neq 1$, this leads to conditions on the parameter ε , in order to exclude that the various z that have to be considered are very close to 1. Hence, this stability result turns out to be rather different from the KAM result for the maximal tori. Namely the series can be given a

meaning via summation rules provided f and β_0 satisfy certain additional conditions and provided certain values of ε are excluded. An example of a theorem is the following:

Theorem 6 *Given the Hamiltonian [80] and a resonant torus $T_{A'_0, \beta_0}$ with $\omega = A'_0 \in \mathbb{R}^r$ satisfying a Diophantine property let β_0 be a nondegenerate maximum point for the average potential $\bar{f}(\beta) \stackrel{\text{def}}{=} (2\pi)^{-r} \int_{\mathbb{T}^r} f(\alpha', \beta) d^r \alpha'$. Consider the Lindstedt series solution for eqns [89] of the perturbed resonant torus with spectrum $(\omega, 0)$. It is possible to express the single n th-order term of the series as a sum of many terms and then rearrange the series thus obtained so that the resummed series converges for ε in a domain \mathcal{E} which contains a segment $[0, \varepsilon_0]$ and also a subset of $[-\varepsilon_0, 0]$ which, although with open dense complement, is so large that it has 0 as a Lebesgue density point. Furthermore, the resummed series for $h_\varepsilon, k_\varepsilon$ define an invariant r -dimensional analytic torus with spectrum ω .*

More generally, if β_0 is only a nondegenerate stationarity point for $\bar{f}(\beta)$, the domain of definition of the resummed series is a set $\mathcal{E} \subset [-\varepsilon_0, \varepsilon_0]$ which on both sides of the origin has an open dense complement although it has 0 as a Lebesgue density point.

Theorem 6 can be naturally extended to the general case in which the Hamiltonian is the most general perturbation of an anisochronous integrable system $\mathcal{H}_\varepsilon(A, \alpha) = b(A) + \varepsilon f(A, \alpha)$ if $\partial_{AA}^2 b$ is a nonsingular matrix and the resonance arises from a spectrum $\omega(A_0)$ which has r independent components (while the remaining are not necessarily zero).

We see that the convergence is a delicate problem for the Lindstedt series for nearly integrable resonant motions. They might even be divergent (mathematically, a proof of divergence is an open problem but it is a very reasonable conjecture in view of the above physical interpretation); nevertheless, **Theorem 6** shows that sum rules can be given that sometimes (i.e., for ε in a large set near $\varepsilon = 0$) yield a true solution to the problem.

This is reminiscent of the phenomenon met in discussing perturbations of isochronous systems in [76], but it is a much more complex situation. It leaves many open problems: foremost among them is the question of uniqueness. The sum rules of divergent series always contain some arbitrary choices, which lead to doubts about the uniqueness of the functions parametrizing the invariant tori constructed in this way. It might even be that the convergence set \mathcal{E} may depend upon the arbitrary choices, and that considering several of them no ε with $|\varepsilon| < \varepsilon_0$ is left out.

The case of *a priori* unstable systems has also been widely studied. In this case too resonances with Diophantine r -dimensional spectrum ω are considered. However, in the case $s_2 = 0$ (called *a priori* unstable hyperbolic resonance) the Lindstedt series can be shown to be convergent, while in the case $s_1 = 0$ (called *a priori* unstable elliptic resonance) or in the mixed cases $s_1, s_2 > 0$ extra conditions are needed. They involve ω and $\mu = (\mu_1, \dots, \mu_{s_2})$ (cf. [86]) and properties of the perturbations as well. It is also possible to study a slightly different problem: namely to look for conditions on ω, μ, f which imply that, for small ε , invariant tori with spectrum ε -dependent but close, in a suitable sense, to ω exist.

The literature is vast, but it seems fair to say that, given the above comments, particularly those concerning uniqueness and analyticity, the situation is still quite unsatisfactory. We refer the reader to [Gallavotti et al. \(2004\)](#) for more details.

Diffusion in Phase Space

The KAM theorem implies that a perturbation of an analytic anisochronous integrable system, i.e., with an analytic Hamiltonian $\mathcal{H}_\varepsilon(A, \alpha) = \mathcal{H}_0(A) + \varepsilon f(A, \alpha)$ and nondegenerate Hessian matrix $\partial_{AA}^2 b(A)$, generates large families of maximal invariant tori. Such tori lie on the energy surfaces but do not have codimension 1 on them, i.e., they do not split the $(2\ell - 1)$ -dimensional energy surfaces into disconnected regions except, of course, in the case of systems with two degrees of freedom (see the section “[Quasiperiodicity and KAM stability](#)”).

Therefore, there might exist trajectories with initial data close to A^i in action space which reach phase space points close to $A^f \neq A^i$ in action space for $\varepsilon \neq 0$, *no matter how small*. Such *diffusion* phenomenon would occur in spite of the fact that the corresponding trajectory has to move in a space in which very close to each $\{A\} \times \mathbb{T}^\ell$ there is an invariant surface on which points move keeping A constant within $O(\varepsilon)$, which for ε small can be $\ll |A^f - A^i|$.

In *a priori* unstable systems (cf. the section “[Resonances and their stability](#)”) with $s_1 = 1, s_2 = 0$, it is not difficult to see that the corresponding phenomenon can actually occur: the paradigmatic example (ARNOL'D) is the *a priori* unstable system

$$\mathcal{H}_\varepsilon = \frac{A_1^2}{2} + A_2 + \frac{p^2}{2} + g(\cos q - 1) + \varepsilon(\cos \alpha_1 + \sin \alpha_2)(\cos q - 1) \quad [90]$$

This is a system describing a motion of a “pendulum” ((p, q) coordinates) interacting with a “rotating wheel” ((A_1, α_1) coordinates) and a “clock” ((A_2, α_2) coordinates) *a priori* unstable near the points $p=0, q=0, 2\pi$ ($s_1=1, s_2=0, \lambda_1=\sqrt{g}$, cf. [86]). It can be proved that on the energy surface of energy E and for each $\varepsilon \neq 0$ small enough (no matter how small) there are initial data with action coordinates close to $A^i = (A_1^i, A_2^i)$ with $(1/2)A_1^i + A_2^i$ close to E eventually evolving to a datum $A' = (A_1', A_2')$ with A_1' at a distance from A_1^i smaller than an arbitrarily prefixed distance (of course with energy E). Furthermore, during the whole process the pendulum energy stays close to zero within $o(\varepsilon)$ (i.e., the pendulum swings following closely the unperturbed separatrices).

In other words, [90] describes a machine (the pendulum) which, working approximately in a cycle, extracts energy from a reservoir (the clock) to transfer it to a mechanical device (the wheel). The statement that diffusion is possible means that the machine can work as soon as $\varepsilon \neq 0$, if the initial actions and the initial phases (i.e., α_1, α_2, p, q) are suitably tuned (as functions of ε).

The peculiarity of the system [90] is that the fixed points P_{\pm} of the unperturbed pendulum (i.e., the equilibria $p=0, q=0, 2\pi$) *remain* unstable equilibria even when $\varepsilon \neq 0$ and this is an important simplifying feature.

It is a peculiarity that permits bypassing the obstacle, arising in the analysis of more general cases, represented by the resonance surfaces consisting of the A 's with $A_1\nu_1 + \nu_2 = 0$: the latter correspond to harmonics (ν_1, ν_2) present in the perturbing function, i.e., the harmonics which would lead to division by zero in an attempt to construct (as necessary in studying [90] by Arnol'd's method) the parametric equations of the perturbed invariant tori with action close to such A 's. In the case of [90] the problem arises only on the resonance marked in Figure 6 by a heavy line, i.e., $A_1=0$, corresponding to $\cos \alpha_1$ in [90].

If $\varepsilon=0$, the points P_- with $p=0, q=0$ and the point P_+ with $p=0, q=2\pi$ are both unstable equilibria (and they are, of course, the same point, if q is an angular variable). The unstable manifold (it is a curve) of P_+ coincides with the stable manifold of P_- and vice versa. So that the unperturbed system admits nontrivial motions leading from P_+ to P_- and from P_- to P_+ , both in a bi-infinite time interval $(-\infty, \infty)$: the p, q variables describe a pendulum and P_{\pm} are its unstable equilibria which are connected by the separatrices (which constitute the zero-energy surfaces for the pendulum).

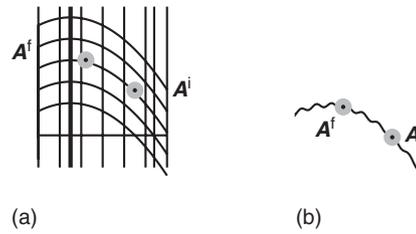


Figure 6 (a) The $\varepsilon=0$ geometry: the “partial energy” lines are parabolas, $(1/2)A_1^2 + A_2 = \text{const}$. The vertical lines are the resonances $A_1 = \text{rational}$ (i.e., $\nu_1 A_1 + \nu_2 = 0$). The disks are neighborhoods of the points A^i and A^f (the dots at their centers). (b) $\varepsilon \neq 0$; an artist's rendering of a trajectory in A space, driven by the pendulum swings to accelerate the wheel from A_1^i to A_1^f at the expenses of the clock energy, sneaking through invariant tori not represented and (approximately) located “away” from the intersections between resonances and partial energy lines (a dense set, however). The pendulum coordinates are not shown: its energy stays close to zero, within a power of ε . Hence the pendulum swings, staying close to the separatrix. The oscillations symbolize the wiggly behavior of the partial energy $(1/2)A_1^2 + A_2$ in the process of sneaking between invariant tori which, because of their invariance, would be impossible without the pendulum. The energy $(1/2)A_1^2$ of the wheel increases slightly at each pendulum swing: accurate estimates yield an increase of the wheel speed A_1 of the order of $\varepsilon/(\log \varepsilon^{-1})$ at each swing of the pendulum implying a transition time of the order of $g^{-1/2}\varepsilon^{-1} \log \varepsilon^{-1}$.

The latter property remains true for more general *a priori* unstable Hamiltonians

$$\mathcal{H}_\varepsilon = \mathcal{H}_0(A) + \mathcal{H}_u(p, q) + \varepsilon f(A, \alpha, p, q) \quad [91]$$

$$\text{in } (U \times \mathbb{T}^\ell) \times (\mathbb{R}^2)$$

where \mathcal{H}_u is a one-dimensional Hamiltonian which has two unstable equilibrium points P_+ and P_- linearly repulsive in one direction and linearly attractive in another which are connected by two heteroclinic trajectories which, as time tends to $\pm\infty$, approach P_- and P_+ and vice versa.

Actually, the points need not be different but, if coinciding, the trajectories linking them must be nontrivial: in the case [90] the variable q can be considered an angle and then P_+ and P_- would coincide (but are connected by nontrivial trajectories, i.e., by trajectories that also visit points different from P_{\pm}). Such trajectories are called heteroclinic if $P_+ \neq P_-$ and homoclinic if $P_+ = P_-$.

In the general case, besides the homoclinicity (or heteroclinicity) condition, certain weak genericity conditions, automatically satisfied in the example [90], have to be imposed in order to show that, given A^i and A^f with the same unperturbed energy E , one can find, for all ε small enough but not equal to zero, initial data (ε -dependent) with actions arbitrarily close to A^i which evolve to data with actions arbitrarily close to A^f . This is a phenomenon

called the *Arnol'd diffusion*. Simple sufficient conditions for a transition from near A^i to near A^f are expressed by the following result:

Theorem 7 *Given the Hamiltonian [91] with \mathcal{H}_u admitting two hyperbolic fixed points P_{\pm} with heteroclinic connections, $t \rightarrow (p_a(t), q_a(t))$, $a=1,2$, suppose that:*

- (i) *On the unperturbed energy surface of energy $E = \mathcal{H}(A^i) + \mathcal{H}_u(P_{\pm})$ there is a regular curve $\gamma: s \rightarrow A(s)$ joining A^i to A^f such that the unperturbed tori $\{A(s)\} \times \mathbb{T}^{\ell}$ can be continued at $\varepsilon \neq 0$ into invariant tori $T_{A(s),\varepsilon}$ for a set of values of s which fills the curve γ leaving only gaps of size of order $o(\varepsilon)$.*
- (ii) *The $\ell \times \ell$ matrix D_{ij} of the second derivatives of the integral of f over the heteroclinic motions is not degenerate, that is,*

$$\begin{aligned} & |\det D| \\ &= \left| \det \left(\int_{-\infty}^{\infty} dt \partial_{\alpha_i \alpha_j} f(A, \alpha + \omega(A)t, \right. \right. \\ & \quad \left. \left. p_a(t), q_a(t) \right) \right| > c > 0 \end{aligned} \quad [92]$$

for all A 's on the curve γ and all $\alpha \in \mathbb{T}^2$.

Given arbitrary $\rho > 0$, for $\varepsilon \neq 0$ small enough there are initial data with action and energy closer than ρ to A^i and E , respectively, which after a long enough time acquire an action closer than ρ to A^f (keeping the initial energy).

The above two conditions can be shown to hold generically for many pairs $A^i \neq A^f$ (and many choices of the curves γ connecting them) if the number of degrees of freedom is ≥ 3 . Thus, the result, obtained by a simple extension of the argument originally outlined by Arnol'd to discuss the paradigmatic example [90], proves the existence of diffusion in *a priori* unstable systems. The integral in [92] is called *Melnikov integral*.

The real difficulty is to estimate the time needed for the transition: it is a time that obviously has to diverge as $\varepsilon \rightarrow 0$. Assuming g fixed (i.e., ε independent) a naive approach easily leads to estimates which can even be worse than $O(\exp(a\varepsilon^{-b}))$ with some $a, b > 0$. It has finally been shown that in such cases the minimum time can be, for rather general perturbations $\varepsilon f(\alpha, q)$, estimated above by $O(\varepsilon^{-1} \log \varepsilon^{-1})$, which is the best that can be hoped for under generic assumptions.

The reader is referred to Arnol'd (1989) and Chierchia and Valdinoci (2000) for more details.

Long-Time Stability of Quasiperiodic Motions

A more difficult problem is whether the same phenomenon of migration in action space occurs in *a priori* stable systems. The root of the difficulty is a remarkable stability property of quasiperiodic motions. Consider Hamiltonians $\mathcal{H}_{\varepsilon}(A, \alpha) = h(A) + \varepsilon f(A, \alpha)$ with $\mathcal{H}_0(A) = h(A)$ strictly convex, analytic, and anisochronous on the closure \bar{U} of an open bounded region $U \subset \mathbb{R}^{\ell}$, and a perturbation $\varepsilon f(A, \alpha)$ analytic in $\bar{U} \times \mathbb{T}^{\ell}$.

Then *a priori* bounds are available on how long it can possibly take to migrate from an action close to A_1 to one close to A_2 : and the bound is of “exponential type” as $\varepsilon \rightarrow 0$ (i.e., it admits a lower bound which behaves as the exponential of an inverse power of ε). The simplest theorem is (NEKHOROSSEV):

Theorem 7 *There are constants $0 < a, b, d, g, \tau$ such that any initial datum (A, α) evolves so that A will not change by more than $a\varepsilon^g$ before a long time bounded below by $\tau \exp(b\varepsilon^{-d})$.*

Thus, this puts an exponential bound, i.e., a bound exponential in an inverse power of ε , to the diffusion time: before a time $\tau \exp(b\varepsilon^{-d})$ actions can only change by $O(\varepsilon^g)$ so that their variation cannot be large no matter how small $\varepsilon \neq 0$ is chosen. This places a (long) lower bound to the time of diffusion in *a priori* stable systems.

The proof of the theorem provides, actually, an interesting and detailed picture of the variations in actions showing that some actions may vary more slowly than others.

The theorem is constructive, i.e., all constants $0 < a, b, d, \tau$ can be explicitly chosen and depend on ℓ, \mathcal{H}_0, f although some of them can be fixed to depend only on ℓ and on the minimum curvature of the convex graph of \mathcal{H}_0 . Its proof can be adapted to cover many cases which do not fall in the class of systems with strictly convex unperturbed Hamiltonian, and even to cases with a resonant unperturbed Hamiltonian.

However, in important problems (e.g., in the three-body problems met in celestial mechanics) there is empirical evidence that diffusion takes place at a fast pace (i.e., not exponentially slow in the above sense) while the above results would forbid a rapid migration in phase space if they applied: however, in such problems the assumptions of the theorem are not satisfied, because the unperturbed system is strongly resonant (as in the celestial mechanics problems, where the number of independent frequencies is a fraction of the number

of degrees of freedom and $b(\mathbf{A})$ is far from strictly convex), leaving wide open the possibility of observing rapid diffusion.

Further, changing the assumptions can dramatically change the results. For instance, rapid diffusion can sometimes be proved even though it might be feared that it should require exponentially long times: an example that has been proposed is the case of a three-timescales system, with Hamiltonian

$$\omega_1 A_1 + \omega_2 A_2 + \frac{p^2}{2} + g(1 + \cos q) + \varepsilon f(\alpha_1, \alpha_2, p, q) \quad [93]$$

with $\boldsymbol{\omega}_\varepsilon \stackrel{\text{def}}{=} (\omega_1, \omega_2)$, where $\omega_1 = \varepsilon^{-1/2} \bar{\omega}$, $\omega_2 = \varepsilon^{1/2} \tilde{\omega}$ and $\bar{\omega}, \tilde{\omega} > 0$ constants. The three scales are $\omega_1^{-1}, \sqrt{g^{-1}}, \omega_2^{-1}$. In this case, there are many (although by no means all) pairs A_1, A_2 which can be connected within a time that can be estimated to be of order $O(\varepsilon^{-1} \log \varepsilon^{-1})$.

This is a rapid-diffusion case in an *a priori* unstable system in which condition [92] is not satisfied: because the ε -dependence of $\boldsymbol{\omega}(\mathbf{A})$ implies that the lower bound c in [92] must depend on ε (and be exponentially small with an inverse power of ε as $\varepsilon \rightarrow 0$).

The unperturbed system in [93] is nonresonant in the \mathcal{H}_0 part for $\varepsilon > 0$ outside a set of zero measure (i.e., where the vector $\boldsymbol{\omega}_\varepsilon$ satisfies a suitable Diophantine property) and, furthermore, it is *a priori* unstable: cases met in applications can be *a priori* stable and resonant (and often not anisochronous) in the \mathcal{H}_0 part. In such a system, not only the speed of diffusion is not understood but proposals to prove its existence, if present (as expected), have so far not given really satisfactory results.

For more details, the reader is referred to Nekhoroshev (1977).

The Three-Body Problem

Mechanics and the three-body problem can be almost identified with each other, in the sense that the motion of three gravitating masses has long been a key astronomical problem and at the same time the source of inspiration for many techniques: foremost among them the theory of perturbations.

As an introduction, consider a special case. Let three masses $m_S = m_0, m_J = m_1, m_M = m_2$ interact via gravity, that is, with interaction potential $-km_i m_j |\mathbf{x}_i - \mathbf{x}_j|^{-1}$: the simplest problem arises when the third body has a negligible mass compared to the two others and the latter are supposed to be on a circular orbit; furthermore, the mass m_j is εm_S

with ε small and the mass m_M moves in the plane of the circular orbit. This will be called the “circular restricted three-body problem.”

In a reference system with center S and rotating at the angular speed of J around S inertial forces (centrifugal and Coriolis) act. Supposing that the body J is located on the axis with unit vector \mathbf{i} at distance R from the origin S , the acceleration of the point M is

$$\ddot{\mathbf{q}} = \mathbf{F} + \omega_0^2 \left(\mathbf{q} - \frac{\varepsilon R}{1 + \varepsilon} \mathbf{i} \right) - 2\boldsymbol{\omega}_0 \wedge \dot{\mathbf{q}}$$

if \mathbf{F} is the force of attraction and $\boldsymbol{\omega}_0 \wedge \dot{\mathbf{q}} \equiv \omega_0 \dot{\mathbf{q}}^\perp$ where $\boldsymbol{\omega}_0$ is a vector with $|\boldsymbol{\omega}_0| = \omega_0$ and perpendicular to the orbital plane and $\mathbf{q}^\perp \stackrel{\text{def}}{=} (-\rho_2, \rho_1)$ if $\mathbf{q} = (\rho_1, \rho_2)$. Here, taking into account that the origin S rotates around the fixed center of mass, $\omega_0^2 (\mathbf{q} - \varepsilon R / (1 + \varepsilon) \mathbf{i})$ is the centrifugal force while $-2\boldsymbol{\omega}_0 \wedge \dot{\mathbf{q}}$ is the Coriolis force. The equations of motion can therefore be derived from a Lagrangian

$$\mathcal{L} = \frac{1}{2} \dot{\mathbf{q}}^2 - W + \omega_0 \mathbf{q}^\perp \cdot \dot{\mathbf{q}} + \frac{1}{2} \omega_0^2 \mathbf{q}^2 - \omega_0^2 \frac{\varepsilon R}{1 + \varepsilon} \mathbf{q} \cdot \mathbf{i} \quad [94]$$

with

$$\omega_0^2 R^3 = km_S (1 + \varepsilon) \stackrel{\text{def}}{=} g_0$$

$$W = -\frac{km_S}{|\mathbf{q}|} - \frac{km_S \varepsilon}{|\mathbf{q} - R\mathbf{i}|}$$

where k is the gravitational constant, R the distance between S and J , and finally the last three terms in [94] come from the Coriolis force (the first) and from the centripetal force (the other two, taking into account that the origin S rotates around the fixed center of mass).

Setting $g = g_0 / (1 + \varepsilon) \equiv km_S$, the Hamiltonian of the system is

$$\mathcal{H} = \frac{1}{2} (\mathbf{p} - \omega_0 \mathbf{q}^\perp)^2 - \frac{g}{|\mathbf{q}|} - \frac{1}{2} \omega_0^2 \mathbf{q}^2 - \varepsilon \frac{g}{R} \left(\left| \frac{\mathbf{q}}{R} - \mathbf{i} \right|^{-1} - \frac{\mathbf{q}}{R} \cdot \mathbf{i} \right) \quad [95]$$

The first part can be expressed immediately in the action–angle coordinates for the two-body problem (cf. the section “Newtonian potential and Kepler’s laws”). Calling such coordinates $(L_0, \lambda_0, G_0, \gamma_0)$ and θ_0 the polar angle of M with respect to the major axis of the ellipse and λ_0 the mean anomaly of M on its ellipse, the Hamiltonian becomes, taking into account that for $\varepsilon = 0$ the ellipse axis rotates at speed $-\omega_0$,

$$\mathcal{H} = -\frac{g^2}{2L_0^2} - \omega_0 G_0 - \varepsilon \frac{g}{R} \left(\left| \frac{\mathbf{q}}{R} - \mathbf{i} \right|^{-1} - \frac{\mathbf{q}}{R} \cdot \mathbf{i} \right) \quad [96]$$

which is convenient if we study the interior problem, i.e., $|\varrho| < R$. This can be expressed in the action-angle coordinates via [41], [42]:

$$\begin{aligned} \theta_0 &= \lambda_0 + f_{\lambda_0}, & \theta_0 + \gamma_0 &= \lambda_0 + \gamma_0 + f_{\lambda_0} \\ e &= \left(1 - \frac{G_0^2}{L_0^2}\right)^{1/2}, & \frac{|\varrho|}{R} &= \frac{G_0^2}{gR} \frac{1}{1 + e \cos(\lambda_0 + f_{\lambda_0})} \end{aligned} \quad [97]$$

where (see [42]), $f_\lambda = f(e \sin \lambda, e \cos \lambda)$ and

$$f(x, y) = 2x \left(1 + \frac{5}{4}y + \dots\right)$$

with the ellipsis denoting higher orders in x, y even in x . The Hamiltonian takes the form, if $\omega^2 = gR^{-3}$,

$$\mathcal{H}_\varepsilon = -\frac{g^2}{2L_0^2} - \omega G_0 + \varepsilon \frac{g}{R} F(G_0, L_0, \lambda_0, \lambda_0 + \gamma_0) \quad [98]$$

where the only important feature (for our purposes) is that $F(L, G, \alpha, \beta)$ is an analytic function of L, G, α, β near a datum with $|G| < L$ (i.e., $e > 0$) and $|\varrho| < R$. However, the domain of analyticity in G is rather small as it is constrained by $|G| < L$ excluding in particular the circular orbit case $G = \pm L$.

Note that apparently the KAM theorem fails to be applicable to [98] because the matrix of the second derivatives of $\mathcal{H}_0(L, G)$ has vanishing determinant. Nevertheless, the proof of the theorem also goes through in this case, with minor changes. This can be checked by studying the proof or, following a remark by Poincaré, by simply noting that the “squared” Hamiltonian $\mathcal{H}'_\varepsilon \stackrel{\text{def}}{=} (\mathcal{H}_\varepsilon)^2$ has the form

$$\mathcal{H}'_\varepsilon = \left(-\frac{g^2}{2L_0^2} - \omega G_0\right)^2 + \varepsilon F'(G_0, L_0, \lambda_0, \lambda_0 + \gamma_0) \quad [99]$$

with F' still analytic. *But* this time

$$\begin{aligned} \det \frac{\partial^2 \mathcal{H}'_0}{\partial(G_0, L_0)} &= -6g^2 L_0^{-4} \omega_0^2 b \neq 0 \\ \text{if } b &= -g^2 L_0^{-2} - 2\omega G_0 \neq 0 \end{aligned}$$

Therefore, the KAM theorem applies to \mathcal{H}'_ε and the key observation is that the orbits generated by the Hamiltonian $(\mathcal{H}'_\varepsilon)^2$ are geometrically the same as those generated by the Hamiltonian \mathcal{H}_ε : they are only run at a different speed because of the need of a time rescaling by the constant factor $2\mathcal{H}_\varepsilon$.

This shows that, given an unperturbed ellipse of parameters (L_0, G_0) such that $\boldsymbol{\omega} = (g^2/L_0^3, -\omega)$, $G_0 > 0$, with ω_1/ω_2 Diophantine, then the perturbed system admits a motion which is quasiperiodic with spectrum proportional to $\boldsymbol{\omega}$ and takes place on an orbit which wraps around a torus remaining forever close to the unperturbed torus (which can be visualized as described by a point moving, according to the area law

on an ellipse rotating at a rate $-\omega_0$) with actions (L_0, G_0) , provided ε is small enough. Hence,

The KAM theorem answers, at least conceptually, the classical question: can a solution of the three-body problem remain close to an unperturbed one forever? That is, is it possible that a solar system is stable forever?

Assuming $e, |\varrho|/R \ll 1$ and retaining only the lowest orders in e and $|\varrho|/R \ll 1$ the Hamiltonian [98] simplifies into

$$\begin{aligned} \mathcal{H} &= -\frac{g^2}{2L_0^2} - \omega G_0 + \delta_\varepsilon(G_0) - \frac{\varepsilon g}{2R} \frac{G_0^4}{g^2 R^2} \left(3 \cos 2(\lambda_0 + \gamma_0) \right. \\ &\quad \left. - e \cos \lambda_0 - \frac{9}{2} e \cos(\lambda_0 + 2\gamma_0) \right. \\ &\quad \left. + \frac{3}{2} e \cos(3\lambda_0 + 2\gamma_0)\right) \end{aligned} \quad [100]$$

where

$$\begin{aligned} \delta_\varepsilon(G_0) &= -((1 + \varepsilon)^{1/2} - 1)\omega G_0 - \frac{\varepsilon g}{2R} \frac{G_0^4}{g^2 R^2} \\ e &= \left(1 - \frac{G_0^2}{L_0^2}\right)^{1/2} \end{aligned}$$

It is an interesting exercise to estimate, assuming as model the Hamiltonian [100] and following the proof of the KAM theorem, how small has ε to be if a planet with the data of Mercury can be stable forever on a (slowly precessing) orbit with actions close to the present-day values under the influence of a mass ε times the solar mass orbiting on a circle, at a distance from the Sun equal to that of Jupiter. It is possible to follow either the above reduction to the ordinary KAM theorem or to apply directly to [100] the Lindstedt series expansion, proceeding along the lines of the section “Quasiperiodicity and KAM stability.” The first approach is easy but the second is more efficient: in both cases, unless the estimates are done in a particularly careful manner, the value found for εm_S is not interesting from the viewpoint of astronomy.

The reader is referred to Arnol’d (1989) for more details.

Rationalization and Regularization of Singularities

Often integrable systems have interesting data which lie on the boundary of the integrability domain. For instance, the central motion when $L = G$ (circular orbits) or the rigid body in a rotation around one of the principal axes or the two-body problem when $G = 0$ (collisional data). In such cases, perturbation

theory cannot be applied as discussed above. Typically, the perturbation depends on quantities like $\sqrt{L-G}$ and is *not analytic* at $L=G$. Nevertheless, it is sometimes possible to enlarge phase space and introduce new coordinates in the vicinity of the data which in the initial phase space are singular.

A notable example is the failure of the analysis of the circular restricted three-body problem: it apparently fails when the orbit that we want to perturb is circular.

It is convenient to introduce the canonical coordinates L, λ and G, γ :

$$\begin{aligned} L &= L_0, & G &= L_0 - G_0 \\ \lambda &= \lambda_0 + \gamma_0, & \gamma &= -\gamma_0 \end{aligned} \quad [101]$$

so that $e = \sqrt{2GL^{-1}}\sqrt{1 - G(2L)^{-1}}$ and $\lambda_0 = \lambda + \gamma$ and $\theta_0 = \lambda_0 + f_{\lambda_0}$, where f_{λ_0} is defined in [42] (see also [97]). Hence,

$$\begin{aligned} \theta_0 &= \lambda + \gamma + f_{\lambda+\gamma}, & \theta_0 + \gamma_0 &= \lambda + f_{\lambda+\gamma} \\ e &= \sqrt{2G}\sqrt{\frac{1}{L}\left(1 - \frac{G}{2L}\right)} \\ \frac{|\varrho|}{R} &= \frac{L^2(1 - e^2)}{gR} \frac{1}{1 + e \cos(\lambda + \gamma + f_{\lambda+\gamma})} \end{aligned} \quad [102]$$

and the Hamiltonian [100] takes the form

$$\begin{aligned} \mathcal{H}_\varepsilon &= -\frac{g^2}{2L^2} - \omega L + \omega G \\ &+ \varepsilon \frac{g}{R} F(L - G, L, \lambda + \gamma, \lambda) \end{aligned} \quad [103]$$

In the coordinates L, G of [101] the unperturbed circular case corresponds to $G=0$ and [96], once expressed in the action-angle variables G, L, γ, λ , is analytic in a domain whose size is controlled by \sqrt{G} . Nevertheless, very often problems of perturbation theory can be “regularized.”

This is done by “enlarging the integrability” domain by adding to it points (one or more) around the singularity (a boundary point of the domain of the coordinates) and introducing new coordinates to describe simultaneously the data close to the singularity and the newly added points: in many interesting cases, the equations of motion are no longer singular (i.e., become analytic) in the new coordinates and are therefore apt to describe the motions that reach the singularity in a finite time. One can say that the singularity was only apparent.

Perhaps this is best illustrated precisely in the above circular restricted three-body problem, with the singularity occurring where $G=0$, that is, at a circular unperturbed orbit. If we describe the points with G small in a new system of coordinates

obtained from the one in [101] by letting alone L, λ and setting

$$p = \sqrt{2G} \cos \gamma, \quad q = \sqrt{2G} \sin \gamma \quad [104]$$

then p, q vary in a neighborhood of the origin with the origin itself excluded.

Adding the origin of the p - q plane then in a full neighborhood of the origin, the Hamiltonian [96] is analytic in L, λ, p, q . This is because it is analytic (cf. [96], [97]) as a function of L, λ and $e \cos \theta_0$ and of $\cos(\lambda_0 + \theta_0)$. Since $\theta_0 = \lambda + \gamma + f_{\lambda+\gamma}$ and $\theta_0 + \lambda_0 = \lambda + f_{\lambda+\gamma}$ by [97], the Hamiltonian [96] is analytic in $L, \lambda, e \cos(\lambda + \gamma + f_{\lambda+\gamma}), \cos(\lambda + f_{\lambda+\gamma})$ for e small (i.e., for G small) and, by [42], $f_{\lambda+\gamma}$ is analytic in $e \sin(\lambda + \gamma)$ and $e \cos(\lambda + \gamma)$. Hence the trigonometric identities

$$\begin{aligned} e \sin(\lambda + \gamma) &= \frac{p \sin \lambda + q \cos \lambda}{\sqrt{L}} \sqrt{1 - \frac{G}{2L}} \\ e \cos(\lambda + \gamma) &= \frac{p \cos \lambda - q \sin \lambda}{\sqrt{L}} \sqrt{1 - \frac{G}{2L}} \end{aligned} \quad [105]$$

together with $G = (1/2)(p^2 + q^2)$ imply that [103] is analytic near $p=q=0$ and $L > 0, \lambda \in [0, 2\pi]$. The Hamiltonian becomes analytic and the new coordinates are suitable to describe motions crossing the origin: for example, by setting

$$C \stackrel{\text{def}}{=} \frac{1}{2} \left(1 - \frac{p^2 + q^2}{4L} \right) L^{-1/2}$$

[100] becomes

$$\begin{aligned} \mathcal{H} &= -\frac{g^2}{2L^2} - \omega L + \omega \frac{1}{2}(p^2 + q^2) \\ &+ \delta_\varepsilon \left(\frac{1}{2}(p^2 + q^2) \right) - \frac{\varepsilon g}{2R} \frac{(L - \frac{1}{2}(p^2 + q^2))^4}{g^2 R^2} \\ &\times (3 \cos 2\lambda - ((-11 \cos \lambda + 3 \cos 3\lambda)p \\ &- (7 \sin \lambda + 3 \sin 3\lambda)q)C) \end{aligned} \quad [106]$$

The KAM theorem does not apply in the form discussed above to “Cartesian coordinates,” that is, when, as in [106], the unperturbed system is not assigned in action-angle variables. However, there are versions of the theorem (actually its corollaries) which do apply and therefore it becomes possible to obtain some results even for the perturbations of circular motions by the techniques that have been illustrated here.

Likewise, the Hamiltonian of the rigid body with a fixed point O and subject to analytic external forces becomes singular, if expressed in the action-angle coordinates of Deprit, when the body motion nears a rotation around a principal axis or, more generally, nears a configuration in which any two of

the axes i_3, z , or z_0 coincide (i.e., any two among the principal axis, the angular momentum axis and the inertial z -axis coincide; see the section “Rigid body”). Nevertheless, by imitating the procedure just described in the simpler cases of the circular three-body problem, it is possible to enlarge the phase space so that in the new coordinates the Hamiltonian is analytic near the singular configurations.

A regularization also arises when considering collisional orbits in the unrestricted planar three-body problem. In this respect, a very remarkable result is the regularization of collisional orbits in the planar three-body problem. After proving that if the total angular momentum does not vanish, simultaneous collisions of the three masses cannot occur within any finite time interval, the question is reduced to the regularization of two-body collisions, under the assumption that the total angular momentum does not vanish.

The local change of coordinates, which changes the relative position coordinates (x, y) of two colliding bodies as $(x, y) \rightarrow (\xi, \eta)$, with $x + iy = (\xi + i\eta)^2$, is not one to one, hence it has to be regarded as an enlargement of the positions space, if points with different (ξ, η) are considered different. However, the equations of motion written in the variables ξ, η have no singularity at $\xi, \eta = 0$ (LEVI-CIVITA).

Another celebrated regularization is the regularization of the Schwarzschild metric, i.e., of the general relativity version of the two-body problem: it is, however, somewhat out of the scope of this review (SYNGE, KRUSKAL).

For more details, the reader is referred to [Levi-Civita \(1956\)](#).

Appendix 1: KAM Resummation Scheme

The idea to control the “remaining contributions” is to reduce the problem to the case in which there are no pairs of lines that follow each other in the tree order and which have the same current. Mark by a scale label “0” the lines, see [74], [83], of a tree whose divisors $C/\omega_0 \cdot \mathbf{v}(l)$ are >1 : these are lines which give no problems in the estimates. Then mark by a scale label “ ≥ 1 ” the lines with current $\mathbf{v}(l)$ such that $|\omega_0 \cdot \mathbf{v}(l)| \leq 2^{-n+1}$ for $n = 1$ (i.e., the remaining lines).

The lines labeled 0 are said to be on scale 0, while those labeled ≥ 1 are said to be on scale ≥ 1 . A cluster of scale 0 will be a maximal collection of lines of scale 0 forming a connected subgraph of a tree θ .

Consider only trees $\theta_0 \in \Theta_0$ of the family Θ_0 of trees containing no clusters of lines with scale label 0 which have only one line entering the cluster and one exiting it with equal current.

It is useful to introduce the notion of a line ℓ_1 situated “between” two lines ℓ, ℓ' with $\ell' > \ell$: this will mean that ℓ_1 precedes ℓ' but not ℓ .

All trees θ in which there are some pairs $\ell' > \ell$ of consecutive lines of scale label ≥ 1 which have equal current and such that all lines between them bear scale label 0 are obtained by “inserting” on the lines of trees in Θ_0 with label ≥ 1 any number of clusters of lines and nodes, with lines of scale 0 and with the property that the sum of the harmonics of the nodes inserted vanishes.

Consider a line $l_0 \in \theta_0 \in \Theta_0$ linking nodes $v_1 < v_2$ and labeled ≥ 1 and imagine inserting on it a cluster γ of lines of scale 0 with sum of the node harmonics vanishing and out of which emerges one line connecting a node v_{out} in γ to v_2 and into which enters one line linking v_1 to a node $v_{in} \in \gamma$. The insertion of a k -lines, $|\gamma| = (k+1)$ -nodes, cluster changes the tree value by replacing the line factor, that will be briefly called “value of the cluster γ ”, as

$$\frac{\mathbf{v}_{v_1} \cdot \mathbf{v}_{v_2}}{\omega_0 \cdot \mathbf{v}(l_0)^2} \rightarrow \frac{(\mathbf{v}_{v_1} \cdot M(\gamma; \mathbf{v}(l_0)) \mathbf{v}_{v_2})}{\omega_0 \cdot \mathbf{v}(l_0)^2} \frac{1}{\omega_0 \cdot \mathbf{v}(l_0)^2} \quad [107]$$

where M is an $\ell \times \ell$ matrix

$$M_{rs}(\gamma, \mathbf{v}(l_0)) = \frac{\varepsilon^{|\gamma|}}{k!} \nu_{out, r} \nu_{in, s} \prod_{v \in \gamma} (-f_{v'}) \prod_{l \in \gamma} \frac{\mathbf{v}_v \cdot \mathbf{v}_{v'}}{\omega_0 \cdot \mathbf{v}(l)^2}$$

if $\ell = v'v$ denotes a line linking v' and v . Therefore, if all possible connected clusters are inserted and the resulting values are added up, the result can be taken into account by attributing to the original line l_0 a factor like [107] with $M^{(0)}(\mathbf{v}(l_0)) \stackrel{\text{def}}{=} \sum_{\gamma} M(\gamma; \mathbf{v}(l_0))$ replacing $M(\gamma; \mathbf{v}(l_0))$.

If several connected clusters γ are inserted on the same line and their values are summed, the result is a modification of the factor associated with the line l_0 into

$$\begin{aligned} & \sum_{k=0}^{\infty} \mathbf{v}_{v_1} \cdot \left(\frac{M^{(0)}(\mathbf{v}(l_0))}{\omega_0 \cdot \mathbf{v}(l_0)^2} \right)^k \mathbf{v}_{v_2} \frac{1}{\omega_0 \cdot \mathbf{v}(l_0)^2} \\ & = \left(\mathbf{v}_{v_1} \cdot \frac{1}{\omega_0 \cdot \mathbf{v}(l_0)^2 - M^{(0)}(\mathbf{v}(l_0))} \mathbf{v}_{v_2} \right) \quad [108] \end{aligned}$$

The series defining $M^{(0)}$ involves, by construction, only trees with lines of scale 0, hence with large divisors, so that it converges to a matrix of small size of order ε (actually ε^2 , more precisely) if ε is small enough.

Convergence can be established by simply remarking that the series defining $M^{(1)}$ is built with lines with values $>(1/2)$ of the propagator, so that it certainly converges for ε small enough (by the estimates in the section “Perturbing functions,” where the propagators were identically 1) and the

sum is of order ε (actually ε^2), hence <1 . However, such an argument cannot be repeated when dealing with lines with smaller propagators (which still have to be discussed). Therefore, a method not relying on so trivial a remark on the size of the propagators has eventually to be used when considering lines of scale higher than 1, as it will soon become necessary.

The advantage of the collection of terms achieved with [108] is that we can represent h as a sum of values of trees which are simpler because they contain no pair of lines of scale ≥ 1 with in between lines of scale 0 with total sum of the node harmonics vanishing. The price is that the divisors are now more involved and we even have a problem due to the fact that we have not proved that the series in [108] converges. In fact, it is a geometric series whose value is the RHS of [108] obtained by the sum rule [79] unless we can prove that the ratio of the geometric series is <1 . This is trivial in this case by the previous remark: but it is better to note that there is another reason for convergence, whose use is not really necessary here but will become essential later.

The property that the ratio of the geometric series is <1 can be regarded as due to the consequence of the cancellation mentioned in the section “Quasi-periodicity and KAM stability” which can be shown to imply that the ratio is <1 because $M^{(0)}(\mathbf{v}) = \varepsilon^2(\boldsymbol{\omega}_0 \cdot \mathbf{v})^2 m^{(0)}(\mathbf{v})$ with $C|m^{(0)}(\mathbf{v})| < D_0$ for some $D_0 > 0$ and for all $|\varepsilon| < \varepsilon_0$ for some ε_0 . Then for small ε the divisor in [108] is essentially still what it was before starting the resummation.

At this point, an induction can be started. Consider trees evaluated with the new rule and place a scale level “ ≥ 2 ” on the lines with $C|\boldsymbol{\omega}_0 \cdot \mathbf{v}(l)| \leq 2^{-n+1}$ for $n=2$: leave the label “0” on the lines already marked so and label by “1” the other lines. The lines of scale “1” will satisfy $2^{-n} < |\boldsymbol{\omega}_0 \cdot \mathbf{v}(l)| \leq 2^{-n+1}$ for $n=1$. The graphs will now possibly contain lines of scale 0, 1 or ≥ 2 while lines with label “ ≥ 1 ” no longer can appear, by construction.

A cluster of scale 1 will be a maximal collection of lines of scales 0, 1 forming a connected subgraph of a tree θ and containing at least one line of scale 1.

The construction carried out by considering clusters of scale 0 can be repeated by considering trees $\theta_1 \in \Theta_1$, with Θ_1 the collection of trees with lines marked 0, 1, or ≥ 2 and in which no pairs of lines with equal momentum appear to follow each other if between them there are only lines marked 0 or 1.

Insertion of connected clusters γ of such lines on a line l_0 of θ_1 leads to define a matrix $M^{(1)}$ formed by summing tree values of clusters γ with lines of scales 0 or 1 evaluated with the line factors defined in [107] and with the restriction that in γ there are no pairs of lines $\ell < \ell'$ with the same current and which

follow each other while any line between them has lower scale (i.e., 0), here “between” means “preceding l' but not preceding l ,” as above.

Therefore, a scale-independent method has to be devised to check the convergence for $M^{(1)}$ and for the matrices to be introduced later to deal with even smaller propagators. This is achieved by the following extension of Siegel’s theorem mentioned in the section “Quasi-periodicity and KAM stability”:

Theorem 8 *Let $\boldsymbol{\omega}_0$ satisfy [74] and set $\boldsymbol{\omega} = C\boldsymbol{\omega}_0$. Consider the contribution to the sum in [82] from graphs θ in which*

- (i) *no pairs $\ell' > \ell$ of lines which lie on the same path to the root carry the same current \mathbf{v} if all lines ℓ_1 between them have current $\mathbf{v}(\ell_1)$ such that $|\boldsymbol{\omega} \cdot \mathbf{v}(\ell_1)| > 2|\boldsymbol{\omega} \cdot \mathbf{v}|$;*
- (ii) *the node harmonics are bounded by $|\mathbf{v}| \leq N$ for some N .*

Then the number of lines ℓ in θ with divisor $\boldsymbol{\omega} \cdot \mathbf{v}_\ell$ satisfying $2^{-n} < |\boldsymbol{\omega} \cdot \mathbf{v}_\ell| \leq 2^{-n+1}$ does not exceed $4Nk2^{-n/\tau}$, $n = 1, 2, \dots$

This implies, by the same estimates in [85], that the series defining $M^{(1)}$ converges. Again, it must be checked that there are cancellations implying that $M^{(1)}(\mathbf{v}) = \varepsilon^2(\boldsymbol{\omega}_0 \cdot \mathbf{v})^2 m^{(1)}(\mathbf{v})$ with $|m^{(1)}(\mathbf{v})| < D_0$ for the same $D_0 > 0$ and the same ε_0 .

At this point, one deals with trees containing only lines carrying labels 0, 1, ≥ 2 , and the line factors for the lines $\ell = v'v$ of scale 0 are $\mathbf{v}_{v'} \cdot \mathbf{v}_v / (\boldsymbol{\omega}_0 \cdot \mathbf{v}(\ell))^2$, those of the lines $\ell = v'v$ of scale 1 have line factors $\mathbf{v}_{v'} \cdot (\boldsymbol{\omega}_0 \cdot \mathbf{v}(\ell))^2 - M^{(0)}(\mathbf{v}(\ell))^{-1} \mathbf{v}_v$, and those of the lines $\ell = v'v$ of scale ≥ 2 have line factors

$$\mathbf{v}_{v'} \cdot (\boldsymbol{\omega}_0 \cdot \mathbf{v}(\ell))^2 - M^{(1)}(\mathbf{v}(\ell))^{-1} \mathbf{v}_v$$

Furthermore, no pair of lines of scale “1” or of scale “ ≥ 2 ” with the same momentum and with only lines of lower scale (i.e., of scale “0” in the first case or of scale “0”, “1” in the second) between them can follow each other.

This procedure can be iterated until, after infinitely many steps, the problem is reduced to the evaluation of tree values in which each line carries a scale label n and there are no pairs of lines which follow each other and which have only lines of lower scale in between. Then the Siegel argument applies once more and the series so resumed is an absolutely convergent series of functions analytic in ε : hence the original series is convergent.

Although at each step there is a lower bound on the denominators, it would not be possible to avoid using Siegel’s theorem. In fact, the lower bound would become worse and worse as the scale increases. In order to check

the estimates of the constants D_0, ε_0 which control the scale independence of the convergence of the various series, it is necessary to take advantage of the theorem, and of the absence (at each step) of the necessity of considering trees with pairs of consecutive lines with equal momentum and intermediate lines of higher scale.

One could also perform the analysis by bounding $h^{(k)}$ order by order with no resummations (i.e., without changing the line factors) and exhibiting the necessary cancellations. Alternatively, the paths that Kolmogorov, Arnol'd and Moser used to prove the first three (somewhat different) versions of the theorem, by successive approximations of the equations for the tori, can be followed.

The invariant tori are Lagrangian manifolds just as the unperturbed ones (cf. comments after [31]) and, in the case of the Hamiltonian [80], the generating function $\mathbf{A} \cdot \boldsymbol{\psi} + \Phi(\mathbf{A}, \boldsymbol{\psi})$ can be expressed in terms of their parametric equations

$$\begin{aligned} \Phi(\mathbf{A}, \boldsymbol{\psi}) &= G(\boldsymbol{\psi}) + \mathbf{a} \cdot \boldsymbol{\psi} + \mathbf{h}(\boldsymbol{\psi}) \cdot (\mathbf{A} - \boldsymbol{\omega} - \Delta \mathbf{h}(\boldsymbol{\psi})) \\ \partial_{\boldsymbol{\psi}} G(\boldsymbol{\psi}) &\stackrel{\text{def}}{=} -\Delta \mathbf{h}(\boldsymbol{\psi}) + \tilde{h}(\boldsymbol{\psi}) \partial_{\boldsymbol{\psi}} \Delta \tilde{h}(\boldsymbol{\psi}) - \mathbf{a} \\ \mathbf{a} &\stackrel{\text{def}}{=} \int (-\Delta \mathbf{h}(\boldsymbol{\psi}) + \tilde{h}(\boldsymbol{\psi}) \partial_{\boldsymbol{\psi}} \Delta \tilde{h}(\boldsymbol{\psi})) \frac{d\boldsymbol{\psi}}{(2\pi)^\ell} \\ &= \int \tilde{h}(\boldsymbol{\psi}) \partial_{\boldsymbol{\psi}} \Delta \tilde{h}(\boldsymbol{\psi}) \frac{d\boldsymbol{\psi}}{(2\pi)^\ell} \end{aligned} \quad [109]$$

where $\Delta = (\boldsymbol{\omega} \cdot \partial_{\boldsymbol{\psi}})$ and the invariant torus corresponds to $\mathbf{A}' = \boldsymbol{\omega}$ in the map $\boldsymbol{\alpha} = \boldsymbol{\psi} + \partial_{\mathbf{A}} \Phi(\mathbf{A}, \boldsymbol{\psi})$ and $\mathbf{A}' = \mathbf{A} + \partial_{\boldsymbol{\psi}} \Phi(\mathbf{A}, \boldsymbol{\psi})$. In fact, by [109] the latter becomes $\mathbf{A}' = \mathbf{A} - \Delta \mathbf{h}$ and, from the second of [75] written for f depending only on the angles $\boldsymbol{\alpha}$, it is $\mathbf{A} = \boldsymbol{\omega} + \Delta \mathbf{h}$ when $\mathbf{A}, \boldsymbol{\alpha}$ are on the invariant torus.

Note that if \mathbf{a} exists it is necessarily determined by the third relation in [109] but the check that the second equation in [109] is soluble (i.e., that the RHS is an exact gradient up to a constant) is nontrivial. The canonical map generated by $\mathbf{A} \cdot \boldsymbol{\psi} + \Phi(\mathbf{A}, \boldsymbol{\psi})$ is also defined for \mathbf{A}' close to $\boldsymbol{\omega}$ and foliates the neighborhood of the invariant torus with other tori: of course, for $\mathbf{A}' \neq \boldsymbol{\omega}$ the tori defined in this way are, in general, not invariant.

The reader is referred to Gallavotti *et al.* (2004) for more details.

Appendix 2: Coriolis and Lorentz Forces – Larmor Precession

Larmor precession refers to the motion of an electrically charged particle in a magnetic field \mathbf{H} (in an inertial frame of reference). It is due to the Lorentz force which, on a unit mass with unit charge, produces an acceleration $\ddot{\mathbf{q}} = \mathbf{v} \wedge \mathbf{H}$ if the speed of light is $c = 1$.

Therefore, if $\mathbf{H} = H\mathbf{k}$ is directed along the \mathbf{k} -axis, the acceleration it produces is the same that the Coriolis force would impress on a unit mass located in a reference frame which rotates with angular velocity $\omega_0 \mathbf{k}$ around the \mathbf{k} -axis if $\mathbf{H} = 2\omega_0 \mathbf{k}$.

The above remarks imply that a homogeneous sphere electrically charged uniformly with a unit charge and freely pivoting about its center in a constant magnetic field H directed along the \mathbf{k} -axis undergoes the same motion as it would follow if not subject to the magnetic field but seen in a noninertial reference frame rotating at constant angular velocity ω_0 around the \mathbf{k} -axis if H and ω_0 are related by $H = 2\omega_0$: in this frame, the Coriolis force is interpreted as a magnetic field.

This holds, however, only if the centrifugal force has zero moment with respect to the center: true in the spherical symmetry case only. In spherically nonsymmetric cases, the centrifugal forces have in general nonzero moment, so the equivalence between Coriolis force and the Lorentz force is only approximate.

The Larmor theorem makes this more precise. It gives a quantitative estimate of the difference between the motion of a general system of particles of mass m in a magnetic field and the motion of the same particles in a rotating frame of reference but in the absence of a magnetic field. The approximation is estimated in terms of the size of the Larmor frequency $eH/2mc$, which should be small compared to the other characteristic frequencies of the motion of the system: the physical meaning is that the centrifugal force should be small compared to the other forces.

The vector potential \mathbf{A} for a constant magnetic field in the \mathbf{k} -direction, $\mathbf{H} = 2\omega_0 \mathbf{k}$, is $\mathbf{A} = 2\omega_0 \mathbf{k} \wedge \mathbf{q} \equiv 2\omega_0 \mathbf{q}^\perp$. Therefore, from the treatment of the Coriolis force in the section “Three-body problem” (see [95]), the motion of a charge e with mass m in a magnetic field \mathbf{H} with vector potential \mathbf{A} and subject to other forces with potential W can be described, in an inertial frame and in generic units, in which the speed of light is c , by a Hamiltonian

$$\mathcal{H} = \frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A} \right)^2 + W(\mathbf{q}) \quad [110]$$

where $\mathbf{p} = m\dot{\mathbf{q}} + (e/c)\mathbf{A}$ and \mathbf{q} are canonically conjugate variables.

Further Reading

- Arnol'd VI (1989) *Mathematical Methods of Classical Mechanics*. Berlin: Springer.
 Calogero F and Degasperis A (1982) *Spectral Transform and Solitons*. Amsterdam: North-Holland.

- Chierchia L and Valdinoci E (2000) A note on the construction of Hamiltonian trajectories along heteroclinic chains. *Forum Mathematicum* 12: 247–255.
- Fassò F (1998) Quasi-periodicity of motions and complete integrability of Hamiltonian systems. *Ergodic Theory and Dynamical Systems* 18: 1349–1362.
- Gallavotti G (1983) *The Elements of Mechanics*. New York: Springer.
- Gallavotti G, Bonetto F, and Gentile G (2004) *Aspects of the Ergodic, Qualitative and Statistical Properties of Motion*. Berlin: Springer.
- Kolmogorov N (1954) On the preservation of conditionally periodic motions. *Doklady Akademii Nauk SSSR* 96: 527–530.
- Landau LD and Lifshitz EM (1976) *Mechanics*. New York: Pergamon Press.
- Levi-Civita T (1956) *Opere Matematiche*. *Accademia Nazionale dei Lincei*. Bologna: Zanichelli.
- Moser J (1962) On invariant curves of an area preserving mapping of the annulus. *Nachrichten Akademie Wissenschaften Göttingen* 11: 1–20.
- Nekhoroshev V (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Russian Mathematical Surveys* 32(6): 1–65.
- Poincaré H (1987) *Méthodes nouvelles de la mécanique céleste* vol. I. Paris: Gauthier-Villars. (reprinted by Gabay, Paris, 1987).

Introductory Article: Differential Geometry

S Paycha, Université Blaise Pascal, Aubière, France

© 2006 Elsevier Ltd. All rights reserved.

Differential geometry is the study of differential properties of geometric objects such as curves, surfaces and higher-dimensional manifolds endowed with additional structures such as metrics and connections. One of the main ideas of differential geometry is to apply the tools of analysis to investigate geometric problems; in particular, it studies their “infinitesimal parts,” thereby linearizing the problem. However, historically, geometric concepts often anticipated the analytic tools required to define them from a differential geometric point of view; the notion of tangent to a curve, for example, arose well before the notion of derivative.

In its barely more than two centuries of existence, differential geometry has always had strong (often two-way) interactions with physics. Just to name a few examples, the theory of curves is used in kinematics, symplectic manifolds arise in Hamiltonian mechanics, pseudo-Riemannian manifolds in general relativity, spinors in quantum mechanics, Lie groups and principal bundles in gauge theory, and infinite-dimensional manifolds in the path-integral approach to quantum field theory.

Curves and Surfaces

The study of differential properties of curves and surfaces resulted from a combination of the coordinate method (or analytic geometry) developed by Descartes and Fermat during the first half of the seventeenth century and infinitesimal calculus developed by Leibniz and Newton during the second half of the seventeenth and beginning of the eighteenth century.

Differential geometry appeared later in the eighteenth century with the works of Euler *Recherches sur la courbure des surfaces* (1760) (Investigations on the curvature of surfaces) and Monge *Une application de l'analyse à la géométrie* (1795) (An application of analysis to geometry). Until Gauss' fundamental article *Disquisitiones generales circa superficies curvas* (General investigations of curved surfaces) published in Latin in 1827 (of which one can find a partial translation to English in Spivak (1979)), surfaces embedded in \mathbb{R}^3 were either described by an equation, $W(x, y, z) = 0$, or by expressing one variable in terms of the others. Although Euler had already noticed that the coordinates of a point on a surface could be expressed as functions of two independent variables, it was Gauss who first made a systematic use of such a parametric representation, thereby initiating the concept of “local chart” which underlies differential geometry.

Differentiable Manifolds

The actual notion of n -manifold independent of a particular embedding in a Euclidean space goes back to a lecture *Über die Hypothesen, welche der Geometrie zu Grunde liegen* (On the hypotheses which lie at the foundations of geometry) (of which one can find a translation to English and comments in Spivak (1979)) delivered by Riemann at Göttingen University in 1854, in which he makes clear the fact that n -manifolds are locally like n -dimensional Euclidean space. In his work, Riemann mentions the existence of infinite-dimensional manifolds, such as function spaces, which today play an important role since they naturally arise as configuration spaces in quantum field theories.

In modern language a differentiable manifold modeled on a topological space V (which can be

finite dimensional, Fréchet, Banach, or Hilbert for example) is a topological space M equipped with a family of local coordinate charts $(U_i, \phi_i)_{i \in I}$ such that the open subsets $U_i \subset M$ cover M and where $\phi_i: U_i \rightarrow V$, $i \in I$, are homeomorphisms which give rise to smooth transition maps $\phi_i \circ \phi_j^{-1}: \phi_j(U_i \cap U_j) \rightarrow \phi_i(U_i \cap U_j)$. An n -dimensional differentiable manifold is a differentiable manifold modeled on \mathbb{R}^n . The sphere $S^{n-1} := \{(x_1, \dots, x_n) \in \mathbb{R}^n, \sum_{i=1}^n x_i^2 = 1\}$ is a differentiable manifold of dimension $n - 1$.

Simple differentiable curves in \mathbb{R}^n are one-dimensional differentiable manifolds locally specified by coordinates $x(t) = (x_1(t), \dots, x_n(t)) \in \mathbb{R}^n$, where $t \mapsto x_j(t)$ is of class C^k . The tangent at point $x(t_0)$ to such a curve, which is a straight line passing through this point with direction given by the vector $x'(t_0)$, generalizes to the concept of tangent space $T_m M$ at point $m \in M$ of a smooth manifold M modeled on V which is a vector space isomorphic to V spanned by tangent vectors at point m to curves $\gamma(t)$ of class C^1 on M such that $\gamma(t_0) = m$.

In order to make this more precise, one needs the notion of differentiable mapping. Given two differentiable manifolds M and N , a mapping $f: M \rightarrow N$ is differentiable at point m if, for every chart (U, ϕ) of M containing m and every chart (V, ψ) of N such that $f(U) \subset V$, the mapping $\psi \circ f \circ \phi^{-1}: \phi(U) \rightarrow \psi(V)$ is differentiable at point $\phi(m)$. In particular, differentiable mappings $f: M \rightarrow \mathbb{R}$ form the algebra $C^\infty(M, \mathbb{R})$ of smooth real-valued functions on M . Differentiable mappings $\gamma: [a, b] \rightarrow M$ from an interval $[a, b] \subset \mathbb{R}$ to a differentiable manifold M are called “differentiable curves” on M . A differentiable mapping $f: M \rightarrow N$ which is invertible and with differentiable inverse $f^{-1}: N \rightarrow M$ is called a diffeomorphism.

The derivative of a function $f \in C^\infty(M, \mathbb{R})$ along a curve $\gamma: [a, b] \rightarrow M$ at point $\gamma(t_0) \in M$ with $t_0 \in [a, b]$ is given by

$$Xf := \frac{d}{dt}\bigg|_{t=t_0} f \circ \gamma(t)$$

and the map $f \mapsto Xf$ is called the tangent vector to the curve γ at point $\gamma(t_0)$. Tangent vectors to some curve $\gamma: [a, b] \rightarrow M$ at a given point $m \in \gamma([a, b])$ form a vector space $T_m M$ called the “tangent space” to M at point m .

A (smooth) map which, to a point $m \in M$, assigns a tangent vector $X \in T_m M$ is called a (smooth) vector field. It can also be seen as a derivation $\tilde{X}: f \mapsto Xf$ on $C^\infty(M, \mathbb{R})$ defined by $(\tilde{X}f)(m) := X(m)f$ for any $m \in M$ and the bracket of vector fields is thereby defined from the operator bracket $[\tilde{X}, \tilde{Y}] := \tilde{X} \circ \tilde{Y} - \tilde{Y} \circ \tilde{X}$. The linear operations on tangent vectors carry out to vector fields $(X + Y)(m) := X(m) + Y(m)$, $(\lambda X)(m) := \lambda X(m)$ for any

$m \in M$ and for any $X, Y \in T_m M, \lambda \in \mathbb{R}$ so that vector fields on M build a linear space.

One can generate tangent vectors to M via local one-parameter groups of differentiable transformations of M , that is, mappings $(t, m) \mapsto \phi_t(m)$ from $]-\epsilon, \epsilon[\times U$ to U (with $\epsilon > 0$ and $U \subset M$ an open subset of M) such that $\phi_0 = \text{Id}$, $\phi_{t+s} = \phi_t \circ \phi_s \forall s, t \in]-\epsilon, \epsilon[$ with $t + s \in]-\epsilon, \epsilon[$ and $m \mapsto \phi_t(m)$ is a diffeomorphism of U onto an open subset $\phi_t(U)$. The tangent vector at $t=0$ to the curve $\gamma(t) = \phi_t(m)$ yields a tangent vector to M at point $m = \gamma(0)$. Conversely, when M is finite dimensional, the fundamental theorem for systems of ordinary equations yields, for any vector field X on M , the existence (around any point $m \in M$) of a local one-parameter group of local transformations $\phi:]-\epsilon, \epsilon[\times U \rightarrow M$ (with U an open subset containing m) which induces the tangent vector $X(m) \in T_m M$.

A differentiable mapping $\phi: M \rightarrow N$ induces a map $\phi_*(m): T_m M \rightarrow T_{\phi(m)} N$ defined by $\phi_* Xf = X(f \circ \phi)$. An “immersion” of a manifold M in a manifold N is a differentiable mapping $\phi: M \rightarrow N$ such that the maps $\phi_*(m)$ are injective at any point $m \in M$. Such a map is an embedding if it is moreover injective in which case $\phi(M) \subset N$ is a submanifold of N . The unit sphere S^n is a submanifold of \mathbb{R}^{n+1} . Whitney showed that every smooth real n -dimensional manifold can be embedded in \mathbb{R}^{2n+1} .

A differentiable manifold whose coordinate charts take values in a complex vector space V and whose transition maps are holomorphic is called a complex manifold, which is complex n -dimensional if $V = \mathbb{C}^n$. The complex projective space CP^n , the union of complex straight lines through 0 in \mathbb{C}^{n+1} , is a compact complex manifold of dimension n . Similarly to the notion of differentiable mapping between differentiable manifolds, we have the notion of holomorphic mapping between complex manifolds.

A smooth family $m \mapsto J_m$ of endomorphisms of the tangent spaces $T_m M$ to a differentiable manifold M such that $J_m^2 = -\text{Id}$ gives rise to an almost-complex manifold. The prototype is the almost-complex structure on \mathbb{C}^n defined by $J(\partial_{x_i}) = \partial_{y_i}$; $J(\partial_{y_i}) = -\partial_{x_i}$ with $z = (x_1 + iy_1, \dots, x_n + iy_n) \in \mathbb{C}^n$ which can be transferred to a complex manifold M by means of local charts. An almost-complex structure J on a manifold M is called complex if M is the underlying differentiable manifold of a complex manifold which induces J in this way.

Studying smooth functions on a differentiable manifold can provide information on the topology of the manifold: for example, the behavior of a smooth function on a compact manifold as its critical points strongly restricted by the topological properties of the manifold. This leads to the Morse

critical point theory which extends to infinite-dimensional manifolds and, among other consequences, leads to conclusions on extremals or closed extremals of variational problems. Rather than privileging points on a manifold, one can study instead the geometry of manifolds from the point of view of spaces of functions, which leads to an algebraic approach to differential geometry. The initial concept there is a commutative ring (which becomes a possibly noncommutative algebra in the framework of noncommutative geometry), namely the ring of smooth functions on the manifold, while the manifold itself is defined in terms of the ring as the space of maximal ideals. In particular, this point of view proves to be fruitful to understand supermanifolds, a generalization of manifolds which is important for supersymmetric field theories.

One can further consider the sheaf of smooth functions on an open subset of the manifold; this point of view leads to sheaf theory which provides a unified approach to establishing connections between local and global properties of topological spaces.

Metric Properties

Riemann focused on the metric properties of manifolds but the first clear formulation of the concept of a manifold equipped with a metric was given by Weyl in *Die Idee der Riemannsche Fläche*. A Riemannian metric on a differentiable manifold M is a positive-definite scalar product g_m on T_mM for every point $m \in M$ depending smoothly on the point m . A manifold equipped with a Riemannian metric is called a Riemannian manifold. A Weyl transformation, which is multiplying the metric by a smooth positive function, yields a new Riemannian metric with the same angle measurement as the original one, and hence leaves the “conformal” structure on M unchanged.

Riemann also suggested considering metrics on the tangent spaces that are not induced from scalar products; metrics on the manifold built this way were first systematically investigated by Finsler and are therefore called Finsler metrics. Geodesics on a Riemannian manifold M which correspond to smooth curves $\gamma: [a, b] \rightarrow M$ that minimize the length functional

$$L(\gamma) := \frac{1}{2} \int_a^b \sqrt{g_{\gamma(t)} \left(\frac{d\gamma}{dt}, \frac{d\gamma}{dt} \right)} dt$$

then generalize to curves which realize the shortest distance between two points chosen sufficiently close.

Euclid’s axioms which naturally lead to Riemannian geometry are also satisfied up to the axiom of parallelism by a geometry developed by

Lobatchevsky in 1829 and Bolyai in 1832. Non-Euclidean geometries actually played a major role in the development of differential geometry and Lobachevsky’s work inspired Riemann and later Klein.

Dropping the positivity assumption for the bilinear forms g_m on T_mM leads to Lorentzian manifolds which are $(n+1)$ -dimensional smooth manifolds equipped with bilinear forms on the tangent spaces with signature $(1, n)$. These occur in general relativity and tangent vectors with negative, positive, or vanishing squared length are called timelike, spacelike, and lightlike, respectively.

Just as complex vector spaces can be equipped with positive-definite Hermitian products, a complex manifold M can come equipped with a Hermitian metric, namely a positive-definite Hermitian product h_m on T_mM for every point $m \in M$ depending smoothly on the point m ; every Hermitian metric induces a Riemannian one given by its real part. The complex projective space $\mathbb{C}P^n$ comes naturally equipped with the Fubini–Study Hermitian metric.

Transformation Groups

Metric properties can be seen from the point of view of transformation groups. Poncelet in his *Traité projectif des figures* (1822) had investigated classical Euclidean geometry from a projective geometric point of view, but it was not until Cayley (1858) that metric properties were interpreted as those stable under any “projective” transformation which leaves “cyclic points” (points at infinity on the imaginary axis of the complex plane) invariant. Transformation groups were further investigated by Lie, leading to the modern concept of Lie group, a smooth manifold endowed with a group structure such that the group operations are smooth.

A vector field X on a Lie group G is called left- (resp. right-) invariant if it is invariant under left translations $L_g: h \mapsto gh$ (resp. right translations $R_g: h \mapsto hg$) for every $g \in G$, that is, if $(L_g)_*X(h) = X(gh) \forall (g, h) \in G^2$ (resp. $(R_g)_*X(h) = X(hg) \forall (g, h) \in G^2$). The set of all left-invariant vector fields equipped with the sum, scalar multiplication, and the bracket operation on vector fields form an algebra called the Lie algebra of G .

The group $\text{GL}_n(\mathbb{R})$ (resp. $\text{GL}_n(\mathbb{C})$) of all real (resp. complex) invertible $n \times n$ matrices is a Lie group with Lie algebra, the algebra $\mathfrak{gl}_n(\mathbb{R})$ (resp. $\mathfrak{gl}_n(\mathbb{C})$) of all real (resp. complex) $n \times n$ matrices and the bracket operation reads $[A, B] = AB - BA$.

The orthogonal (resp. unitary) group $\text{O}_n(\mathbb{R}) := \{A \in \text{GL}_n(\mathbb{R}), A^t A = 1\}$, where A^t denotes the transposed matrix (resp. $\text{U}_n(\mathbb{C}) := \{A \in \text{GL}_n(\mathbb{C}), A^* A = 1\}$, where $A^* = \bar{A}^t$), is a compact Lie group with Lie

algebra $\mathfrak{o}_n(\mathbb{R}) := \{A \in \text{Gl}_n(\mathbb{R}), A^t = -A\}$ (resp. $\mathfrak{u}_n(\mathbb{C}) := \{A \in \text{Gl}_n(\mathbb{C}), A^* = -A\}$).

A left-invariant vector field X on a finite-dimensional Lie group G (or equivalently an element X of the Lie algebra of G) generates a global one-parameter group of transformations $\phi_X(t), t \in \mathbb{R}$. The mapping from the Lie algebra of G into G defined by $\exp(X) := \phi_X(1)$ is called the exponential mapping. The exponential mapping on $\text{Gl}_n(\mathbb{R})$ (resp. $\text{Gl}_n(\mathbb{C})$) is given by the series $\exp(A) = \sum_{i=0}^{\infty} A^i / i!$.

As symmetry groups of physical systems, Lie groups play an important role in physics, in particular in quantum mechanics and Yang–Mills theory. Infinite-dimensional Lie groups arise as symmetry groups, such as the group of diffeomorphisms of a manifold in general relativity, the group of gauge transformations in Yang–Mills theory, and the group of Weyl transformations of metrics on a surface in string theory. The principle “the physics should not depend on how it is described” translates to an invariance under the action of the (possibly infinite-dimensional group) of symmetries of the theory. Anomalies arise when such an invariance holds for the classical action of a physical theory but “breaks” at the quantized level.

In his Erlangen program (1872), Klein puts the concept of transformation group in the foreground introducing a novel idea by which one should consider a space endowed with some properties as a set of objects invariant under a given group of transformations. One thereby reaches a classification of geometric results according to which group is relevant in a particular problem as, for example, the projective linear group for projective geometry, the orthogonal group for Riemannian geometry, or the symplectic group for “symplectic” geometry.

Fiber Bundles

Transformation groups give rise to principal fiber bundles which play a major role in Yang–Mills theory. The notion of fiber bundle first arose out of questions posed in the 1930s on the topology and the geometry of manifolds, and by 1950 the definition of fiber bundle had been clearly formulated by Steenrod.

A smooth fiber bundle with typical fiber a manifold F is a triple (E, π, B) , where E and B are smooth manifolds called the total space and the base space, and $\pi: E \rightarrow B$ is a smooth surjective map called the projection of the bundle such that the preimage $\pi^{-1}(b)$ of a point $b \in B$ called the fiber of the bundle over b is isomorphic to F and any base point b has a neighborhood $U \subset B$ with preimage $\pi^{-1}(U)$ diffeomorphic to $U \times F$, where the diffeomorphisms commute with the projection on the base

space. Smooth sections of E are maps $\sigma: B \rightarrow E$ such that $\pi \circ \sigma = I_B$.

When F is a vector space and when, given open subsets $U_i \subset B$ that cover B with corresponding coordinate charts $(U_i, \phi_i)_{i \in I}$, the local diffeomorphisms $\tau_i: \pi^{-1}(U_i) \simeq \phi_i(U_i) \times F$ give rise to transition maps $\tau_i \circ \tau_j^{-1}: \phi_j(U_i \cap U_j) \times F \rightarrow \phi_i(U_i \cap U_j) \times F$ that are linear in the fiber, the bundle is called a “vector bundle.” The tangent bundle $TM = \bigcup_{m \in M} T_m M$ to a differentiable manifold M modeled on a vector space V is a vector bundle with typical fiber V and transition maps $\tau_{ij} = (\phi_i \circ \phi_j^{-1}, d(\phi_i \circ \phi_j^{-1}))$ expressed in terms of the differentials of the transition maps on the manifold M . So are the cotangent bundle, the dual of the tangent bundle, and tensor products of the tangent and cotangent vector bundles with typical fiber the dual V^* and tensor products of V and V^* . Vector fields defined previously are sections of the tangent bundle, 1-forms on M are sections of the cotangent bundle, and contravariant tensors, resp. covariant tensors are sections of tensor products of the tangent, resp. cotangent bundles. A differentiable mapping $\phi: M \rightarrow N$ takes covariant p -tensor fields on N to their pullbacks by ϕ , covariant p -tensors on M given by

$$(\phi^* T)(X_1, \dots, X_p) := T(\phi_* X_1, \dots, \phi_* X_p)$$

for any vector fields X_1, \dots, X_p on M .

Differentiating a smooth function f on M gives rise to a 1-form df on M . More generally, exterior p -forms are antisymmetric smooth covariant p -tensors so that $\omega(X_{\sigma(1)}, \dots, X_{\sigma(p)}) = \epsilon(\sigma)\omega(X_1, \dots, X_p)$ for any vector fields X_1, \dots, X_p on M and any permutation $\sigma \in \Sigma_p$ with signature $\epsilon(\sigma)$.

Riemannian metrics are covariant 2-tensors and the space of Riemannian metrics on a manifold M is an infinite-dimensional manifold which arises as a configuration space in string theory and general relativity.

A principal bundle is a fiber bundle (P, π, B) with typical fiber a Lie group G acting freely and properly on the total space P via a right action $(p, g) \in P \times G \mapsto pg = R_g(p) \in P$ and such that the local diffeomorphisms $\pi^{-1}(U) \simeq U \times G$ are G -equivariant. Given a principal fiber bundle (P, π, B) with structure group a finite-dimensional Lie group G , the action of G on P induces a homomorphism which to an element X of the Lie algebra of G assigns a vector field X^* on P called the “fundamental vector field” generated by X . It is defined at $p \in P$ by

$$X^*(p) := \frac{d}{dt}\bigg|_{t=0} R_{\exp(tX)}(p)$$

where \exp is the exponential map on G .

Given an action of G on a vector space V , one builds from a principal bundle with typical fiber G an associated vector bundle with typical fiber V . Principal bundles are essential in gauge theory; $U(1)$ -principal bundles arise in electro-magnetism and nonabelian structure groups arise in Yang–Mills theory. There the fields are connections on the principal bundle, and the action of gauge transformations on (irreducible) connections gives rise to an infinite-dimensional principal bundle over the moduli space with structure group given by gauge transformations. Infinite-dimensional bundles arise in other field theories such as string theory where the moduli space corresponds to inequivalent complex structures on a Riemann surface and the infinite-dimensional structure group is built up from Weyl transformations of the metric and diffeomorphisms of the surface.

Connections

On a manifold there is no canonical method to identify tangent spaces at different points. Such an identification, which is needed in order to differentiate vector fields, can be achieved on a Riemannian manifold via “parallel transport” of the vector fields. The basic concepts of the theory of covariant differentiation on a Riemannian manifold were given at the end of the nineteenth century by Ricci and, in a more complete form, in 1901 in collaboration with Levi-Civita in *Méthodes de calcul différentiel absolu et leurs applications*; on a Riemannian manifold, it is possible to define in a canonical manner a parallel displacement of tangent vectors and thereby to differentiate vector field covariantly using the since then called Levi-Civita connection.

More generally, a (linear) connection (or equivalently a covariant derivation) on a vector bundle E over a manifold M provides a way to identify fibers of the vector bundle at different points; it is a map ∇ taking sections σ of E to E -valued 1-forms on M which satisfies a Leibniz rule, $\nabla(f\sigma) = df\sigma + f\nabla\sigma$, for any smooth function f on M . When E is the tangent bundle over M , curves γ on the manifold with covariantly constant velocity $\nabla\dot{\gamma}(t) = 0$ give rise to geodesics. Given an initial velocity $\dot{\gamma}(0) = X \in T_mM$ and provided X has small enough norm, $\gamma_X(1)$ defines a point on the corresponding geodesic and the map $\exp: X \mapsto \gamma_X(1)$ a diffeomorphism from a neighborhood of 0 in T_mM to a neighborhood of $m \in M$ called the “exponential map” of ∇ .

The concept of connection extends to principal bundles where it was developed by Ehresmann building on the work of Cartan. A connection on a principal bundle (P, π, B) with structure group G , which is a smooth equivariant (under the action of

the group G) decomposition of the tangent space $T_pP = H_pP \oplus V_pP$ at each point p into a horizontal space H_pP and the vertical space $V_pP = \text{Ker } d\pi_p$, gives rise to a linear connection on the associated vector bundle.

A connection on P gives rise to a 1-form ω on P with values in the Lie algebra of the structure group G called the connection 1-form and defined as follows. For each $X \in T_pP$, $\omega(X)$ is the unique element U of the Lie algebra of G such that the corresponding fundamental vector field $U^*(p)$ at point p coincides with the vertical component of X . In particular, $\omega(U^*) = U$ for any element U of the Lie algebra of G .

The space of connections which is an infinite-dimensional manifold arises as a configuration space in Yang–Mills theory and also comes into play in the Seiberg–Witten theory.

Geometric Differential Operators

From connections one defines a number of differential operators on a Riemannian manifold, among them second-order Laplacians. In particular, the Laplace–Beltrami operator $f \mapsto -\text{tr}(\nabla^{T^*M} df)$ on smooth functions, where ∇^{T^*M} is the connection on the cotangent bundle induced by the Levi-Civita connection on M , generalizes the ordinary Laplace operator on Euclidean space. This in turn generalizes to second-order operators $\Delta^E := -\text{tr}(\nabla^{T^*M \otimes E} \nabla^E)$ acting on smooth sections of a vector bundle E over a Riemannian manifold M , where ∇^E is a connection on E and $\nabla^{T^*M \otimes E}$ the connection on $T^*M \otimes E$ induced by ∇^E and the Levi-Civita connection on M .

The Dirac operator on a spin Riemannian manifold, a first-order differential operator whose square coincides with the Laplace–Beltrami operator up to zeroth-order terms, can be best understood going back to the initial idea of Dirac. A first-order differential operator with constant matrix coefficients $\sum_{i=1}^n \gamma_i(\partial/\partial x_i)$ has square given by the Laplace operator $-\sum_{i=1}^n \partial^2/\partial x_i^2$ on \mathbb{R}^n if and only if its coefficients satisfy the Clifford relations

$$\begin{aligned} \gamma_i^2 &= -1 \quad \forall i = 1, \dots, n \\ \gamma_i \gamma_j + \gamma_j \gamma_i &= 0 \quad \forall i \neq j \end{aligned}$$

The resulting Clifford algebra, once complexified, is isomorphic in even dimensions $n = 2k$ to the space $\text{End}(S_n)$ (and $\text{End}(S_n) \oplus \text{End}(S_n)$ in odd dimensions $n = 2k + 1$) of endomorphisms of the space $S_n = \mathbb{C}^{2^k}$ of complex n -spinors. When instead of the canonical metric on \mathbb{R}^n one starts from the metric on the tangent bundle TM induced by the Riemannian

metric on M and provided the corresponding spinor spaces patch up to a “spinor bundle” over M , M is called a spin manifold. The Dirac operator on a spin Riemannian manifold M is a first-order differential operator acting on spinors given by $D_g = \sum_{i=1}^n \gamma_i \nabla_{e_i}$, where ∇ is the connection on spinors (sections of the spinor bundle S) induced by the Levi-Civita connection and e_1, \dots, e_n is an orthonormal frame of the tangent bundle TM . This is a particular case of more general twisted Dirac operators D_g^W on a twisted spinor bundle $S \otimes W$ equipped with the connection $\nabla^{S \otimes W}$ which combines the connection ∇ with a connection ∇^W on an auxiliary vector bundle W . Their square $(D_g^W)^2$ relates to the Laplacian $\Delta^{S \otimes W}$ built from this twisted connection via the Lichnerowicz formula which is useful for estimates on the spectrum of the Dirac operator in terms of the underlying geometric data.

When there is no spin structure on M , one can still hope for a Spin^c structure and a Dirac D^c operator associated with a connection compatible with that structure. In particular, every compact orientable 4-manifold can be equipped with a Spin^c structure and one can build invariants of the differentiable manifold called Seiberg–Witten invariants from solutions of a system of two partial differential equations, one of which is the Dirac equation $D^c \Phi = 0$ associated with a connection compatible with the Spin^c structure and the other a nonlinear equation involving the curvature.

Curvature

The concept of “curvature,” which is now understood in terms of connections (the curvature of a connection ∇ is defined by $\Omega = \nabla^2$), historically arose prior to that of connection. In its modern form, the concept of curvature dates back to Gauss. Using a spherical representation of surfaces – the Gauss map ν , which sends a point m of an oriented surface $\Sigma \subset \mathbb{R}^3$ to the outward pointing unit normal vector ν_m – Gauss defined what is since then called the Gaussian curvature K_m at point $m \in U \subset \Sigma$ as the limit when the area of U tends to zero of the ratio $\text{area}(\nu(U))/\text{area}(U)$. It measures the obstruction to finding a distance-preserving map from a piece of the surface around m to a region in the standard plane. Gauss’ *Teorema Egregium* says that the Gaussian curvature of a smooth surface in \mathbb{R}^3 is defined in terms of the metric on the surface so that it agrees for two isometric surfaces.

From the curvature Ω of a connection on a Riemannian manifold (M, g) , one builds the

Riemannian curvature tensor, a 4-tensor which in local coordinates reads

$$R_{ijkl} := g \left(\Omega \left(\frac{\partial}{\partial_i}, \frac{\partial}{\partial_j} \right), \frac{\partial}{\partial_k}, \frac{\partial}{\partial_l} \right)$$

further taking a partial trace leads to the Ricci curvature given by the 2-tensor $\text{Ric}_{ij} = \sum_k R_{ikjk}$, the trace of which gives in turn the scalar curvature $R = \sum_i \text{Ric}_{ii}$. Sectional curvature at a point m in the direction of a two-dimensional plane spanned by two vectors U and V corresponds to $K(U, V) = g(\Omega(U, V)V, U)$. A manifold has constant sectional curvature whenever $K(U, V)/\|U \wedge V\|^2$ is a constant K for all linearly independent vectors U, V . A Riemannian manifold with constant sectional curvature is said to be spherical, flat, or hyperbolic type depending on whether $K > 0$, $K = 0$, or $K < 0$, respectively. One owes to Cartan the discovery of an important class of Riemannian manifolds, symmetric spaces, which contains the spheres, the Euclidean spaces, the hyperbolic spaces, and compact Lie groups. A connected Riemannian manifold M equipped at every point m with an isometry σ_m such that $\sigma_m(m) = m$ and the tangent map $T_m \sigma_m$ equals $-\text{Id}$ on the tangent space (it therefore reverses the geodesics through m) is called symmetric. CP^n equipped with the Fubini–Study metric is a symmetric space with the isometry given by the reflection with respect to a line in \mathbb{C}^{n+1} . A compact symmetric space has non-negative sectional curvature K .

Constraints on the curvature can have topological consequences. Spheres are the only simply connected manifolds with constant positive sectional curvature; if a simply connected complete Riemannian manifold of dimension > 1 has non-positive sectional curvature along every plane, then it is homeomorphic to the Euclidean space.

A manifold with Ricci curvature tensor proportional to the metric tensor is called an Einstein manifold. Since Einstein, curvature is a cornerstone of general relativity with gravitational force being interpreted in terms of curvature. For example, the vacuum Einstein equation reads $\text{Ric}_g = (1/2)R_g g$ with Ric_g the Ricci curvature of a metric g and R_g its scalar curvature. In addition, Kaluza–Klein supergravity is a unified theory modeled on a direct product of the Mikowski four-dimensional space and an Einstein manifold with positive scalar curvature.

The Ricci flow $dg(t)/dt = -2\text{Ric}_{g(t)}$, which is related with the Einstein equation in general relativity, was only fairly recently introduced in the mathematical literature. Hopes are strong to get a classification of closed 3-manifolds using the Ricci flow as an essential ingredient.

Cohomology

Differentiation of functions $f \mapsto df$ on a differentiable manifold M generalizes to exterior differentiation $\alpha \mapsto d\alpha$ of differential forms. A form α is closed whenever it is in the kernel of d and it is exact whenever it lies in the range of d . Since $d^2 = 0$, exact forms are closed.

Cartan’s structure equations $d\omega = -(1/2)[\omega, \omega] + \Omega$ relate the exterior differential of the connection 1-form ω on a principal bundle to its curvature Ω given by the exterior covariant derivative $D\omega := d\omega \circ h$, where $h: T_pP \rightarrow H_pP$ is the projection onto the horizontal space.

On a complex manifold, forms split into sums of (p, q) -forms, those with p -holomorphic and q -antiholomorphic components, and exterior differentiation splits as $d = \partial + \bar{\partial}$ into holomorphic and antiholomorphic derivatives, with $\partial^2 = \bar{\partial}^2 = 0$.

Geometric data are often expressed in terms of closedness conditions on certain differential forms. For example, a “symplectic manifold” is a manifold M equipped with a closed nondegenerate differential 2-form called the “symplectic form.” The theory of J -holomorphic curves on a manifold equipped with an almost-complex structure J has proved fruitful in building invariants on symplectic manifolds. A Kähler manifold is a complex manifold equipped with a Hermitian metric h whose imaginary part $\text{Im } h$ yields a closed $(1, 1)$ -form. The complex projective space CP^n is Kähler.

The exterior differentiation d gives rise to de Rham cohomology as $\text{Ker } d / \text{Im } d$, and de Rham’s theorem establishes an isomorphism between de Rham cohomology and the real singular cohomology of a manifold. Chern (or characteristic) classes are topological invariants associated to fiber bundles and play a crucial role in index theory. Chern–Weil theory builds representatives of these de Rham cohomology classes from a connection ∇ of the form $\text{tr}(f(\nabla^2))$, where f is some analytic function.

When the manifold is Riemannian, the Laplace–Beltrami operator on functions generalizes to differential forms in two different ways, namely to the Bochner Laplacian $\Delta^{\Lambda T^*M}$ on forms (i.e., sections of ΛT^*M), where the cotangent bundle T^*M is equipped with a connection induced by the Levi-Civita connection and to the Laplace–Beltrami operator on forms $(d + d^*)^2 = d^*d + dd^*$, where d^* is the (formal) adjoint of the exterior differential d . These are related via Weitzenböck’s formula which in the particular case of 1-forms states that the difference of those two operators is measured by the Ricci curvature.

When the manifold is compact, Hodge’s theorem asserts that the de Rham cohomology groups are

isomorphic to the space of harmonic (i.e., annihilated by the Laplace–Beltrami operator) differential forms. Thus, the dimension of the set of harmonic k -forms equals the k th Betti numbers from which one can define the Euler characteristic $\chi(M)$ of the manifold M taking their alternate sum. Hodge theory plays an important role in mirror symmetry which posits a duality between different manifolds on the geometric side and between different field theories via their correlation functions on the physics side. Calabi–Yau manifolds, which are Ricci-flat Kähler manifolds, are studied extensively in the context of duality.

Index Theory

While the Gaussian curvature is the solution to a local problem, it has strong influence on the global topology of a surface. The Gauss–Bonnet formula (1850) relates the Euler characteristic on a closed surface to the Gaussian curvature by

$$\chi(M) = \frac{1}{2\pi} \int_M K_m \, dA_m$$

where dA_m is the volume element on M . This is the first result relating curvature to global properties and can be seen as one of the starting points for index theory. It generalizes to the Chern–Gauss–Bonnet theorem (1944) on an even-dimensional closed manifold and can be interpreted as an example of the Atiyah–Singer index theorem (1963)

$$\text{ind}(D_g^W) = \int_M \hat{A}(\Omega_g) e^{-\text{tr}(\Omega^W)}$$

where g denotes a Riemannian metric on a spin manifold M , D_g^W a Dirac operator acting on sections of some twisted bundle $S \otimes W$ with S the spinor bundle on M and W an auxiliary vector bundle over M , $\text{ind}(D_g^W)$ the “index” of the Dirac operator, and Ω_g, Ω^W respectively the curvatures of the Levi-Civita connection and a connection on W , and $\hat{A}(\Omega_g)$ a particular Chern form called the \hat{A} -genus. Index theorems are useful to compute anomalies in gauge theories arising from functional quantisation of classical actions.

Given an even-dimensional closed spin manifold (M, g) and a Hermitian vector bundle W over M , the index of the associated Dirac operator D_g^W yields the so-called Atiyah map $K^0(M) \mapsto \mathbb{Z}$ defined by $W \mapsto \text{ind}(D_g^W)$, where $K^0(M)$ is the group of formal differences of stable homotopy classes of smooth vector bundles over M . This is the starting point for the noncommutative geometry approach to index theory, in which the space of smooth functions on a

manifold which arises here in a disguised form since $K^0(M) \simeq K_0(C^\infty(M))$ (which consists of formal differences of smooth homotopy classes of idempotents in the inductive limit of spaces of matrices $\mathfrak{gl}_n(C^\infty(M))$) is generalized to any noncommutative smooth algebra.

Further Reading

- Bishop R and Crittenden R (2001) *Geometry of Manifolds*. Providence, RI: AMS Chelsea Publishing.
- Chern SS, Chen WH, and Lam KS (2000) *Lectures on Differential Geometry, Series on University Mathematics*. Singapore: World Scientific.
- Choquet-Bruhat Y, de Witt-Morette C, and Dillard-Bleick M (1982) *Analysis, Manifolds and Physics*, 2nd edn. Amsterdam–New York: North Holland.
- Gallot S, Hulin D, and Lafontaine J (1993) *Riemannian Geometry, Universitext*. Berlin: Springer.
- Helgason S (2001) *Differential, Lie Groups and Symmetric Spaces*. Graduate Studies in Mathematics 36. AMS, Providence, RI.

- Husemoller D (1994) *Fibre Bundles*, 3rd edn. Graduate Texts in Mathematics 20. New York: Springer Verlag.
- Jost J (1998) *Riemannian Geometry and Geometric Analysis, Universitext*. Berlin: Springer.
- Klingenberg W (1995) *Riemannian Geometry*, 2nd edn. Berlin: de Gruyter.
- Kobayashi S and Nomizou K (1996) *Foundations of Differential Geometry I, II*. Wiley Classics Library, a Wiley-Interscience Publication. New York: Wiley.
- Lang S (1995) *Differential and Riemannian Manifolds*, 3rd edn. Graduate Texts in Mathematics, 160. New York: Springer Verlag.
- Milnor J (1997) *Topology from the Differentiate Viewpoint*. Princeton Landmarks in Mathematics. Princeton, NJ: Princeton University Press.
- Nakahara M (2003) *Geometry, Topology and Physics*, 2nd edn. Bristol: Institute of Physics.
- Spivak M (1979) *A Comprehensive Introduction to Differential Geometry*, vols. 1, 2 and 3. Publish or Perish Inc., Wilmington, Delaware.
- Sternberg S (1983) *Lectures on Differential Geometry*, 2nd edn. New York: Chelsea Publishing Co.

Introductory Article: Electromagnetism

N M J Woodhouse, University of Oxford, Oxford, UK

© 2006 Springer-Verlag. Published by Elsevier Ltd.
All rights reserved.

This article is adapted from Chapters 2 and 3 of *Special Relativity*, N M J Woodhouse, Springer-Verlag, 2002, by kind permission of the publisher.

Introduction

The modern theory of electromagnetism is built on the foundations of Maxwell's equations:

$$\operatorname{div} \mathbf{E} = \frac{\rho}{\epsilon_0} \quad [1]$$

$$\operatorname{div} \mathbf{B} = 0 \quad [2]$$

$$\operatorname{curl} \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J} \quad [3]$$

$$\operatorname{curl} \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad [4]$$

On the left-hand side are the electric and magnetic fields, \mathbf{E} and \mathbf{B} , which are vector-valued functions of position and time. On the right are the sources, the charge density ρ , which is a scalar function of position and time, and the current density \mathbf{J} . The source terms encode the distribution and velocities of charges, and the equations, together with boundary conditions at infinity, determine the fields

that they generate. From these equations, one can derive the familiar predictions of electrostatics and magnetostatics, as well as the dynamical behavior of fields and charges, in particular, the generation and propagation of electromagnetic waves – light waves.

Maxwell would not have recognized the equations in this compact vector notation – still less in the tensorial form that they take in special relativity. It is notable that although his contribution is universally acknowledged in the naming of the equations, it is rare to see references to “Maxwell's theory.” This is for a good reason. In his early studies of electromagnetism, Maxwell worked with elaborate mechanical models, which he saw as analogies rather than as literal descriptions of the underlying physical reality. In his later work, the mechanical models, in particular the mechanical properties of the “luminiferous ether” through which light waves propagate, were put forward more literally as the foundations of his electromagnetic theory. The equations survive in the modern theory, but the mechanical models with which Maxwell, Faraday, and others wrestled live on only in the survival of archaic terminology, such as “lines of force” and “magnetic flux.” The luminiferous ether evaporated with the advent of special relativity.

Maxwell's legacy is not his “theory,” but his equations: a consistent system of partial differential equations that describe the whole range of known interactions of electric and magnetic fields with

moving charges. They unify the treatment of electricity and magnetism by revealing for the first time the full duality between the electric and magnetic fields. They have been verified over an almost unimaginable variety of physical processes, from the propagation of light over cosmological distances, through the behavior of the magnetic fields of stars and the everyday applications in electrical engineering and laboratory experiments, down – in their quantum version – to the exchange of photons between individual electrons.

The history of Maxwell's equations is convoluted, with many false turns. Maxwell himself wrote down an inconsistent form of the equations, with a different sign for ρ in the first equation, in his 1865 work "A dynamical theory of the electromagnetic field." The consistent form appeared later in his *Treatise on Electricity and Magnetism* (1873); see [Chalmers \(1975\)](#).

In this article, we shall not follow the historical route to the equations. Some of the complex story of the development hinted at in the remarks above can be found in the articles by [Chalmers \(1975\)](#), [Siegel \(1985\)](#), and [Roche \(1998\)](#). Neither shall we follow the traditional pedagogic route of many textbooks in building up to the full dynamical equations through the study of basic electrical and magnetic phenomena. Instead, we shall follow a path to Maxwell's equations that is informed by knowledge of their most critical feature, invariance under Lorentz transformations. Maxwell, of course, knew nothing of this.

We shall start with a summary of basic facts about the behavior of charges in electric and magnetic fields, and then establish the full dynamical framework by considering this behavior as seen from moving frames of reference. It is impossible, of course, to do this consistently within the framework of classical ideas of space and time since Maxwell's equations are inconsistent with Galilean relativity. But it is at least possible to understand some of the key features of the equations, in particular the need for the term involving the time derivative of E , the so-called "displacement current," in the third of Maxwell's equations.

We shall begin with some remarks concerning the role of relativity in classical dynamics.

Relativity in Newtonian Dynamics

Newton's laws hold in all inertial frames. The formalism of classical mechanics is invariant under Galilean transformations and it is impossible to tell by observing the dynamical behavior of particles and other bodies whether a frame of reference is at

rest or in uniform motion. In the world of classical mechanics, therefore:

Principle of Relativity There is no absolute standard of rest; only relative motion is observable.

In his "Dialogue concerning the two chief world systems," Galileo illustrated the principle by arguing that the uniform motion of a ship on a calm sea does not affect the behavior of fish, butterflies, and other moving objects, as observed in a cabin below deck.

Relativity theory takes the principle as fundamental, as a statement about the nature of space and time as much as about the properties of the Newtonian equations of motion. But if it is to be given such universal significance, then it must apply to all of physics, and not just to Newtonian dynamics. At first this seems unproblematic – it is hard to imagine that it holds at such a basic level, but not for more complex physical interactions. Nonetheless, deep problems emerge when we try to extend it to electromagnetism since Galilean invariance conflicts with Maxwell's equations.

All appears straightforward for systems involving slow-moving charges and slowly varying electric and magnetic fields. These are governed by laws that appear to be invariant under transformations between uniformly moving frames of reference. One can imagine a modern version of Galileo's ship also carrying some magnets, batteries, semi-conductors, and other electrical components. Galilei's argument for relativity would seem just as compelling.

The problem arises when we include rapidly varying fields – in particular, when we consider the propagation of light. As [Einstein \(1905\)](#) put it, "Maxwell's electrodynamics . . . , when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena." The central difficulty is that Maxwell's equations give light, along with other electromagnetic waves, a definite velocity: in empty space, it travels with the same speed in every direction, independently of the motion of the source – a fact that is incompatible with Galilean invariance. Light traveling with speed c in one frame should have speed $c + u$ in a frame moving towards the source of the light with speed u . Thus, it should be possible for light to travel with any speed. Light that travels with speed c in a frame in which its source is at rest should have some other speed in a moving frame; so Galilean invariance would imply dependence of the velocity of light on the motion of the source.

A full resolution of the conflict can only be achieved within the special theory of relativity: here, remarkably, Maxwell's equations retain exactly

their classical form, but the transformations between the space and time coordinates of frames of reference in relative motion do not. The difference appears when the velocities involved are not insignificant when compared with the velocity of light. So long as one can ignore terms of order u^2/c^2 , Maxwell's equations are compatible with the Galilean principle of relativity.

Charges, Fields, and the Lorentz-Force Law

The basic objects in the modern form of electromagnetic theory are

- charged particles; and
- the electric and magnetic fields E and B , which are vector quantities that depend on position and time.

The charge e of a particle, which can be positive or negative, is an intrinsic quantity analogous to gravitational mass. It determines the strength of the particle's interaction with the electric and magnetic fields – as its mass determines the strength of its interaction with gravitational fields.

The interaction is in two directions. First, electric and magnetic fields exert a force on a charged particle which depends on the value of the charge, the particle's velocity, and the values of E and B at the location of the particle. The force is given by the Lorentz-force law

$$\mathbf{f} = e(\mathbf{E} + \mathbf{u} \wedge \mathbf{B}) \quad [5]$$

in which e is the charge and \mathbf{u} is the velocity. It is analogous to the gravitational force

$$\mathbf{f} = m\mathbf{g} \quad [6]$$

on a particle of mass m in a gravitational field \mathbf{g} . It is through the force law that an observer can, in principle, measure the electric and magnetic fields at a point, by measuring the force on a standard charge moving with known velocity.

Second, moving charges generate electric and magnetic fields. We shall not yet consider in detail the way in which they do this, beyond stating the following basic principles.

EM1. The fields depend linearly on the charges.

This means that if we superimpose two distributions of charge, then the resultant E and B fields are the sums of the respective fields that the two distributions generate separately.

EM2. A stationary point charge e generates an electric field, but no magnetic field. The electric field is given by

$$\mathbf{E} = \frac{k e \mathbf{r}}{r^3} \quad [7]$$

where \mathbf{r} is the position vector from the charge, $r = |\mathbf{r}|$, and k is a positive constant, analogous to the gravitational constant.

By combining [7] and [5], we obtain an inverse-square law electrostatic force

$$\frac{k e e'}{r^2} \quad [8]$$

between two stationary charges; unlike gravity, it is repulsive when the charges have the same sign.

EM3. A point charge moving with velocity \mathbf{v} generates a magnetic field

$$\mathbf{B} = \frac{k' e \mathbf{v} \wedge \mathbf{r}}{r^3} \quad [9]$$

where k' is a second positive constant.

This is extrapolated from measurements of the magnetic field generated by currents flowing in electrical circuits.

The constants k and k' in EM2 and EM3 determine the strengths of electric and magnetic interactions. They are usually denoted by

$$k = \frac{1}{4\pi\epsilon_0}, \quad k' = \frac{\mu_0}{4\pi} \quad [10]$$

Charge e is measured in coulombs, $|\mathbf{B}|$ in teslas, and $|\mathbf{E}|$ in volts per meter. With other quantities in SI units,

$$\epsilon_0 = 8.9 \times 10^{-12}, \quad \mu_0 = 1.3 \times 10^{-6} \quad [11]$$

The charge of an electron is -1.6×10^{-19} C; the current through an electric fire is a flow of $5\text{--}10$ C s⁻¹. The earth's magnetic field is about 4×10^{-5} T; a bar magnet's is about 1 T; there is a field of about 50 T on the second floor of the Clarendon Laboratory in Oxford; and the magnetic field on the surface of a neutron star is about 10^8 T.

Although we are more aware of gravity in everyday life, it is very much weaker than the electrostatic force – the electrostatic repulsion between two protons is a factor of 1.2×10^{36} greater than their gravitational attraction (at any separation, both forces obey the inverse-square law).

Our aim is to pass from EM1–EM3 to Maxwell's equations, by replacing [7] and [9] by partial differential equations that relate the field strengths to the charge and current densities ρ and \mathbf{J} of a

continuous distribution of charge. The densities are defined as the limits

$$\rho = \lim_{V \rightarrow 0} \left(\frac{\sum e}{V} \right), \quad \mathbf{J} = \lim_{V \rightarrow 0} \left(\frac{\sum e\mathbf{v}}{V} \right) \quad [12]$$

where V is a small volume containing the point, e is a charge within the volume, and \mathbf{v} is its velocity; the sums are over the charges in V and the limits are taken as the volume is shrunk (although we shall not worry too much about the precise details of the limiting process).

Stationary Distributions of Charge

We begin the task of converting the basic principles into partial differential equations by looking at the electric field of a stationary distribution of charge, where the passage to the continuous limit is made by using the Gauss theorem to restate the inverse-square law.

The Gauss theorem relates the integral of the electric field over a closed surface to the total charge contained within it. For a point charge, the electric field is given by EM2:

$$\mathbf{E} = \frac{e\mathbf{r}}{4\pi\epsilon_0 r^3}$$

Since $\text{div } \mathbf{r} = 3$ and $\text{grad } r = \mathbf{r}/r$, we have

$$\text{div}(\mathbf{E}) = \text{div} \left(\frac{e\mathbf{r}}{\pi\epsilon_0 r^3} \right) = \frac{e}{4\pi\epsilon_0} \left(\frac{3}{r^3} - \frac{3\mathbf{r} \cdot \mathbf{r}}{r^5} \right) = 0$$

everywhere except at $r=0$. Therefore, by the divergence theorem,

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = 0 \quad [13]$$

for any closed surface ∂V bounding a volume V that does not contain the charge.

What if the volume does contain the charge? Consider the region bounded by the sphere S_R of radius R centered on the charge; S_R has outward unit normal \mathbf{r}/r . Therefore,

$$\int_{S_R} \mathbf{E} \cdot d\mathbf{S} = \frac{e}{4\pi R^2 \epsilon_0} \int_{S_R} d\mathbf{S} = \frac{e}{\epsilon_0}$$

In particular, the value of the surface integral on the left-hand side does not depend on R .

Now consider arbitrary finite volume bounded by a closed surface S . If the charge is not inside the volume, then the integral of \mathbf{E} over S vanishes by [13]. If it is, then we can apply [13] to the

volume V between S and a small sphere S_R to deduce that

$$\int_S \mathbf{E} \cdot d\mathbf{S} - \int_{S_R} \mathbf{E} \cdot d\mathbf{S} = \int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = 0$$

and that the integrals of \mathbf{E} over S and S_R are the same. Therefore,

$$\int_S \mathbf{E} \cdot d\mathbf{S} = \begin{cases} e/\epsilon_0 & \text{if the charge is in} \\ & \text{the volume bounded by } S \\ 0 & \text{otherwise} \end{cases}$$

When we sum over a distribution of charges, the integral on the left picks out the total charge within S . Therefore, we have the Gauss theorem.

The Gauss theorem. For any closed surface ∂V bounding a volume V ,

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = Q/\epsilon_0$$

where \mathbf{E} is the total electric field and Q is the total charge within V .

Now we can pass to the continuous limit. Suppose that \mathbf{E} is generated by a distribution of charges with density ρ (charge per unit volume). Then by the Gauss theorem,

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} \int_V \rho dV$$

for any volume V . But then, by the divergence theorem,

$$\int_V (\text{div } \mathbf{E} - \rho/\epsilon_0) dV = 0$$

Since this holds for any volume V , it follows that

$$\text{div } \mathbf{E} = \rho/\epsilon_0 \quad [14]$$

By an argument in a similar spirit, we can also show that the electric field of a stationary distribution of charge is conservative in the sense that the total work done by the field when a charge is moved around a closed loop vanishes; that is,

$$\oint \mathbf{E} \cdot d\mathbf{s} = 0$$

for any closed path. This is equivalent to

$$\text{curl } \mathbf{E} = 0 \quad [15]$$

since, by Stokes' theorem,

$$\oint \mathbf{E} \cdot d\mathbf{s} = \int_S \text{curl } \mathbf{E} \cdot d\mathbf{S}$$

where S is any surface spanning the path. This vanishes for every path and for every S if and only if [15] holds.

The field of a single stationary charge is conservative since

$$\mathbf{E} = -\text{grad } \phi, \quad \phi = \frac{e}{4\pi\epsilon_0 r}$$

and therefore $\text{curl } \mathbf{E} = 0$ since the curl of a gradient vanishes identically. For a continuous distribution, $\mathbf{E} = -\text{grad } \phi$, where

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_{r' \in V} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \quad [16]$$

In the integral, \mathbf{r} (the position of the point at which ϕ is evaluated) is fixed, and the integration is over the positions \mathbf{r}' of the individual charges. In spite of the singularity at $\mathbf{r} = \mathbf{r}'$, the integral is well defined. So, [15] also holds for a continuous distribution of stationary charge.

The Divergence of the Magnetic Field

We can apply the same argument that established the Gauss theorem to the magnetic field of a slow-moving charge. Here,

$$\mathbf{B} = \frac{\mu_0 e \mathbf{v} \wedge \mathbf{r}}{4\pi r^3}$$

where \mathbf{r} is the vector from the charge to the point at which the field is measured. Since $\mathbf{r}/r^3 = -\text{grad}(1/r)$, we have

$$\text{div} \left(\mathbf{v} \wedge \frac{\mathbf{r}}{r^3} \right) = \mathbf{v} \wedge \text{curl} \left(\text{grad} \frac{1}{r} \right) = 0$$

Therefore, $\text{div } \mathbf{B} = 0$ except at $r = 0$, as in the case of the electric field. However, in the magnetic case, the integral of the field over a surface surrounding the charge also vanishes, since if S_R is a sphere of radius R centered on the charge, then

$$\int_{S_R} \mathbf{B} \cdot d\mathbf{S} = \frac{\mu_0 e}{4\pi} \int_{S_R} \frac{\mathbf{v} \wedge \mathbf{r}}{r^3} \cdot \frac{\mathbf{r}}{r} dS = 0$$

By the divergence theorem, the same is true for any surface surrounding the charge. We deduce that if magnetic fields are generated only by moving charges, then

$$\int_{\partial V} \mathbf{B} \cdot d\mathbf{S} = 0$$

for any volume V , and hence that

$$\text{div } \mathbf{B} = 0 \quad [17]$$

Of course, if there were free “magnetic poles” generating magnetic fields in the same way that charges generate electric fields, then this would not hold; there would be a “magnetic pole density” on

the right-hand side, by analogy with the charge density in [14].

Inconsistency with Galilean Relativity

Our central concern is the compatibility of the laws of electromagnetism with the principle of relativity. As Einstein observed, simple electromagnetic interactions do indeed depend only on relative motion; the current induced in a conductor moving through the field of a magnet is the same as that generated in a stationary conductor when a magnet is moved past it with the same relative velocity (Einstein 1905). Unfortunately, this symmetry is not reflected in our basic principles. We very quickly come up against contradictions if we assume that they hold in every inertial frame of reference.

One emerges as follows. An observer O can measure the values of \mathbf{B} and \mathbf{E} at a point by measuring the force on a particle of standard charge, which is related to the velocity \mathbf{v} of the charge by the Lorentz-force law,

$$\mathbf{f} = e(\mathbf{E} + \mathbf{v} \wedge \mathbf{B})$$

A second observer O' moving relative to the first with velocity \mathbf{v} will see the same force, but now acting on a particle at rest. He will therefore measure the electric field to be $\mathbf{E}' = \mathbf{f}/e$. We conclude that an observer moving with velocity \mathbf{v} through a magnetic field \mathbf{B} and an electric field \mathbf{E} should see an electric field

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \wedge \mathbf{B} \quad [18]$$

By interchanging the roles of the two observers, we should also have

$$\mathbf{E} = \mathbf{E}' - \mathbf{v} \wedge \mathbf{B}' \quad [19]$$

where \mathbf{B}' is the magnetic field measured by the second observer. If both are to hold, then $\mathbf{B} - \mathbf{B}'$ must be a scalar multiple of \mathbf{v} .

But this is incompatible with EM3; if the fields are those of a point charge at rest relative to the first observer, then \mathbf{E} is given by [7], and

$$\mathbf{B} = 0$$

On the other hand, the second observer sees the field of a point charge moving with velocity $-\mathbf{v}$. Therefore,

$$\mathbf{B}' = -\frac{\mu_0 e \mathbf{v} \wedge \mathbf{r}}{4\pi r^3}$$

So $\mathbf{B} - \mathbf{B}'$ is orthogonal to \mathbf{v} , not parallel to it.

This conspicuous paradox is resolved, in part, by the realization that EM3 is not exact; it holds only when the velocities are small enough for the magnetic force between two particles to be negligible in comparison with the electrostatic force. If v is a typical velocity, then the condition is that $v^2 \mu_0$

should be much less than $1/\epsilon_0$. That is, the velocities involved should be much less than

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3 \times 10^8 \text{ m s}^{-1}$$

This, of course, is the velocity of light.

The Limits of Galilean Invariance

Our basic principles EM1–EM3 must now be seen to be approximations – they describe the interactions of particles and fields when the particles are moving relative to each other at speeds much less than that of light. To emphasize that we cannot expect, in particular, EM3 to hold for particles moving at speeds comparable with c , we must replace it by

EM3'. A charge moving with velocity \mathbf{v} , where $v \ll c$, generates a magnetic field

$$\mathbf{B} = \frac{\mu_0 e \mathbf{v} \wedge \mathbf{r}}{4\pi r^3} + O(v^2/c^2) \quad [20]$$

The magnetic field of a system of charges in general motion satisfies

$$\text{div } \mathbf{B} = 0 \quad [21]$$

In the second part, we have retained [21] as a differential form of the statement that there are no free magnetic poles; the magnetic field is generated only by the motion of the charges. With this change, the theory is consistent with the principle of relativity, provided that we ignore terms of order v^2/c^2 . The substitution of EM3' for EM3 resolves the conspicuous paradox; the symmetry noted by Einstein between the current generated by the motion of the conductor in a magnetic field and by the motion of a magnet past a conductor is explained, provided that the velocities are much less than that of light.

The central problem remains however; the equations of electromagnetism are not invariant under a Galilean transformation with velocity comparable to c . The paradox is still there, but it is more subtle than it appeared to be at first. There are three possible ways out: (1) the noninvariance is real and has observable effects (necessarily of order v^2/c^2 or smaller); (2) Maxwell's theory is wrong; or (3) the Galilean transformation is wrong. Disconcertingly, it is the last path that physics has taken. But that is to jump ahead in the story. Our task is to complete the derivation of Maxwell's equations.

Faraday's Law of Induction

The magnetic field of a slow-moving charge will always be small in relation to its electric field (even

when we replace \mathbf{B} by $c\mathbf{B}$ to put it into the same units as \mathbf{E}). The magnetic fields generated by currents in electrical circuits are not, however, dominated by large electric fields. This is because the currents are created by the flow, at slow velocity, of electrons, while overall the matter in the wire is roughly electrically neutral, with the electric fields of the positively charged nuclei and negatively charged electrons canceling.

This is the physical context to keep in mind in the following deduction of Faraday's law of induction from Galilean invariance for velocities much less than c . The law relates the electromotive force or "voltage" around an electrical circuit to the rate of change of the magnetic field \mathbf{B} over a surface spanning the circuit. In its differential form, the law becomes one of Maxwell's equations.

Suppose first that the fields are generated by charges all moving relative to a given inertial frame of reference R with the same velocity \mathbf{v} . Then in a second frame R' moving relative to R with velocity \mathbf{v} , there is a stationary distribution of charge. If the velocity is much less than that of light, then the electric field \mathbf{E}' measured in R' is related to the electric and magnetic \mathbf{E} and \mathbf{B} measured in R by

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \wedge \mathbf{B}$$

Since the field measured in R' is that of a stationary distribution of charge, we have

$$\text{curl } \mathbf{E}' = 0$$

In R , the charges are all moving with velocity \mathbf{v} , so their configuration looks exactly the same from the point \mathbf{r} at time t as it does from the point $\mathbf{r} + \mathbf{v}\tau$ at time $t + \tau$. Therefore,

$$\mathbf{B}(\mathbf{r} + \mathbf{v}\tau, t + \tau) = \mathbf{B}(\mathbf{r}, t)$$

$$\mathbf{E}(\mathbf{r} + \mathbf{v}\tau, t + \tau) = \mathbf{E}(\mathbf{r}, t)$$

and hence by taking derivatives with respect to τ at $\tau = 0$,

$$\begin{aligned} \mathbf{v} \cdot \text{grad } \mathbf{B} + \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \mathbf{v} \cdot \text{grad } \mathbf{E} + \frac{\partial \mathbf{E}}{\partial t} &= 0 \end{aligned} \quad [22]$$

So we must have

$$\begin{aligned} 0 &= \text{curl } \mathbf{E}' \\ &= \text{curl } \mathbf{E} + \text{curl}(\mathbf{v} \wedge \mathbf{B}) \\ &= \text{curl } \mathbf{E} + \mathbf{v} \text{div } \mathbf{B} - \mathbf{v} \cdot \text{grad } \mathbf{B} \\ &= \text{curl } \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} \end{aligned} \quad [23]$$

since $\text{div } \mathbf{B} = 0$. It follows that

$$\text{curl } \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad [24]$$

Equation [24] is linear in \mathbf{B} and \mathbf{E} ; so by adding the magnetic and electric fields of different streams of charges moving relative to R with different velocities, we deduce that it holds generally for the electric and magnetic fields generated by moving charges.

Equation [24] encodes Faraday's law of electromagnetic induction, which describes how changing magnetic fields can generate currents. In the static case

$$\frac{\partial \mathbf{B}}{\partial t} = 0$$

and the equation reduces to $\text{curl } \mathbf{E} = 0$ – the condition that the electrostatic field should be conservative; that is, it should do no net work when a charge is moved around a closed loop.

More generally, consider a wire loop in the shape of a closed curve γ . Let S be a fixed surface spanning γ . Then we can deduce from eqn [24] that

$$\begin{aligned} \oint_{\gamma} \mathbf{E} \cdot d\mathbf{s} &= \int_S \text{curl } \mathbf{E} \cdot d\mathbf{S} \\ &= - \int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \\ &= - \frac{d}{dt} \int_S (\mathbf{B} \cdot d\mathbf{S}) \end{aligned} \quad [25]$$

If the magnetic field is varying, so that the integral of \mathbf{B} over S is not constant, then the integral of \mathbf{E} around the loop will not be zero. There will be a nonzero electric field along the wire, which will exert a force on the electrons in the wire and cause a current to flow.

The quantity

$$\oint \mathbf{E} \cdot d\mathbf{s}$$

which is measured in volts, is the work done by the electric field when a unit charge makes one circuit of the wire. It is called the electromotive force around the circuit. The integral is the magnetic flux linking the circuit. The relationship [25] between electromotive force and rate of change of magnetic flux is Faraday's law.

The Field of Charges in Uniform Motion

We can extract another of Maxwell's equations from this argument. By EM3', a single charge e with velocity \mathbf{v} generates an electric field \mathbf{E} and a magnetic field

$$\mathbf{B} = \frac{\mu_0 e \mathbf{v} \wedge \mathbf{r}}{4\pi r^3} + O(v^2/c^2)$$

where \mathbf{r} is the vector from the charge to the point at which the field is measured. In the frame of reference R' in which the charge is at rest, its electric field is

$$\mathbf{E}' = \frac{e\mathbf{r}}{4\pi\epsilon_0 r^3}$$

In the frame in which it is moving with velocity \mathbf{v} , $\mathbf{E} = \mathbf{E}' + O(v/c)$. Therefore,

$$c\mathbf{B} = \frac{\mathbf{v} \wedge \mathbf{E}'}{c} = \frac{\mathbf{v} \wedge \mathbf{E}}{c} + O\left(\frac{v^2}{c^2}\right)$$

By taking the curl of both sides, and dropping terms of order v^2/c^2 ,

$$\begin{aligned} \text{curl}(c\mathbf{B}) &= \text{curl}\left(\frac{\mathbf{v} \wedge \mathbf{E}}{c}\right) \\ &= \frac{1}{c}(\mathbf{v} \text{ div } \mathbf{E} - \mathbf{v} \cdot \text{grad } \mathbf{E}) \end{aligned}$$

But

$$\text{div } \mathbf{E} = \rho/\epsilon_0, \quad \mathbf{v} \cdot \text{grad } \mathbf{E} = -\frac{\partial \mathbf{E}}{\partial t}$$

by [22]. Therefore,

$$\text{curl}(c\mathbf{B}) - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = \frac{1}{c\epsilon_0} \mathbf{J} = c\mu_0 \mathbf{J}$$

where $\mathbf{J} = \rho\mathbf{v}$. By summing over the separate particle velocities, we conclude that

$$\text{curl } \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}$$

holds for an arbitrary distribution of charges, provided that their velocities are much less than that of light.

Maxwell's Equations

The basic principles, together with the assumption of Galilean invariance for velocities much less than that of light, have allowed us to deduce that the electric and magnetic fields generated by a continuous distribution of moving charges in otherwise empty space satisfy

$$\text{div } \mathbf{E} = \frac{\rho}{\epsilon_0} \quad [26]$$

$$\text{div } \mathbf{B} = 0 \quad [27]$$

$$\text{curl } \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J} \quad [28]$$

$$\text{curl } \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad [29]$$

where ρ is the charge density, \mathbf{J} is the current density, and $c^2 = 1/\epsilon_0\mu_0$. These are Maxwell's equations, the basis of modern electrodynamics. Together with the Lorentz-force law, they describe the dynamics of charges and electromagnetic fields.

We have arrived at them by considering how basic electromagnetic processes appear in moving frames of reference – an unsatisfactory route because we have seen on the way that the principles on which we based the derivation are incompatible with Galilean invariance for velocities comparable with that of light. Maxwell derived them by analyzing an elaborate mechanical model of electric and magnetic fields – as displacements in the luminiferous ether. That is also unsatisfactory because the model has long been abandoned. The reason that they are accepted today as the basis of theoretical and practical applications of electromagnetism has little to do with either argument. It is first that they are self-consistent, and second that they describe the behavior of real fields with unreasonable accuracy.

The Continuity Equation

It is not immediately obvious that the equations are self-consistent. Given ρ and \mathbf{J} as functions of the coordinates and time, Maxwell's equations are two scalar and two vector equations in the unknown components of \mathbf{E} and \mathbf{B} . That is, a total of eight equations for six unknowns – more equations than unknowns. Therefore, it is possible that they are in fact inconsistent.

If we take the divergence of eqn [29], then we obtain

$$\frac{\partial}{\partial t}(\text{div } \mathbf{B}) = 0$$

which is consistent with eqn [27]; so no problem arises here. However, by taking the divergence of eqn [28] and substituting from eqn [26], we get

$$\begin{aligned} 0 &= \text{div } \text{curl } \mathbf{B} \\ &= \frac{1}{c^2} \frac{\partial}{\partial t}(\text{div } \mathbf{E}) + \mu_0 \text{div } \mathbf{J} \\ &= \mu_0 \left(\frac{\partial \rho}{\partial t} + \text{div } \mathbf{J} \right) \end{aligned}$$

This gives a contradiction unless

$$\frac{\partial \rho}{\partial t} + \text{div } \mathbf{J} = 0 \quad [30]$$

So the choice of ρ and \mathbf{J} is not unconstrained; they must be related by the continuity equation [30]. This holds for physically reasonable distributions of

charge; it is a differential form of the statement that charges are neither created nor destroyed.

Conservation of Charge

To see the connection between the continuity equation and charge conservation, let us look at the total charge within a fixed V bounded by a surface S . If charge is conserved, then any increase or decrease in a short period of time must be exactly balanced by an inflow or outflow of charge across S .

Consider a small element dS of S with outward unit normal and consider all the particles that have a particular charge e and a particular velocity \mathbf{v} at time t . Suppose that there are σ of these per unit volume (σ is a function of position). Those that cross the surface element between t and $t + \delta t$ are those that at time t lie in the region of volume

$$|\mathbf{v} \cdot \mathbf{n} dS \delta t|$$

shown in Figure 1. They contribute $e\sigma\mathbf{v} \cdot dS\delta t$ to the outflow of charge through the surface element. But the value of \mathbf{J} at the surface element is the sum of $e\sigma\mathbf{v}$ over all possible values of \mathbf{v} and e . By summing over \mathbf{v} , e , and the elements of the surface, therefore, and by passing to the limit of a continuous distribution, the total rate of outflow is

$$\int_S \mathbf{J} \cdot d\mathbf{S}$$

Charge conservation implies that the rate of outflow should be equal to the rate of decrease in the total charge within V . That is,

$$\frac{d}{dt} \int_V \rho dV + \int_S \mathbf{J} \cdot d\mathbf{S} = 0 \quad [31]$$

By differentiating the first term under the integral sign and by applying the divergence theorem to the second integral,

$$\int_V \left(\frac{\partial \rho}{\partial t} + \text{div } \mathbf{J} \right) dV = 0 \quad [32]$$

If this is to hold for any choice of V , then ρ and \mathbf{J} must satisfy the continuity equation. Conversely, the continuity equation implies charge conservation.

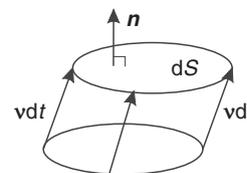


Figure 1 The outflow through a surface element.

The Displacement Current

The third of Maxwell's equations can be written as

$$\text{curl } \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \quad [33]$$

in which form it can be read as an equation for an unknown magnetic field \mathbf{B} in terms of a known current distribution \mathbf{J} and electric field \mathbf{E} . When \mathbf{E} and \mathbf{J} are independent of t , it reduces to

$$\text{curl } \mathbf{B} = \mu_0 \mathbf{J}$$

which determines the magnetic field of a steady current, in a way that was already familiar to Maxwell's contemporaries. But his second term on the right-hand side of [33] was new; it adds to \mathbf{J} the so-called vacuum displacement current

$$\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

The name comes from an analogy with the behavior of charges in an insulating material. Here no steady current can flow, but the distribution of charges within the material is distorted by an external electric field. When the field changes, the distortion also changes, and the result appears as a current – the displacement current – which flows during the period of change. Maxwell's central insight was that the same term should be present even in empty space. The consequence was profound; it allowed him to explain the propagation of light as an electromagnetic phenomenon.

The Source-Free Equations

In a region of empty space, away from the charges generating the electric and magnetic fields, we have $\rho=0=\mathbf{J}$, and Maxwell's equations reduce to

$$\text{div } \mathbf{E} = 0 \quad [34]$$

$$\text{div } \mathbf{B} = 0 \quad [35]$$

$$\text{curl } \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = 0 \quad [36]$$

$$\text{curl } \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad [37]$$

where $c = 1/\sqrt{\epsilon_0 \mu_0}$. By taking the curl of eqn [36] and by substituting from eqns [35] and [37], we obtain

$$\begin{aligned} 0 &= \text{grad}(\text{div } \mathbf{B}) - \nabla^2 \mathbf{B} - \frac{1}{c^2} \text{curl} \left(\frac{\partial \mathbf{E}}{\partial t} \right) \\ &= -\nabla^2 \mathbf{B} - \frac{1}{c^2} \frac{\partial}{\partial t} (\text{curl } \mathbf{E}) \\ &= -\nabla^2 \mathbf{B} + \frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} \end{aligned} \quad [38]$$

Therefore, the three components of \mathbf{B} in empty space satisfy the (scalar) wave equation

$$\square u = 0$$

Here \square is the d'Alembertian operator, defined by

$$\square = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$$

By taking the curl of eqn [37], we also obtain $\square \mathbf{E} = 0$.

Monochromatic Plane Waves

The fact that \mathbf{E} and \mathbf{B} are vector-valued solutions of the wave equation in empty space suggests that we look for “plane wave” solutions of Maxwell's equations in which

$$\mathbf{E} = \boldsymbol{\alpha} \cos \Omega + \boldsymbol{\beta} \sin \Omega \quad [39]$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are constant vectors and

$$\Omega = \frac{\omega}{c} (ct - \mathbf{r} \cdot \mathbf{e}), \quad \mathbf{e} \cdot \mathbf{e} = 1 \quad [40]$$

with $\omega > 0$, $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and \mathbf{e} constant; ω is the frequency and \mathbf{e} is a unit vector that gives the direction of propagation (adding τ to t and $c\tau\mathbf{e}$ to \mathbf{r} leaves Ω unchanged). This satisfies the wave equation, but for a general choice of the constants, it will not be possible to find \mathbf{B} such that eqns [34]–[37] also hold.

By taking the divergence of eqn [39], we obtain

$$\text{div } \mathbf{E} = \frac{\omega}{c} (\mathbf{e} \cdot \boldsymbol{\alpha} \sin \Omega - \mathbf{e} \cdot \boldsymbol{\beta} \cos \Omega) \quad [41]$$

For eqn [34] to hold, therefore, we must choose $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ orthogonal to \mathbf{e} . For eqn [37] to hold, we must find \mathbf{B} such that

$$\text{curl } \mathbf{E} = \frac{\omega}{c} (\mathbf{e} \wedge \boldsymbol{\alpha} \sin \Omega - \mathbf{e} \wedge \boldsymbol{\beta} \cos \Omega) = -\frac{\partial \mathbf{B}}{\partial t} \quad [42]$$

A possible choice is

$$\mathbf{B} = \frac{\mathbf{e} \wedge \mathbf{E}}{c} = \frac{1}{c} (\mathbf{e} \wedge \boldsymbol{\alpha} \cos \Omega + \mathbf{e} \wedge \boldsymbol{\beta} \sin \Omega) \quad [43]$$

and it is not hard to see that \mathbf{E} and \mathbf{B} then satisfy [35] and [36] as well.

The solutions obtained in this way are called “monochromatic electromagnetic plane waves.”

Note that such waves are transverse in the sense that \mathbf{E} and \mathbf{B} are orthogonal to the direction of propagation. The definition \mathbf{E} can be written more concisely in the form

$$\mathbf{E} = \text{Re}[(\boldsymbol{\alpha} + i\boldsymbol{\beta})e^{-i\Omega t}] \quad [44]$$

It is an exercise in Fourier analysis to show every solution in empty space is a combination of monochromatic plane waves. A plane wave has “plane” or “linear” polarization if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are proportional. It has “circular” polarization if $\boldsymbol{\alpha} \cdot \boldsymbol{\alpha} = \boldsymbol{\beta} \cdot \boldsymbol{\beta}, \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 0$.

At the heart of Maxwell’s theory was the idea that a light wave with definite frequency or color is represented by a monochromatic plane solution of his equations.

Potentials

For every solution of Maxwell’s equations *in vacuo*, the components of \mathbf{E} and \mathbf{B} satisfy the three-dimensional wave equation; but the converse is not true. That is, it is not true in general that if

$$\square \mathbf{B} = 0, \quad \square \mathbf{E} = 0$$

then \mathbf{E} and \mathbf{B} satisfy Maxwell’s equations. For this to happen, the divergence of both fields must vanish, and they must be related by [36] and [37]. These additional constraints are somewhat simpler to handle if we work not with the fields themselves, but with auxiliary quantities called “potentials.”

The definition of the potentials depends on standard integrability conditions from vector calculus. Suppose that \mathbf{v} is a vector field, which may depend on time. If $\text{curl } \mathbf{v} = 0$, then there exists a function ϕ such that

$$\mathbf{v} = \text{grad } \phi \quad [45]$$

If $\text{div } \mathbf{v} = 0$, then there exists a second vector field \mathbf{a} such that

$$\mathbf{v} = \text{curl } \mathbf{a} \quad [46]$$

Neither ϕ nor \mathbf{a} is uniquely determined by \mathbf{v} . In the first case, if [45] holds, then it also holds when ϕ is replaced by $\phi' = \phi + f$, where f is a function of time alone; in the second, if [46] holds, then it also holds when \mathbf{a} is replaced by

$$\mathbf{a}' = \mathbf{a} + \text{grad } u$$

for any scalar function u of position and time. It should be kept in mind that the existence statements are local. If \mathbf{v} is defined on a region U with

nontrivial topology, then it may not be possible to find a suitable ϕ or \mathbf{a} throughout the whole of U .

Suppose now that we are given fields \mathbf{E} and \mathbf{B} satisfying Maxwell’s equations [26]–[29] with sources represented by the charge density ρ and the current density \mathbf{J} . Since $\text{div } \mathbf{B} = 0$, there exists a time-dependent vector field $\mathbf{A}(t, x, y, z)$ such that

$$\mathbf{B} = \text{curl } \mathbf{A}$$

If we substitute $\mathbf{B} = \text{curl } \mathbf{A}$ into [29] and interchange curl with the time derivative, then we obtain

$$\text{curl} \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0$$

It follows that there exists a scalar $\phi(t, x, y, z)$ such that

$$\mathbf{E} = -\text{grad } \phi - \frac{\partial \mathbf{A}}{\partial t} \quad [47]$$

Such a vector field \mathbf{A} is called a “magnetic vector potential”; a function ϕ such that eqn [47] holds is called an “electric scalar potential.”

Conversely, given scalar and vector functions ϕ and \mathbf{A} of t, x, y, z , we can define \mathbf{B} and \mathbf{E} by

$$\mathbf{B} = \text{curl } \mathbf{A}, \quad \mathbf{E} = -\text{grad } \phi - \frac{\partial \mathbf{A}}{\partial t} \quad [48]$$

Then two of Maxwell’s equations hold automatically, since

$$\text{div } \mathbf{B} = 0, \quad \text{curl } \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0$$

The remaining pair translate into conditions on \mathbf{A} and ϕ . Equation [26] becomes

$$\text{div } \mathbf{E} = -\nabla^2 \phi - \frac{\partial}{\partial t} (\text{div } \mathbf{A}) = \frac{\rho}{\epsilon_0}$$

and eqn [28] becomes

$$\begin{aligned} \text{curl } \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} &= -\nabla^2 \mathbf{A} + \text{grad } \text{div } \mathbf{A} \\ &+ \frac{1}{c^2} \frac{\partial}{\partial t} \left(\text{grad } \phi + \frac{\partial \mathbf{A}}{\partial t} \right) \\ &= \mu_0 \mathbf{J} \end{aligned}$$

If we put

$$\boldsymbol{\alpha} = \frac{1}{c^2} \frac{\partial \phi}{\partial t} + \text{div } (\mathbf{A})$$

then we can rewrite the equations for \mathbf{A} and ϕ more simply as

$$\begin{aligned} \square \phi - \frac{\partial \boldsymbol{\alpha}}{\partial t} &= \frac{\rho}{\epsilon_0} \\ \square \mathbf{A} + \text{grad } \boldsymbol{\alpha} &= \mu_0 \mathbf{J} \end{aligned}$$

Here we have four equations (one scalar, one vector) in four unknowns (ϕ and the components of \mathbf{A}). Any set of solutions ϕ, \mathbf{A} determines a solution of Maxwell's equations via [48].

Gauge Transformations

Given solutions \mathbf{E} and \mathbf{B} of Maxwell's equations, what freedom is there in the choice of \mathbf{A} and ϕ ? First, \mathbf{A} is determined by $\text{curl}\mathbf{A}=\mathbf{B}$ up to the replacement of \mathbf{A} by

$$\mathbf{A}' = \mathbf{A} + \text{grad } u$$

for some function u of position and time. The scalar potential ϕ' corresponding to \mathbf{A}' must be chosen so that

$$\begin{aligned} -\text{grad } \phi' &= \mathbf{E} + \frac{\partial \mathbf{A}'}{\partial t} \\ &= \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} + \text{grad} \left(\frac{\partial u}{\partial t} \right) \\ &= -\text{grad} \left(\phi - \frac{\partial u}{\partial t} \right) \end{aligned}$$

That is, $\phi' = \phi - \partial u / \partial t + f(t)$, where f is a function of t alone. We can absorb f into u by subtracting

$$\int f \, dt$$

(this does not alter \mathbf{A}'). So the freedom in the choice of \mathbf{A} and ϕ is to make the transformation

$$\mathbf{A} \mapsto \mathbf{A}' = \mathbf{A} + \text{grad } u, \quad \phi \mapsto \phi' = \phi - \frac{\partial u}{\partial t} \quad [49]$$

for any $u = u(t, x, y, z)$. The transformation [49] is called a “gauge transformation.”

Under [49],

$$\alpha \mapsto \alpha' = \frac{1}{c^2} \frac{\partial \phi'}{\partial t} + \text{div}(\mathbf{A}') = \alpha - \square u$$

It is possible to show, under certain very mild conditions on α , that the inhomogeneous wave equation

$$\square u = \alpha \quad [50]$$

has a solution $u = u(t, x, y, z)$. If we choose u so that [50] holds, then the transformed potentials \mathbf{A}' and ϕ' satisfy

$$\text{div}(\mathbf{A}') + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} = 0$$

This is the “Lorenz gauge condition,” named after L Lorenz (not the H A Lorentz of the “Lorentz contraction”).

If we impose the Lorenz condition, then the only remaining freedom in the choice of \mathbf{A} and ϕ is to make gauge transformations [49] in which u is a solution of the wave equation $\square u = 0$. Under the Lorenz condition, Maxwell's equations take the form

$$\square \phi = \rho / \epsilon_0, \quad \square \mathbf{A} = \mu_0 \mathbf{J} \quad [51]$$

Consistency with the Lorenz condition follows from the continuity equation on ϕ and \mathbf{J} .

In the absence of sources, therefore, Maxwell's equations for the potential in the Lorenz gauge reduce to

$$\square \phi = 0, \quad \square \mathbf{A} = 0 \quad [52]$$

together with the constraint

$$\text{div } \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0$$

We can, for example, choose three arbitrary solutions of the scalar wave equation for the components of the vector potential, and then define ϕ by

$$\phi = c^2 \int \text{div } \mathbf{A} \, dt$$

Whatever choice we make, we shall get a solution of Maxwell's equations, and every solution of Maxwell's equations (without sources) will arise from some such choice.

Historical Note

At the end of the eighteenth century, four types of electromagnetic phenomena were known, but not the connections between them.

- *Magnetism*, the word derives from the Greek for “stone from Magnesia.”
- *Static electricity*, produced by rubbing amber with fur; the word “electricity” derives from the Greek for “amber.”
- *Light*.
- *Galvanism* or “animal electricity” – the electricity produced by batteries, discovered by Luigi Galvani.

The construction of a unified theory was a slow and painful business. It was hindered by attempts, which seem bizarre in retrospect, to understand electromagnetism in terms of underlying mechanical models involving such inventions as “electric fluids” and “magnetic vortices.” We can see the legacy of this period, which ended with Einstein's work in 1905, in the misleading and archaic terms that still survive in modern terminology: “magnetic flux,” “lines of force,” “electric displacement,” and so on.

Maxwell's contribution was decisive, although much of what we now call "Maxwell's theory" is due to his successors (Lorentz, Hertz, Einstein, and so on); and, as we shall see, a key element in Maxwell's own description of electromagnetism – the "electromagnetic ether," an all-pervasive medium which was supposed to transmit electromagnetic waves – was thrown out by Einstein.

A rough chronology is as follows.

- 1800 Volta demonstrated the connection between galvanism and static electricity.
- 1820 Oersted showed that the current from a battery generates a force on a magnet.
- 1822 Ampère suggested that light was a wave motion in a "luminiferous ether" made up of two types of electric fluid. In the same year, Galileo's "Dialogue concerning the two chief world systems" was removed from the index of prohibited books.
- 1831 Faraday showed that moving magnets can induce currents.

- 1846 Faraday suggested that light is a vibration in magnetic lines of force.
- 1863 Maxwell published the equations that describe the dynamics of electric and magnetic fields.
- 1905 Einstein's paper "On the electrodynamics of moving bodies."

Further Reading

- Chalmers AF (1975) Maxwell and the displacement current. *Physics Education* January 1975: 45–49.
- Einstein A (1905) *On the Electrodynamics of Moving Bodies*. A translation of the paper can be found in *The Principle of Relativity* by Lorentz HA, Einstein A, Minkowski H, and Weyl H, with notes by Sommerfeld A. New York: Dover, 1952.
- Roche J (1998) The present status of Maxwell's displacement current. *European Journal of Physics* 19: 155–166.
- Siegel DM (1985) Mechanical image and reality in Maxwell's electromagnetic theory. In: Harman PM (ed.) *Wranglers and Physicists*. Manchester: Manchester University Press.

Introductory Article: Equilibrium Statistical Mechanics

G Gallavotti, Università di Roma "La Sapienza," Rome, Italy

© 2006 G Gallavotti. Published by Elsevier Ltd.
All rights reserved.

Foundations: Atoms and Molecules

Classical statistical mechanics studies properties of macroscopic aggregates of particles, atoms, and molecules, based on the assumption that they are point masses subject to the laws of classical mechanics. Distinction between macroscopic and microscopic systems is evanescent and in fact the foundations of statistical mechanics have been laid on properties, proved or assumed, of few-particle systems.

Macroscopic systems are often considered in stationary states, which means that their microscopic configurations follow each other as time evolves while looking the same macroscopically. Observing time evolution is the same as sampling ("not too closely" time-wise) independent copies of the system prepared in the same way.

A basic distinction is necessary: a stationary state may or may not be in equilibrium. The first case arises when the particles are enclosed in a container Ω and are subject only to their mutual conservative

interactions and, possibly, to external conservative forces: a typical example is a gas in a container subject to forces due to the walls of Ω and gravity, besides the internal interactions. This is a very restricted class of systems and states.

A more general case is when the system is in a stationary state but it is also subject to nonconservative forces: a typical example is a gas or fluid in which a wheel rotates, as in the Joule experiment, with some device acting to keep the temperature constant. The device is called a thermostat and in statistical mechanics it has to be modeled by forces, including nonconservative ones, which prevent an indefinite energy transfer from the external forcing to the system: such a transfer would impede the occurrence of stationary states. For instance, the thermostat could simply be a constant friction force (as in stirred incompressible liquids or as in electric wires in which current circulates because of an electromotive force).

A more fundamental approach would be to imagine that the thermostat device is not a phenomenologically introduced nonconservative force (e.g., a friction force) but is due to the interaction with an external infinite system which is in "equilibrium at infinity."

In any event nonequilibrium stationary states are intrinsically more complex than equilibrium states. Here attention will be confined to equilibrium

statistical mechanics of systems of N identical point particles $\mathbf{Q} = (q_1, \dots, q_N)$ enclosed in a cubic box Ω , with volume V and side L , normally assumed to have perfectly reflecting walls.

Particles of mass m located at \mathbf{q}, \mathbf{q}' will be supposed to interact via a pair potential $\varphi(\mathbf{q} - \mathbf{q}')$. The microscopic motion follows the equations

$$m\ddot{\mathbf{q}}_i = -\sum_{j=1}^N \partial_{\mathbf{q}_i} \varphi(\mathbf{q}_i - \mathbf{q}_j) + \sum_i W_{\text{wall}}(\mathbf{q}_i) \stackrel{\text{def}}{=} -\partial_{\mathbf{q}_i} \Phi(\mathbf{Q}) \quad [1]$$

where the potential φ is assumed to be smooth except, possibly, for $|\mathbf{q} - \mathbf{q}'| \leq r_0$ where it could be $+\infty$, that is, the particles cannot come closer than r_0 , and at r_0 [1] is interpreted by imagining that they undergo elastic collisions; the potential W_{wall} models the container and it will be replaced, unless explicitly stated, by an elastic collision rule.

The time evolution $(\mathbf{Q}, \dot{\mathbf{Q}}) \rightarrow S_t(\mathbf{Q}, \dot{\mathbf{Q}})$ will, therefore, be described on the position – velocity space, $\widehat{\mathcal{F}}(N)$, of the N particles or, more conveniently, on the phase space, i.e., by a time evolution S_t on the momentum – position (\mathbf{P}, \mathbf{Q}) , with $\mathbf{P} = m\dot{\mathbf{Q}}$ space, $\mathcal{F}(N)$. The motion being conservative, the energy

$$U \stackrel{\text{def}}{=} \sum_i \frac{1}{2m} p_i^2 + \sum_{i < j} \varphi(\mathbf{q}_i - \mathbf{q}_j) + \sum_i W_{\text{wall}}(\mathbf{q}_i) \stackrel{\text{def}}{=} K(\mathbf{P}) + \Phi(\mathbf{Q})$$

will be a constant of motion; the last term in Φ is missing if walls are perfect. This makes it convenient to regard the dynamics as associated with two dynamical systems $(\mathcal{F}(N), S_t)$ on the $6N$ -dimensional phase space, and $(\mathcal{F}_U(N), S_t)$ on the $(6N - 1)$ -dimensional surface of energy U . Since the dynamics [1] is Hamiltonian on phase space, with Hamiltonian

$$H(\mathbf{P}, \mathbf{Q}) \stackrel{\text{def}}{=} \sum_i \frac{1}{2m} p_i^2 + \Phi(\mathbf{Q}) \stackrel{\text{def}}{=} K + \Phi$$

it follows that the volume $d^{3N} \mathbf{P} d^{3N} \mathbf{Q}$ is conserved (i.e., a region E has the same volume as $S_t E$) and also the area $\delta(H(\mathbf{P}, \mathbf{Q}) - U) d^{3N} \mathbf{P} d^{3N} \mathbf{Q}$ is conserved.

The above dynamical systems are well defined, i.e., S_t is a map on phase space globally defined for all $t \in (-\infty, \infty)$, when the interaction potential is bounded below: this is implied by the *a priori* bounds due to energy conservation. For gravitational or Coulomb interactions, much more has to be said, assumed, and done in order to even define the key quantities needed for a statistical theory of motion.

Although our world is three dimensional (or at least was so believed to be until recent revolutionary

theories), it will be useful to consider also systems of particles in dimension $d \neq 3$: in this case the above $6N$ and $3N$ become, respectively, $2dN$ and dN . Systems with dimension $d=1, 2$ are in fact sometimes very good models for thin filaments or thin films. For the same reason, it is often useful to imagine that space is discrete and particles can only be located on a lattice, for example, on \mathbb{Z}^d (see the section “Lattice models”).

The reader is referred to Gallavotti (1999) for more details.

Pressure, Temperature, and Kinetic Energy

The beginning was BERNOULLI’s derivation of the perfect gas law via the identification of the *pressure* at *numerical density* ρ with the average momentum transferred per unit time to a surface element of area dS on the walls: that is, the average of the observable $2mv\rho v dS$, with v the normal component of the velocity of the particles that undergo collisions with dS . If $f(v)dv$ is the distribution of the normal component of velocity and $f(v)d^3v \equiv \prod_i f(v_i)d^3v$, $\mathbf{v} = (v_1, v_2, v_3)$, is the total velocity distribution, the average of the momentum transferred is $p dS$ given by

$$dS \int_{v>0} 2mv^2 \rho f(v) dv = dS \int mv^2 \rho f(v) dv = \rho \frac{2}{3} dS \int \frac{m}{2} \mathbf{v}^2 f(\mathbf{v}) d^3\mathbf{v} = \rho \frac{2}{3} \left\langle \frac{K}{N} \right\rangle dS \quad [2]$$

Furthermore $(2/3)\langle K/N \rangle$ was identified as proportional to the absolute temperature $\langle K/N \rangle \stackrel{\text{def}}{=} \text{const} (3/2)T$ which, with present-day notations, is written as $(2/3)\langle K/N \rangle = k_B T$. The constant k_B was (later) called Boltzmann’s constant and it is the same for at least all perfect gases. Its independence on the particular nature of the gas is a consequence of *Avogadro’s law* stating that equal volumes of gases at the same conditions of temperature and pressure contain equal number of molecules.

Proportionality between average kinetic energy and temperature via the universal constant k_B became in fact a fundamental assumption extending to all aggregates of particles gaseous or not, never challenged in all later works (until quantum mechanics, where this is no longer true, see the section “Quantum statistics”).

For more details, we refer the reader to Gallavotti (1999).

Heat and Entropy

After Clausius' discovery of entropy, BOLTZMANN, in order to explain it mechanically, introduced the *heat theorem*, which he developed to full generality between 1866 and 1884. Together with the mentioned identification of absolute temperature with average kinetic energy, the heat theorem can also be considered a founding element of statistical mechanics.

The theorem makes precise the notion of time average and then states in great generality that given any mechanical system one can associate with its dynamics four quantities U, V, p, T , defined as time averages of suitable mechanical observables (i.e., functions on phase space), so that when the external conditions are infinitesimally varied and the quantities U, V change by dU, dV , respectively, the ratio $(dU + pdV)/T$ is exact, i.e., there is a function $S(U, V)$ whose corresponding variation equals the ratio. It will be better, for the purpose of considering very large boxes ($V \rightarrow \infty$) to write this relation in terms of intensive quantities $u \stackrel{\text{def}}{=} U/N$ and $v = V/N$ as

$$\frac{du + pdv}{T} \text{ is exact} \quad [3]$$

i.e., the ratio equals the variation ds of $s(U/N, V/N) \equiv (1/N)S(U, V)$.

The proof originally dealt with *monocyclic* systems, i.e., systems in which all motions are periodic. The assumption is clearly much too restrictive and justification for it developed from the early "nonperiodic motions can be regarded as periodic with infinite period" (1866), to the later *ergodic hypothesis* and finally to the realization that, after all, the heat theorem does not really depend on the ergodic hypothesis (1884).

Although for a one-dimensional system the proof of the heat theorem is a simple check, it was a real breakthrough because it led to an answer to the general question as to under which conditions one could define mechanical quantities whose variations were constrained to satisfy [3] and therefore could be interpreted as a mechanical model of Clausius' macroscopic thermodynamics. It is reproduced in the following.

Consider a one-dimensional system subject to forces with a confining potential $\varphi(x)$ such that $|\varphi'(x)| > 0$ for $|x| > 0, \varphi''(0) > 0$ and $\varphi(x) \xrightarrow{x \rightarrow \pm\infty} +\infty$. All motions are periodic, so that the system is monocyclic. Suppose that the potential $\varphi(x)$ depends on a parameter V and define a *state* to be a motion with given energy U and given V ; let

$$\begin{aligned} U &= \text{total energy of the system} \equiv K + \Phi \\ T &= \text{time average of the kinetic energy } K = \langle K \rangle \\ V &= \text{the parameter on which } \varphi \\ &\quad \text{is supposed to depend} \\ p &= -\text{time average of } \partial_V \varphi, -\langle \partial_V \varphi \rangle \end{aligned} \quad [4]$$

A state is thus parametrized by U, V . If such parameters change by dU, dV , respectively, and if $dL \stackrel{\text{def}}{=} -pdV, dQ \stackrel{\text{def}}{=} dU + pdV$, then [3] holds. In fact, let $x_{\pm}(U, V)$ be the extremes of the oscillations of the motion with given U, V and define S as

$$\begin{aligned} S &= 2 \log \int_{x_{-}(U,V)}^{x_{+}(U,V)} \sqrt{(U - \varphi(x))} dx \\ \Rightarrow dS &= \frac{\int (dU - \partial_V \varphi(x) dV) (dx/\sqrt{K})}{\int (dx/\sqrt{K}) K} \end{aligned} \quad [5]$$

Noting that $dx/\sqrt{K} = \sqrt{2/m} dt$, [3] follows because time averages are given by integrating with respect to dx/\sqrt{K} and dividing by the integral of $1/\sqrt{K}$.

For more details, the reader is referred to Boltzmann (1968b) and Gallavotti (1999).

Heat Theorem and Ergodic Hypothesis

Boltzmann tried to extend the result beyond the one-dimensional systems (e.g., to Keplerian motions, which are not monocyclic unless only motions with a fixed eccentricity are considered). However, the early statement that "aperiodic motions can be regarded as periodic with infinite period" is really the heart of the application of the heat theorem for monocyclic systems to the far more complex gas in a box.

Imagine that the gas container Ω is closed by a piston of section A located to the right of the origin at distance L and acting as a lid, so that the volume is $V = AL$. The microscopic model for the piston will be a potential $\bar{\varphi}(L - \xi)$ if $x = (\xi, \eta, \zeta)$ are the coordinates of a particle. The function $\bar{\varphi}(r)$ will vanish for $r > r_0$, for some $r_0 \ll L$, and diverge to $+\infty$ at $r = 0$. Thus, r_0 is the width of the layer near the piston where the force of the wall is felt by the particles that happen to be roaming there.

The contribution to the total potential energy Φ due to the walls is $W_{\text{wall}} = \sum_j \bar{\varphi}(L - \xi_j)$ and $\partial_V \bar{\varphi} = A^{-1} \partial_L \bar{\varphi}$; assuming monocyclicity, it is necessary to evaluate the time average of $\partial_L \Phi(x) = \partial_L W_{\text{wall}} \equiv -\sum_j \bar{\varphi}'(L - \xi_j)$. As time evolves, the particles x_j with ξ_j in the layer within r_0 of the wall will feel the force exercised by the wall and

bounce back. One particle in the layer will contribute to the average of $\partial_L \Phi(x)$ the amount

$$\frac{1}{\text{total time}} 2 \int_{t_0}^{t_1} -\overline{\varphi}'(L - \xi_j) dt \quad [6]$$

if t_0 is the first instant when the point j enters the layer and t_1 is the instant when the ξ -component of the velocity vanishes “against the wall.” Since $-\overline{\varphi}'(L - \xi_j)$ is the ξ -component of the force, the integral is $2m|\dot{\xi}_j|$ (by Newton’s law), provided, of course, $\dot{\xi}_j > 0$.

Suppose that no collisions between particles occur while the particles travel within the range of the potential of the wall, i.e., the mean free path is much greater than the range of the potential $\overline{\varphi}$ defining the wall. The contribution of collisions to the average momentum transfer to the wall per unit time is therefore given by, see [2],

$$\int_{v>0} 2mv f(v) \rho_{\text{wall}} A v dv$$

if $\rho_{\text{wall}}, f(v)$ are the average density near the wall and, respectively, the average fraction of particles with a velocity component normal to the wall between v and $v + dv$. Here p, f are supposed to be independent of the point on the wall: this should be true up to corrections of size $o(A)$.

Thus, writing the average kinetic energy per particle and per velocity component, $\int (m/2)v^2 f(v) dv$, as $(1/2)\beta^{-1}$ (cf. [2]) it follows that

$$p \stackrel{\text{def}}{=} -\langle \partial_V \Phi \rangle = \rho_{\text{wall}} \beta^{-1} \quad [7]$$

has the physical interpretation of pressure. $(1/2)\beta^{-1}$ is the average kinetic energy per degree of freedom: hence, it is proportional to the absolute temperature T (cf. see the section “Pressure, temperature, and kinetic energy”).

On the other hand, if motion on the energy surface takes place on a single periodic orbit, the quantity p in [7] is the right quantity that would make the heat theorem work; see [4]. Hence, regarding the trajectory on each energy surface as periodic (i.e., the system as monocyclic) leads to the heat theorem with p, U, V, T having the right physical interpretation corresponding to their appellations. This shows that monocyclic systems provide natural models of thermodynamic behavior.

Assuming that a chaotic system like a gas in a container of volume V will satisfy, for practical purposes, the above property, a quantity p can be defined such that $dU + p dV$ admits the inverse of the average kinetic energy $\langle K \rangle$ as an integrating factor and, furthermore, $p, U, V, \langle K \rangle$ have the physical interpretations of pressure, energy, volume,

and (up to a proportionality factor) absolute temperature, respectively.

Boltzmann’s conception of space (and time) as discrete allowed him to conceive the property that the energy surface is constituted by “points” all of which belong to a single trajectory: a property that would be impossible if the phase space was really a continuum. Regarding phase space as consisting of a finite number of “cells” of finite volume h^{dN} , for some $h > 0$ (rather than of a continuum of points), allowed him to think, without logical contradiction, that the energy surface consisted of a single trajectory and, hence, that motion was a cyclic permutation of its points (actually cells).

Furthermore, it implied that the time average of an observable $F(\mathbf{P}, \mathbf{Q})$ had to be identified with its average on the energy surface computed via the Liouville distribution

$$C^{-1} \int F(\mathbf{P}, \mathbf{Q}) \delta(H(\mathbf{P}, \mathbf{Q}) - U) d\mathbf{P} d\mathbf{Q}$$

with

$$C = \int \delta(H(\mathbf{P}, \mathbf{Q}) - U) d\mathbf{P} d\mathbf{Q}$$

(the appropriate normalization factor): a property that was written symbolically

$$\frac{dt}{T} = \frac{d\mathbf{P} d\mathbf{Q}}{\int d\mathbf{P} d\mathbf{Q}}$$

or

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T F(S_t(\mathbf{P}, \mathbf{Q})) dt \\ = \frac{\int F(\mathbf{P}', \mathbf{Q}') \delta(H(\mathbf{P}', \mathbf{Q}') - U) d\mathbf{P}' d\mathbf{Q}'}{\int \delta(H(\mathbf{P}', \mathbf{Q}') - U) d\mathbf{P}' d\mathbf{Q}'} \quad [8] \end{aligned}$$

The validity of [8] for all (piecewise smooth) observables F and for all points of the energy surface, with the exception of a set of zero area, is called the ergodic hypothesis.

For more details, the reader is referred to Boltzmann (1968) and Gallavotti (1999).

Ensembles

Eventually Boltzmann in 1884 realized that the validity of the heat theorem for averages computed via the right-hand side (rhs) of [8] held independently of the ergodic hypothesis, that is, [8] was not necessary because the heat theorem (i.e., [3]) could also be derived under the only assumption that the averages involved in its formulation were computed

as averages over phase space with respect to the probability distribution on the rhs of [8].

Furthermore, if T was identified with the average kinetic energy, U with the average energy, and p with the average force per unit surface on the walls of the container Ω with volume V , the relation [3] held for a variety of families of probability distributions on phase space, besides [8]. Among these are:

1. The “microcanonical ensemble,” which is the collection of probability distributions on the rhs of [8] parametrized by $u = U/N, v = V/N$ (energy and volume per particle),

$$\begin{aligned} \mu_{u,v}^{\text{mc}}(\text{dP dQ}) \\ = \frac{1}{Z_{\text{mc}}(U, N, V)} \delta(H(\mathbf{P}, \mathbf{Q}) - U) \frac{\text{dP dQ}}{N! h^{dN}} \quad [9] \end{aligned}$$

where h is a constant with the dimensions of an action which, in the discrete representation of phase space mentioned in the previous section, can be taken such that h^{dN} equals the volume of the cells and, therefore, the integrals with respect to [9] can be interpreted as an (approximate) sum over the cells conceived as microscopic configurations of N indistinguishable particles (whence the $N!$).

2. The “canonical ensemble,” which is the collection of probability distributions parametrized by $\beta, v = V/N$,

$$\mu_{\beta,v}^{\text{c}}(\text{dPdQ}) = \frac{1}{Z_{\text{c}}(\beta, N, V)} e^{-\beta H(\mathbf{P}, \mathbf{Q})} \frac{\text{dPdQ}}{N! h^{dN}} \quad [10]$$

to which more ensembles can be added, such as the grand canonical ensemble (Gibbs).

3. The “grand canonical ensemble” which is the collection of probability distributions parameterized by β, λ and defined over the space $\mathcal{F}_{\text{gc}} = \cup_{N=0}^{\infty} \mathcal{F}(N)$,

$$\begin{aligned} \mu_{\beta,\lambda}^{\text{gc}}(\text{dPdQ}) \\ = \frac{1}{Z_{\text{gc}}(\beta, \lambda, V)} e^{\beta \lambda N - \beta H(\mathbf{P}, \mathbf{Q})} \frac{\text{dPdQ}}{N! h^{dN}} \quad [11] \end{aligned}$$

Hence, there are several different models of thermodynamics. The key tests for accepting them as real microscopic descriptions of macroscopic thermodynamics are as follows.

1. A correspondence between the macroscopic states of thermodynamic equilibrium and the elements of a collection of probability distributions on phase space can be established by identifying, on the one hand, macroscopic thermodynamic states with given values of the thermodynamic functions and, on the other,

probability distributions attributing the same average values to the corresponding microscopic observables (i.e., whose averages have the interpretation of thermodynamic functions).

2. Once the correct correspondence between the elements of the different ensembles is established, that is, once the pairs $(u, v), (\beta, v), (\beta, \lambda)$ are so related to produce the same values for the averages $U, V, k_{\text{B}} T \stackrel{\text{def}}{=} \beta^{-1}, p|\partial\Omega|$ of

$$H(\mathbf{P}, \mathbf{Q}), V, \frac{2K(\mathbf{P})}{3N}, \int \delta_{\partial\Omega}(\mathbf{q}_1) 2m(\mathbf{v}_1 \cdot \mathbf{n})^2 \text{d}\mathbf{q}_1 \quad [12]$$

where $(\delta_{\partial\Omega}(\mathbf{q}_1))$ is a delta-function pinning \mathbf{q}_1 to the surface $\partial\Omega$, then the averages of all physically interesting observables *should coincide* at least in the thermodynamic limit, $\Omega \rightarrow \infty$. In this way, the elements μ of the considered collection of probability distributions can be identified with the states of macroscopic equilibrium of the system. The μ 's depend on parameters and therefore they form an ensemble: each of them corresponds to a macroscopic equilibrium state whose thermodynamic functions are appropriate averages of microscopic observables and therefore are functions of the parameters identifying μ .

Remark The word “ensemble” is often used to indicate the individual probability distributions of what has been called here an ensemble. The meaning used here seems closer to the original sense in the 1884 paper of Boltzmann (in other words, often by “ensemble” one means that collection of the phase space points on which a given probability distribution is considered, and this does not seem to be the original sense).

For instance, in the case of the microcanonical distributions this means interpreting energy, volume, temperature, and pressure of the equilibrium state with specific energy u and specific volume v as proportional, through appropriate universal proportionality constants, to the integrals with respect to $\mu_{u,v}^{\text{mc}}(\text{dP dQ})$ of the mechanical quantities in [12]. The averages of other thermodynamic observables in the state with specific energy u and specific volume v should be given by their integrals with respect to $\mu_{u,v}^{\text{mc}}$.

Likewise, one can interpret energy, volume, temperature, and pressure of the equilibrium state with specific energy u and specific volume v as the averages of the mechanical quantities [12] with respect to the canonical distribution $\mu_{\beta,v}^{\text{c}}(\text{dP dQ})$ which has average specific energy precisely u . The averages of other thermodynamic observables in the state with specific energy and volume u and v are

given by their integrals with respect to $\mu_{\beta,v}^c$. A similar definition can be given for the description of thermodynamic equilibria via the grand canonical distributions.

For more details, see Gibbs (1981) and Gallavotti (1999).

Equivalence of Ensembles

BOLTZMANN proved that, computing averages via the microcanonical or canonical distributions, the essential property [3] was satisfied when changes in their parameters (i.e., u, v or β, v , respectively) induced changes du and dv on energy and volume, respectively. He also proved that the function s , whose existence is implied by [3], was the same function once expressed as a function of u, v (or of any pair of thermodynamic parameters, e.g., of T, v or p, u). A close examination of Boltzmann's proof shows that the [3] holds exactly in the canonical ensemble and up to corrections tending to 0 as $\Omega \rightarrow \infty$ in the microcanonical ensemble. Identity of thermodynamic functions evaluated in the two ensembles holds, as a consequence, up to corrections of this order. In addition, Gibbs added that the same held for the grand canonical ensemble.

Of course, not every collection of stationary probability distributions on phase space would provide a model for thermodynamics: Boltzmann called "orthodic" the collections of stationary distributions which generated models of thermodynamics through the above-mentioned identification of its elements with macroscopic equilibrium states. The microcanonical, canonical, and the later grand canonical ensembles are the chief examples of orthodic ensembles. Boltzmann and Gibbs proved these ensembles to be not only orthodic but to generate the same thermodynamic functions, that is to generate the same thermodynamics.

This meant freedom from the analysis of the truth of the doubtful ergodic hypothesis (still unproved in any generality) or of the monocyclicity (manifestly false if understood literally rather than regarding the phase space as consisting of finitely many small, discrete cells), and allowed Gibbs to formulate the problem of statistical mechanics of equilibrium as follows.

Problem Study the properties of the collection of probability distributions constituting (any) one of the above ensembles.

However, by no means the three ensembles just introduced exhaust the class of orthodic ensembles producing the same models of thermodynamics in the limit of infinitely large systems. The wealth of

ensembles with the orthodicity property, hence leading to equivalent mechanical models of thermodynamics, can be naturally interpreted in connection with the phenomenon of phase transition (see the section "Phase transitions and boundary conditions").

Clearly, the quoted results do not "prove" that thermodynamic equilibria "are" described by the microcanonical, canonical, or grand canonical ensembles. However, they certainly show that, for most systems, independently of the number of degrees of freedom, one can define quite unambiguously a mechanical model of thermodynamics establishing parameter-free, system-independent, physically important relations between thermodynamic quantities (e.g., $\partial_u(p(u, v)/T(u, v)) \equiv \partial_v(1/T(u, v))$, from [3]).

The ergodic hypothesis which was at the root of the mechanical theorems on heat and entropy cannot be taken as a justification of their validity. Naively one would expect that the time scale necessary to see an equilibrium attained, called recurrence time scale, would have to be at least the time that a phase space point takes to visit all possible microscopic states of given energy: hence, an explanation of why the necessarily enormous size of the recurrence time is not a problem becomes necessary.

In fact, the recurrence time can be estimated once the phase space is regarded as discrete: for the purpose of countering mounting criticism, Boltzmann assumed that momentum was discretized in units of $(2mk_B T)^{1/2}$ (i.e., the average momentum size) and space was discretized in units of $\rho^{-1/3}$ (i.e., the average spacing), implying a volume of cells h^{3N} with $h \stackrel{\text{def}}{=} \rho^{-1/3} (2mk_B T)^{1/2}$; then he calculated that, even with such a gross discretization, a cell representing a microscopic state of 1cm^3 of hydrogen at normal condition would require a time (called "recurrence time") of the order of $\sim 10^{10^{19}}$ times the age of the Universe (!) to visit the entire energy surface. In fact, the phase space volume is $\Gamma = (\rho^{-3} N (2mk_B T)^{3/2})^N \equiv h^{3N}$ and the number of cells of volume h^{3N} is $\Gamma / (N! h^{3N}) \simeq e^{3N}$; and the time to visit all will be $e^{3N} \tau_0$, with τ_0 a typical atomic unit, e.g., 10^{-12} s – but $N = 10^{19}$. In this sense, the statement boldly made by young Boltzmann that "aperiodic motions can be regarded as periodic with infinite period" was even made quantitative.

The recurrence time is clearly so long to be irrelevant for all purposes: nevertheless, the correctness of the microscopic theory of thermodynamics can still rely on the microscopic dynamics once it is understood (as stressed by Boltzmann) that the reason why we observe approach to equilibrium, and equilibrium itself, over "human" timescales

(which are far shorter than the recurrence times) is due to the property that on most of the energy surface the (very few) observables whose averages yield macroscopic thermodynamic functions (namely pressure, temperature, energy, ...) *assume the same value* even if N is only very moderately large (of the order of 10^3 rather than 10^{19}). This implies that this value coincides with the average and therefore satisfies the heat theorem without any contradiction with the length of the recurrence time. The latter rather concerns the time needed to the *generic observable* to thermalize, that is, to reach its time average: the generic observable will indeed take a very long time to “thermalize” but no one will ever notice, because the generic observable (e.g., the position of a pre-identified particle) is not relevant for thermodynamics.

The word “proof” is not used in the mathematical sense so far in this article: the relevance of a mathematically rigorous analysis was widely realized only around the 1960s at the same time when the first numerical studies of the thermodynamic functions became possible and rigorous results were needed to check the correctness of various numerical simulations.

For more details, the reader is referred to Boltzmann (1968a, b) and Gallavotti (1999).

Thermodynamic Limit

Adopting Gibbs axiomatic point of view, it is interesting to see the path to be followed to achieve an equivalence proof of three ensembles introduced in the section “Heat theorem and ergodic hypothesis.”

A preliminary step is to consider, given a cubic box Ω of volume $V = L^d$, the normalization factors $Z^{\text{gc}}(\beta, \lambda, V)$, $Z^{\text{c}}(\beta, N, V)$, and $Z^{\text{mc}}(U, N, V)$ in [9], [10], and [11], respectively, and to check that the following thermodynamic limits exist:

$$\begin{aligned} \beta p_{\text{gc}}(\beta, \lambda) &\stackrel{\text{def}}{=} \lim_{V \rightarrow \infty} \frac{1}{V} \log Z^{\text{gc}}(\beta, \lambda, V) \\ -\beta f_{\text{c}}(\beta, \rho) &\stackrel{\text{def}}{=} \lim_{V \rightarrow \infty, \frac{N}{V} = \rho} \frac{1}{N} \log Z^{\text{c}}(\beta, N, V) \\ k_{\text{B}}^{-1} s_{\text{mc}}(u, \rho) &\stackrel{\text{def}}{=} \lim_{V \rightarrow \infty, \frac{N}{V} = \rho, \frac{U}{N} = u} \frac{1}{N} \log Z^{\text{mc}}(U, N, V) \end{aligned} \quad [13]$$

where the density $\rho \stackrel{\text{def}}{=} v^{-1} \equiv N/V$ is used, instead of v , for later reference. The normalization factors play an important role because they have simple thermodynamic interpretation (see the next section): they are called grand canonical, canonical, and micro-canonical partition functions, respectively.

Not surprisingly, assumptions on the interparticle potential $\varphi(\mathbf{q} - \mathbf{q}')$ are necessary to achieve an existence proof of the limits in [13]. The assumptions on φ are not only quite general but also have a clear physical meaning. They are

1. *stability*: that is, existence of a constant $B \geq 0$ such that $\sum_{i < j}^N \varphi(\mathbf{q}_i - \mathbf{q}_j) \geq -BN$ for all $N \geq 0$, $\mathbf{q}_1, \dots, \mathbf{q}_N \in \mathbb{R}^d$, and
2. *temperedness*: that is, existence of constants $\varepsilon_0, R > 0$ such that $|\varphi(\mathbf{q} - \mathbf{q}')| < B|\mathbf{q} - \mathbf{q}'|^{-d-\varepsilon_0}$ for $|\mathbf{q} - \mathbf{q}'| > R$.

The assumptions are satisfied by essentially all microscopic interactions with the notable exceptions of the gravitational and Coulombic interactions, which require a separate treatment (and lead to somewhat different results on the thermodynamic behavior).

For instance, assumptions (1), (2) are satisfied if $\varphi(\mathbf{q})$ is $+\infty$ for $|\mathbf{q}| < r_0$ and smooth for $|\mathbf{q}| > r_0$, for some $r_0 \geq 0$, and furthermore $\varphi(\mathbf{q}) > B_0|\mathbf{q}|^{-(d+\varepsilon_0)}$ if $r_0 < |\mathbf{q}| \leq R$, while for $|\mathbf{q}| > R$ it is $|\varphi(\mathbf{q})| < B_1|\mathbf{q}|^{-(d+\varepsilon_0)}$, for some $B_0, B_1, \varepsilon_0 > 0, R > r_0$. Briefly, φ is fast diverging at contact and fast approaching 0 at large distance. This is called a (generalized) Lennard–Jones potential. If $r_0 > 0$, φ is called a hard-core potential. If $B_1 = 0$, the potential is said to have finite range. (See Appendix 1 for physical implications of violations of the above stability and temperedness properties.) However, in the following, it will be necessary, both for simplicity and to contain the length of the exposition, to restrict consideration to the case $B_1 = 0$, i.e., to

$$\begin{aligned} \varphi(\mathbf{q}) &> B_0|\mathbf{q}|^{-(d+\varepsilon_0)}, \quad r_0 < |\mathbf{q}| \leq R, \\ |\varphi(\mathbf{q})| &\equiv 0, \quad |\mathbf{q}| > R \end{aligned} \quad [14]$$

unless explicitly stated.

Assuming stability and temperedness, the existence of the limits in [13] can be mathematically proved: in Appendix 2, the proof of the first is analyzed to provide the simplest example of the technique. A remarkable property of the functions $\beta p_{\text{gc}}(\beta, \lambda)$, $-\beta p_{\text{f}}(\beta, \rho)$, and $\rho s_{\text{mc}}(u, \rho)$ is that they are convex functions: hence, they are continuous in the interior of their domains of definition and, at one variable fixed, are differentiable with respect to the other with at most countably many exceptions.

In the case of a potential without hard core ($\rho_{\text{max}} = \infty$), $-\beta p_{\text{f}}(\beta, \rho)$ can be checked to tend to 0 slower than ρ as $\rho \rightarrow 0$, and to $-\infty$ faster than $-\rho$ as $\rho \rightarrow \infty$ (essentially proportionally to $-\rho \log \rho$ in both cases). Likewise, in the same case, $s_{\text{mc}}(u, \rho)$ can be shown to tend to 0 slower than $u - u_{\text{min}}$ as $u \rightarrow u_{\text{min}}$, and to $-\infty$ faster than $-u$ as $u \rightarrow \infty$. The latter

asymptotic properties can be exploited to derive, from the relations between the partition functions in [13],

$$\begin{aligned} Z^{\text{gc}}(\beta, \lambda, V) &= \sum_{N=0}^{\infty} e^{\beta\lambda N} Z^c(\beta, N, V) \\ Z^c(\beta, N, V) &= \int_{-B}^{\infty} e^{-\beta U} Z^{\text{mc}}(U, N, V) dU \end{aligned} \quad [15]$$

and, from the above-mentioned convexity, the consequences

$$\begin{aligned} \beta p_{\text{mc}}(\beta, \lambda) &= \max_v (\beta\lambda v^{-1} - \beta v^{-1} f_c(\beta, v^{-1})) \\ -\beta f_c(\beta, v^{-1}) &= \max_u (-\beta u + k_B^{-1} s_{\text{mc}}(u, v^{-1})) \end{aligned} \quad [16]$$

and that the maxima are attained in points, or intervals, internal to the intervals of definition. Let v_{gc}, u_c be points where the maxima are, respectively, attained in [16].

Note that the quantity $e^{\beta\lambda N} Z^c(\beta, N, V) / Z^{\text{gc}}(\beta, \lambda, V)$ has the interpretation of probability of a density $v^{-1} = N/V$ evaluated in the grand canonical distribution. It follows that, if the maximum in the first of [16] is strict, that is, it is reached at a single point, the values of v^{-1} in closed intervals not containing the maximum point v_{gc}^{-1} have a probability behaving as $\langle e^{-cV}, c > 0$, as $V \rightarrow \infty$, compared to the probability of v^{-1} 's in any interval containing v_{gc}^{-1} . Hence, v_{gc} has the interpretation of average value of v in the grand canonical distribution, in the limit $V \rightarrow \infty$.

Likewise, the interpretation of

$$e^{-\beta u N} Z^{\text{mc}}(u N, N, V) / Z^c(\beta, N, V)$$

as probability in the canonical distribution of an energy density u shows that, if the maximum in the second of [16] is strict, the values of u in closed intervals not containing the maximum point u_c have a probability behaving as $\langle e^{-cV}, c > 0$, as $V \rightarrow \infty$, compared to the probability of u 's in any interval containing u_c . Hence, in the limit $\Omega \rightarrow \infty$, the average value of u in the canonical distribution is u_c .

If the maxima are strict, [16] also establishes a relation between the grand canonical density, the canonical free energy and the grand canonical parameter λ , or between the canonical energy, the microcanonical entropy, and the canonical parameter β :

$$\lambda = \partial_{v^{-1}} (v_{\text{gc}}^{-1} f_c(\beta, v_{\text{gc}}^{-1})), \quad k_B \beta = \partial_u s_{\text{mc}}(u_c, v^{-1}) \quad [17]$$

where convexity and strictness of the maxima imply the derivatives existence.

Remark Therefore, in the equivalence between canonical and microcanonical ensembles, the canonical distribution with parameters (β, v) should correspond with the microcanonical with parameters (u_c, v) . The grand canonical distribution

with parameters (β, λ) should correspond with the canonical with parameters (β, v_{gc}) .

For more details, the reader is referred to Ruelle (1969) and Gallavotti (1999).

Physical Interpretation of Thermodynamic Functions

The existence of the limits [13] implies several properties of interest. The first is the possibility of finding the physical meaning of the functions $p_{\text{gc}}, f_c, s_{\text{mc}}$ and of the parameters λ, β

Note first that, for all V the grand canonical average $\langle K \rangle_{\beta, \lambda}$ is $(d/2)\beta^{-1} \langle N \rangle_{\beta, \lambda}$ so that β^{-1} is proportional to the temperature $T_{\text{gc}} = T(\beta, \lambda)$ in the grand canonical distribution: $\beta^{-1} = k_B T(\beta, \lambda)$. Proceeding heuristically, the physical meaning of $p(\beta, \lambda)$ and λ can be found through the following remarks.

Consider the microcanonical distribution $\mu_{u,v}^{\text{mc}}$ and denote by \int^* the integral over (\mathbf{P}, \mathbf{Q}) extended to the domain of the (\mathbf{P}, \mathbf{Q}) such that $H(\mathbf{P}, \mathbf{Q}) = U$ and, at the same time, $\mathbf{q}_1 \in dV$, where dV is an infinitesimal volume surrounding the region Ω . Then, by the microscopic definition of the pressure p (see the introductory section), it is

$$\begin{aligned} p dV &= \frac{N}{Z(U, N, V)} \int^* \delta \frac{2}{3} \frac{p_1^2}{2m} \frac{d\mathbf{P} d\mathbf{Q}}{N! h^{dN}} \\ &\equiv \frac{2}{3Z(U, N, V)} \int^* \delta K(\mathbf{P}) \frac{d\mathbf{P} d\mathbf{Q}}{N! h^{dN}} \end{aligned} \quad [18]$$

where $\delta \equiv \delta(H(\mathbf{P}, \mathbf{Q}) - U)$. The RHS of [18] can be compared with

$$\frac{\partial_V Z(U, N, V) dV}{Z(U, N, V)} = \frac{N}{Z(U, N, V)} \int^* \frac{d\mathbf{P} d\mathbf{Q}}{N! h^{dN}}$$

to give

$$\frac{\partial_V Z dV}{Z} = N \frac{p dV}{(2/3)\langle K \rangle^*} = \beta p dV$$

because $\langle K \rangle^*$, which denotes the average $\int^* K / \int^* 1$, should be essentially the same as the microcanonical average $\langle K \rangle_{\text{mc}}$ (i.e., insensitive to the fact that one particle is constrained to the volume dV) if N is large. In the limit $V \rightarrow \infty, V/N = v$, the latter remark together with the second of [17] yields

$$\begin{aligned} k_B^{-1} \partial_v s_{\text{mc}}(u, v^{-1}) &= \beta p(u, v), \\ k_B^{-1} \partial_u s_{\text{mc}}(u, v) &= \beta \end{aligned} \quad [19]$$

respectively. Note that $p \geq 0$ and it is not increasing in v because $s_{\text{mc}}(\rho)$ is concave as a function of $v = \rho^{-1}$ (in fact, by the remark following [14] $\rho s_{\text{mc}}(u, \rho)$ is convex in ρ and, in general, if $\rho g(\rho)$ is convex in ρ then $g(v^{-1})$ is always concave in $v = \rho^{-1}$).

Hence, $ds_{\text{mc}}(u, v) = (du + pdv)/T$, so that taking into account the physical meaning of p , T (as pressure and temperature, see the section “Pressure, temperature, and kinetic energy”), s_{mc} is, in thermodynamics, the entropy. Therefore (see the second of [16]), $-\beta f_c(\beta, \rho) = -\beta u_c + k_B^{-1} s_{\text{mc}}(u_c, \rho)$ becomes

$$\begin{aligned} f_c(\beta, \rho) &= u_c - T_c s_{\text{mc}}(u_c, \rho), \\ df_c &= -p dv - s_{\text{mc}} dT \end{aligned} \quad [20]$$

and since u_c has the interpretation (as mentioned in the last section) of average energy in the canonical distribution $\mu_{\beta, v}^c$ it follows that f_c has the thermodynamic interpretation of free energy (once compared with the definition of free energy, $F = U - TS$, in thermodynamics).

By [17] and [20],

$$\lambda = \partial_{v^{-1}}(v_{\text{gc}}^{-1} f_c(\beta, v_{\text{gc}}^{-1})) \equiv u_c - T_c s_{\text{mc}} + p v_{\text{gc}}$$

and v_{gc} has the meaning of specific volume v . Hence, after comparison with the definition of chemical potential, $\lambda V = U - TS + pV$, in thermodynamics, it follows that the thermodynamic interpretation of λ is the chemical potential and (see [16], [17]), the grand canonical relation

$$\beta p_{\text{gc}}(\beta, \lambda) = \beta \lambda v_{\text{gc}}^{-1} - \beta v_{\text{gc}}^{-1} (-\beta u_c + k_B^{-1} s_{\text{mc}}(u_c, v^{-1}))$$

shows that $p_{\text{gc}}(\beta, \lambda) \equiv p$, implying that $p_{\text{gc}}(\beta, \lambda)$ is the pressure expressed, however, as a function of temperature and chemical potential.

To go beyond the heuristic derivations above, it should be remarked that convexity and the property that the maxima in [16], [17] are reached in the interior of the intervals of variability of v or u are sufficient to turn the above arguments into rigorous mathematical deductions: this means that given [19] as definitions of $p(u, v)$, $\beta(u, v)$, the second of [20] follows as well as $p_{\text{gc}}(\beta, \lambda) \equiv p(u_v, v_{\text{gc}}^{-1})$. But the values v_{gc} and u_c in [16] are not necessarily unique: convex functions can contain horizontal segments and therefore the general conclusion is that the maxima may possibly be attained in intervals. Hence, instead of a single v_{gc} , there might be a whole interval $[v_-, v_+]$, where the rhs of [16] reaches the maximum and, instead of a single u_c , there might be a whole interval $[u_-, u_+]$ where the rhs of [17] reaches the maximum.

Convexity implies that the values of λ or β for which the maxima in [16] or [17] are attained in intervals rather than in single points are rare (i.e., at most denumerably many): the interpretation is, in such cases, that the thermodynamic functions show discontinuities, and the corresponding phenomena are called phase transitions (see the next section).

For more details the reader is referred to Ruelle (1969) and Gallavotti (1999).

Phase Transitions and Boundary Conditions

The analysis in the last two sections of the relations between elements of ensembles of distributions describing macroscopic equilibrium states not only allows us to obtain mechanical models of thermodynamics but also shows that the models, for a given system, coincide at least as $\Omega \rightarrow \infty$. Furthermore, the equivalence between the thermodynamic functions computed via corresponding distributions in different ensembles can be extended to a full equivalence of the distributions.

If the maxima in [16] are attained at single points v_{gc} or u_c the equivalence should take place in the sense that a correspondence between $\mu_{\beta, \lambda}^{\text{gc}}, \mu_{\beta, v}^c, \mu_{u, v}^{\text{mc}}$ can be established so that, given any local observable $F(\mathbf{P}, \mathbf{Q})$, defined as an observable depending on (\mathbf{P}, \mathbf{Q}) only through the p_i, q_i with $q_i \in \Lambda$, where $\Lambda \subset \Omega$ is a finite region, has the same average with respect to corresponding distributions in the limit $\Omega \rightarrow \infty$.

The correspondence is established by considering $(\lambda, \beta) \leftrightarrow (\beta, v_{\text{gc}}) \leftrightarrow (u_{\text{mc}}, v)$, where v_{gc} is where the maximum in [16] is attained, $u_{\text{mc}} \equiv u_c$ is where the maximum in [17] is attained and $v_{\text{gc}} \equiv v$, (cf. also [19], [20]). This means that the limits

$$\begin{aligned} \lim_{V \rightarrow \infty} \int F(\mathbf{P}, \mathbf{Q}) \mu^a(d\mathbf{P} d\mathbf{Q}) &\stackrel{\text{def}}{=} \langle F \rangle_a \\ (a - \text{independent}), \quad a &= \text{gc}, c, \text{mc} \end{aligned} \quad [21]$$

coincide if the averages are evaluated by the distributions $\mu_{\beta, \lambda}^{\text{gc}}, \mu_{\beta, v_c}^c, \mu_{u_{\text{mc}}, v_{\text{mc}}}^{\text{mc}}$.

Exceptions to [21] are possible: and are certainly likely to occur at values of u, v where the maxima in [16] or [17] are attained in intervals rather than in isolated points; but this does not exhaust, in general, the cases in which [21] may not hold.

However, no case in which [21] fails has to be regarded as an exception. It rather signals that an interesting and important phenomenon occurs. To understand it properly, it is necessary to realize that the grand canonical, canonical, and microcanonical families of probability distributions are by far not the only ensembles of probability distributions whose elements can be considered to generate models of thermodynamics, that is, which are orthodic in the sense of the discussion in the section “Equivalence of ensembles.” More general families of orthodic statistical ensembles of probability

distributions can be very easily conceived. In particular:

Definition Consider the grand canonical, canonical, and microcanonical distributions associated with an energy function in which the potential energy contains, besides the interaction Φ between particles located inside the container, also the interaction energy $\Phi_{\text{in,out}}$ between particles inside the container and external particles, identical to the ones in the container but not allowed to move and fixed in positions such that in every unit cube Δ external to Ω there is a finite number of them bounded independently of Δ . Such configurations of external particles will be called “boundary conditions of fixed external particles.”

The thermodynamic limit with such boundary conditions is obtained by considering the grand canonical, canonical, and microcanonical distributions constructed with potential energy function $\Phi + \Phi_{\text{in,out}}$ in containers Ω of increasing size taking care that, while the size increases, the fixed particles that would become internal to Ω are eliminated. The argument used in the section “Thermodynamic limit” to show that the three models of thermodynamics, considered there, did define the same thermodynamic functions can be repeated to reach the conclusion that also the (infinitely many) “new” models of thermodynamics in fact give rise to the same thermodynamic functions and averages of local observables. Furthermore, the values of the limits corresponding to [13] can be computed using the new partition functions and coincide with the ones in [13] (i.e., they are independent of the boundary conditions).

However, it may happen, and in general it is the case, for many models and for particular values of the state parameters, that the limits in [21] do not coincide with the analogous limits computed in the new ensembles, that is, the averages of some local observables are unstable with respect to changes of boundary conditions with fixed particles.

There is a very natural interpretation of such apparent ambiguity of the various models of thermodynamics: namely, at the values of the parameters that are selected to describe the macroscopic states under consideration, there may correspond different equilibrium states with the same parameters. When the maximum in [16] is reached on an interval of densities, one should not think of any failure of the microscopic models for thermodynamics: rather one has to think that there are several states possible with the same β, λ and that they can be identified with the probability distributions obtained by forming the grand canonical,

canonical, or microcanonical distributions with different kinds of boundary conditions.

For instance, a boundary condition with high density may produce an equilibrium state with parameters β, λ which also has high density, i.e., the density ν_+^{-1} at the right extreme of the interval in which the maximum in [16] is attained, while using a low-density boundary condition the limit in [21] may describe the averages taken in a state with density ν_-^{-1} at the left extreme of the interval or, perhaps, with a density intermediate between the two extremes. Therefore, the following definition emerges.

Definition If the grand canonical distributions with parameters (β, λ) and different choices of fixed external particles boundary conditions generate for some local observable F average values which are different by more than a quantity $\delta > 0$ for all large enough volumes Ω then one says that the system has a phase transition at (β, λ) . This implies that the limits in [21], when existing, will depend on the boundary condition and their values will represent averages of the observables in “different phases.” A corresponding definition is given in the case of the canonical and microcanonical distributions when, given (β, ν) or (u, ν) , the limit in [21] depends on the boundary conditions for some F .

Remarks

1. The idea is that by fixing one of the thermodynamic ensembles and by varying the boundary conditions one can realize all possible states of equilibrium of the system that can exist with the given values of the parameters determining the state in the chosen ensemble (i.e., (β, λ) , (β, ν) , or (u, ν) in the grand canonical, canonical, or microcanonical cases, respectively).
2. The impression that in order to define a phase transition the thermodynamic limit is necessary is incorrect: the definition does not require considering the limit $\Omega \rightarrow \infty$. The phenomenon that occurs is that by changing boundary conditions the average of a local observable can change at least by amounts independent of the system size. Hence, occurrence of a phase transition is perfectly observable in finite volume: it suffices to check that by changing boundary conditions the average of some observable changes by an amount whose minimal size is volume independent. It is a manifestation of an instability of the averages with respect to changes in boundary conditions: an instability which does not fade away when the boundary recedes to infinity, i.e., boundary perturbations produce

bulk effects and at a phase transition the averages of the local observable, if existing at all, will exhibit a nontrivial dependence on the boundary conditions. This is also called “long range order.”

3. It is possible to show that when this happens then some thermodynamic function whose value is independent of the boundary condition (e.g., the free energy in the canonical distributions) has discontinuous derivatives in terms of the parameters of the ensemble. This is in fact one of the frequently-used alternative definitions of phase transitions: the latter two natural definitions of first-order phase transition are equivalent. However, it is very difficult to prove that a given system shows a phase transition. For instance, existence of a liquid–gas phase transition is still an open problem in systems of the type considered until the section “Lattice models” below.
4. A remarkable unification of the theory of the equilibrium ensembles emerges: all distributions of any ensemble describe equilibrium states. If a boundary condition is fixed once and for all, then some equilibrium states might fail to be described by an element of an ensemble. However, if all boundary conditions are allowed then all equilibrium states should be realizable in a given ensemble by varying the boundary conditions.
5. The analysis leads us to consider as completely equivalent without exceptions grand canonical, canonical, or microcanonical ensembles enlarged by adding to them the distributions with potential energy augmented by the interaction with fixed external particles.
6. The above picture is really proved only for special classes of models (typically in models in which particles are constrained to occupy points of a lattice and in systems with hard core interactions, $r_0 > 0$ in [14]) but it is believed to be correct in general. At least it is consistent with all that is known so far in classical statistical mechanics. The difficulty is that, conceivably, one might even need boundary conditions more complicated than the fixed particles boundary conditions (e.g., putting different particles outside, interacting with the system with an arbitrary potential, rather than via φ).

The discussion of the equivalence of the ensembles and the question of the importance of boundary conditions has already imposed the consideration of several limits as $\Omega \rightarrow \infty$. Occasionally, it will again come up. For conciseness, it is useful to set up a formal definition of equilibrium states of an infinite-volume system: although infinite volume is

an idealization void of physical reality, it is nevertheless useful to define such states because certain notions (e.g., that of pure state) can be sharply defined, with few words and avoiding wide circumvolutions, in terms of them. Therefore, let:

Definition An infinite-volume state with parameters $(\beta, \nu), (u, \nu)$ or (β, λ) is a collection of average values $F \rightarrow \langle F \rangle$ obtained, respectively, as limits of finite-volume averages $\langle F \rangle_{\Omega_n}$ defined from canonical, microcanonical, or grand canonical distributions in Ω_n with fixed parameters $(\beta, \nu), (u, \nu)$ or (β, λ) and with general boundary condition of fixed external particles, on sequences $\Omega_n \rightarrow \infty$ for which such limits exist simultaneously for all local observables F .

Having set the definition of infinite-volume state consider a local observable $G(X)$ and let $\tau_\xi G(X) = G(X + \xi), \xi \in \mathbb{R}^d$, with $X + \xi$ denoting the configuration X in which all particles are translated by ξ : then an infinite-volume state is called a pure state if for any pair of local observables F, G it is

$$\langle F \tau_\xi G \rangle - \langle F \rangle \langle \tau_\xi G \rangle \xrightarrow{\xi \rightarrow \infty} 0 \quad [22]$$

which is called a cluster property of the pair F, G .

The result alluded to in remark (6) is that at least in the case of hard-core systems (or of the simple lattice systems discussed in the section “Lattice models”) the infinite-volume equilibrium states in the above sense exhaust at least the totality of the infinite-volume pure states. Furthermore, the other states that can be obtained in the same way are convex combinations of the pure states, i.e., they are “statistical mixtures” of pure phases. Note that $\langle \tau_\xi G \rangle$ cannot be replaced, in general, by $\langle G \rangle$ because not all infinite-volume states are necessarily translation invariant and in simple cases (e.g., crystals) it is even possible that no translation-invariant state is a pure state.

Remarks

1. This means that, in the latter models, generalizing the boundary conditions, for example considering external particles to be not identical to the ones inside the system, using periodic or partially periodic boundary conditions, or the widely used alternative of introducing a small auxiliary potential and first taking the infinite-volume states in presence of it and then letting the potential vanish, does not enlarge further the set of states (but may sometimes be useful: an example of a study of a phase transition by using the latter method of small fields will be given in the section “Continuous symmetries: ‘no $d=2$ crystal’ theorem”).

2. If χ is the indicator function of a local event, it will make sense to consider the probability of occurrence of the event in an infinite-volume state defining it as $\langle \chi \rangle$. In particular, the probability density for finding p particles at $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, called the p -point correlation function, will thus be defined in an infinite-volume state. For instance, if the state is obtained as a limit of canonical states $\langle \cdot \rangle_{\Omega_n}$ with parameters $\beta, \rho, \rho = N_n/V_n$, in a sequence of containers Ω_n , then

$$\rho(\mathbf{x}) = \lim_n \left\langle \sum_{j=1}^{N_n} \delta(\mathbf{x} - \mathbf{q}_j) \right\rangle_{\Omega_n}$$

$$\rho(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \lim_n \left\langle \sum_{i_1, \dots, i_p}^{N_n} \prod_{j=1}^p \delta(\mathbf{x}_j - \mathbf{q}_{i_j}) \right\rangle_{\Omega_n}$$

where the sum is over the ordered p -ples (j_1, \dots, j_p) . Thus, the *pair correlation* $\rho(\mathbf{q}, \mathbf{q}')$ and its possible *cluster property* are

$$\rho(\mathbf{q}, \mathbf{q}')$$

$$\stackrel{\text{def}}{=} \lim_n \frac{\int_{\Omega_n} \exp(-\beta U(\mathbf{q}, \mathbf{q}', \mathbf{q}_1, \dots, \mathbf{q}_{N_n-2})) d\mathbf{q}_1 \cdots d\mathbf{q}_{N_n-2}}{(N_n-2)! Z_0^c(\beta, \rho, V_n)}$$

$$\rho(\mathbf{q}, (\mathbf{q}' + \boldsymbol{\xi})) - \rho(\mathbf{q})\rho(\mathbf{q}' + \boldsymbol{\xi}) \xrightarrow{\xi \rightarrow \infty} 0 \quad [23]$$

where

$$Z_0^c \stackrel{\text{def}}{=} \int e^{-\beta U(\mathbf{Q})} d\mathbf{Q}$$

is the ‘‘configurational’’ partition function.

The reader is referred to Ruelle (1969), Dobrushin (1968), Lanford and Ruelle (1969), and Gallavotti (1999).

Virial Theorem and Atomic Dimensions

For a long time it has been doubted that ‘‘just changing boundary conditions’’ could produce such dramatic changes as macroscopically different states (i.e., phase transitions in the sense of the definition in the last section). The first evidence that by taking the thermodynamic limit very regular analytic functions like $N^{-1} \log Z^c(\beta, N, V)$ (as a function of $\beta, v = V/N$) could develop, in the limit $\Omega \rightarrow \infty$, singularities like discontinuous derivatives (corresponding to the maximum in [16] being reached on a plateau and to a consequent existence of several pure phases) arose in the van der Waals’ theory of liquid–gas transition.

Consider a real gas with N identical particles with mass m in a container Ω with volume V . Let the force acting on the i th particle be \mathbf{f}_i ; multiplying

both sides of the equations of motion, $m\ddot{\mathbf{q}}_i = \mathbf{f}_i$, by $-(1/2)\mathbf{q}_i$ and summing over i , it follows that

$$-\frac{1}{2} \sum_{i=1}^N m \mathbf{q}_i \cdot \ddot{\mathbf{q}}_i = -\frac{1}{2} \sum_{i=1}^N \mathbf{q}_i \cdot \mathbf{f}_i \stackrel{\text{def}}{=} \frac{1}{2} C(\mathbf{q})$$

and the quantity $C(\mathbf{q})$ defines the *virial of the forces* in the configuration \mathbf{q} . Note that $C(\mathbf{q})$ is not translation invariant because of the presence of the forces due to the walls.

Writing the force \mathbf{f}_i as a sum of the internal and the external forces (due to the walls) the virial C can be expressed naturally as sum of the virial C_{int} of the internal forces (translation invariant) and of the virial C_{ext} of the external forces.

By dividing both sides of the definition of the virial by τ and integrating over the time interval $[0, \tau]$, one finds in the limit $\tau \rightarrow +\infty$, that is, up to quantities relatively infinitesimal as $\tau \rightarrow \infty$, that

$$\langle K \rangle = \frac{1}{2} \langle C \rangle \quad \text{and} \quad \langle C_{\text{ext}} \rangle = 3pV$$

where p is the pressure and V the volume. Hence

$$\langle K \rangle = \frac{3}{2} pV + \frac{1}{2} \langle C_{\text{int}} \rangle$$

or

$$\frac{1}{\beta} = pv + \frac{\langle C_{\text{int}} \rangle}{3N} \quad [24]$$

Equation [24] is Clausius’ *virial theorem*: in the case of no internal forces, it yields $\beta pv = 1$, the ideal-gas equation.

The internal virial C_{int} can be written, if $\mathbf{f}_{j \rightarrow i} = -\partial_{\mathbf{q}_i} \varphi(\mathbf{q}_i - \mathbf{q}_j)$, as

$$C_{\text{int}} = - \sum_{i=1}^N \sum_{i \neq j} \mathbf{f}_{j \rightarrow i} \cdot \mathbf{q}_i$$

$$\equiv - \sum_{i < j} \partial_{\mathbf{q}_i} \varphi(\mathbf{q}_i - \mathbf{q}_j) \cdot (\mathbf{q}_i - \mathbf{q}_j)$$

which shows that the contribution to the virial by the internal repulsive forces is negative while that of the attractive forces is positive. The average of C_{int} can be computed by the canonical distribution, which is convenient for the purpose. van der Waals first used the virial theorem to perform an actual computation of the corrections to the perfect-gas laws. Simply neglect the third-order term in the density and use the approximation $\rho(\mathbf{q}_1, \mathbf{q}_2) = \rho^2 e^{-\beta \varphi(\mathbf{q}_1 - \mathbf{q}_2)}$ for the pair correlation function, [23], then

$$\frac{1}{2} \langle C_{\text{int}} \rangle = V \frac{3}{2\beta} \rho^2 I(\beta) + VO(\rho^3) \quad [25]$$

where

$$I(\beta) = \frac{1}{2} \int (e^{-\beta\varphi(q)} - 1) d^3q$$

and the equation of state [24] becomes

$$pv + \frac{I(\beta)}{\beta v} + O(v^{-2}) = \beta^{-1}$$

For the purpose of illustration, the calculation of I can be performed approximately at “high temperature” (β small) in the case

$$\varphi(r) = 4\varepsilon \left(\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r} \right)^6 \right)$$

(the classical Lennard–Jones potential), $\varepsilon, r_0 > 0$. The result is

$$I \cong -(b - \beta a) \\ b = 4v_0, \quad a = \frac{32}{3} \varepsilon v_0, \quad v_0 = \frac{4\pi}{3} \left(\frac{r_0}{2} \right)^3$$

Hence,

$$pv + \frac{a}{v} - \frac{b}{\beta v} = \frac{1}{\beta} + O\left(\frac{1}{\beta v^2}\right) \\ \left(p + \frac{a}{v^2}\right)v = \left(1 + \frac{b}{v}\right)\frac{1}{\beta} = \frac{1}{1 - b/v} \frac{1}{\beta} + O\left(\frac{1}{\beta v^2}\right)$$

or

$$\left(p + \frac{a}{v^2}\right)(v - b)\beta = 1 + O(v^{-2}) \quad [26]$$

which gives the equation of state for $\beta\varepsilon \ll 1$. Equation [26] can be compared with the well-known empirical *van der Waals equation* of state:

$$\beta \left(p + \frac{a}{v^2} \right) (v - b) = 1$$

or

$$(p + An^2/V^2)(V - nB) = nRT \quad [27]$$

where, if N_A is Avogadro’s number, $A = aN_A^2$, $B = bN_A$, $R = k_B N_A$, $n = N/N_A$. It shows the possibility of accessing the microscopic parameters ε and r_0 of the potential φ via measurements detecting deviations from the *Boyle–Mariotte law*, $\beta pv = 1$, of the rarefied gases: $\varepsilon = 3a/8b = 3A/8BN_A$, $r_0 = (3b/2\pi)^{1/3} = (3B/2\pi N_A)^{1/3}$.

As a final comment, it is worth stressing that the virial theorem gives in principle the exact corrections to the equation of state, in a rather direct and simple form, as time averages of the virial of the internal forces. Since the virial of the internal forces is easy to calculate from the positions of the particles as a function of time, the theorem provides a method for computing the equation of state in

numerical simulations. In fact, this idea has been exploited in many numerical experiments, in which [24] plays a key role.

For more details, the reader is referred to Gallavotti (1999).

van der Waals Theory

Equation [27] is empirically used beyond its validity region (small density and small β) by regarding A, B as phenomenological parameters to be experimentally determined by measuring them near generic values of p, V, T . The measured values of A, B do not “usually vary too much” as functions of v, T and, apart from this small variability, the predictions of [27] have reasonably agreed with experience until, as experimental precision increased over the years, serious inadequacies eventually emerged.

Certain consequences of [27] are appealing: for example, Figure 1 shows that it does not give a p monotonic nonincreasing in v if the temperature is small enough. A critical temperature can be defined as the largest value, T_c , of the temperature below which the graph of p as a function of v is not monotonic decreasing; the critical volume V_c is the value of v at the horizontal inflection point occurring for $T = T_c$.

For $T < T_c$ the van der Waals interpretation of the equation of state is that the function $p(v)$ may describe metastable states while the actual equilibrium states would follow an equation with a monotonic dependence on v and $p(v)$ becoming horizontal in the coexistence region of specific volumes. The precise value of p where to draw the plateau (see Figure 1) would then be fixed by experiment or theoretically predicted via the simple rule that the plateau associated with the represented isotherm is drawn at a height such that the area of the two cycles in the resulting loop are equal.

This is *Maxwell’s rule*: obtained by assuming that the isotherm curve joining the extreme points of the plateau and the plateau itself define a cycle

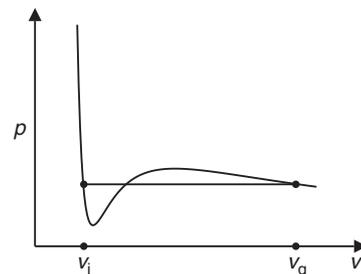


Figure 1 The van der Waals equation of state at a temperature $T < T_c$ where the pressure is not monotonic. The horizontal line illustrates the “Maxwell rule.”

(see [Figure 1](#)) representing a sequence of possible macroscopic equilibrium states (the ones corresponding to the plateau) or states with extremely long time of stability (“metastable”) represented by the curved part. This would be an isothermal Carnot cycle which, therefore, could not produce work: since the work produced in the cycle (i.e., $\oint p dv$) is the signed area enclosed by the cycle the rule just means that the area is zero. The argument is doubtful at least because it is not clear that the intermediate states with p increasing with v could be realized experimentally or could even be theoretically possible.

A striking prediction of [27], taken literally, is that the gas undergoes a gas–liquid phase transition with a critical point at a temperature T_c , volume v_c , and pressure p_c that can be computed via [27] and are given by $RT_c = 8A/27B$, $V_c = 3B$ ($n = 1$).

At the same time, the above prediction is interesting as it shows that there are simple relations between the critical parameters and the microscopic interaction constants, i.e., $\varepsilon \simeq k_B T_c$ and $r_0 \simeq (V_c/N_A)^{1/3}$; or more precisely $\varepsilon = 81k_B T_c/64$, $r_0 = (V_c/2\pi N_A)^{1/3}$ if a classical Lennard–Jones potential (i.e., $\varphi = 4\varepsilon ((r_0/|q|)^{12} - (r_0/|q|)^6)$; see the last section) is used for the interaction potential φ .

However, [27] cannot be accepted acritically not only because of the approximations (essentially the neglecting of $O(v^{-1})$ in the equation of state), but mainly because, as remarked above, for $T < T_c$ the function p is no longer monotonic in v as it must be; see comment following [19].

The van der Waals equation, refined and complemented by Maxwell’s rule, predicts the following behavior:

$$\begin{aligned} (p - p_c) &\propto (v - v_c)^\delta, \quad \delta = 3, \quad T = T_c \\ (v_g - v_l) &\propto (T_c - T)^\beta, \quad \beta = 1/2, \quad \text{for } T \rightarrow T_c^- \end{aligned} \quad [28]$$

which are in sharp contrast with the experimental data gathered in the twentieth century. For the simplest substances, one finds instead $\delta \cong 5$, $\beta \cong 1/3$.

Finally, blind faith in the equation of state [27] is untenable, last but not least, also because nothing in the analysis would change if the space dimension was $d = 2$ or $d = 1$: but for $d = 1$, it is easily proved that the system, if the interaction decays rapidly at infinity, does not undergo phase transitions (see next section).

In fact, it is now understood that van der Waals’ equation represents rigorously only a limiting situation, in which particles have a hard-core interaction (or a strongly repulsive one at close distance) and a further smooth interaction φ with very long range. More precisely, suppose that the part of the potential outside a hard-core radius $r_0 > 0$ is attractive (i.e., non-negative) and has the form $\gamma^d \varphi_1(\gamma^{-1}|q|) \leq 0$

and call $P_0(v)$ the (β -independent) product of β times the pressure of the hard-core system without any attractive tail ($P_0(v)$ is not explicitly known except if $d = 1$, in which case it is $P_0(v)(v - b) = 1$, $b = r_0$), and let

$$a = -\frac{1}{2} \int_{|q|>r_0} |\varphi_1(q)| dq$$

If $p(\beta, v; \gamma)$ is the pressure when $\gamma > 0$ then it can be proved that

$$\begin{aligned} \beta p(\beta, v) &\stackrel{\text{def}}{=} \lim_{\gamma \rightarrow 0} \beta p(\beta, v; \gamma) \\ &= \left[-\frac{\beta a}{v^2} + P_0(v) \right]_{\text{Maxwell's rule}} \end{aligned} \quad [29]$$

where the subscript means that the graph of $p(\beta, v)$ as a function of v is obtained from the function in square bracket by applying to it Maxwell’s rule, described above in the case of the van der Waals equation. Equation [29] reduces exactly to the van der Waals equation for $d = 1$, and for $d > 1$ it leads to an equation with identical critical behavior (even though $P_0(v)$ cannot be explicitly computed).

The reader is referred to [Lebowitz and Penrose \(1979\)](#) and [Gallavotti \(1999\)](#) for more details.

Absence of Phase Transitions: $d = 1$

One of the most quoted no-go theorems in statistical mechanics is that one-dimensional systems of particles interacting via short-range forces do not exhibit phase transitions (cf. the next section) unless the somewhat unphysical situation of having zero absolute temperature is considered. This is particularly easy to check in the case of “nearest-neighbor hard-core interactions.” Let the hard-core size be r_0 , so that the interaction potential $\varphi(r) = +\infty$ if $r \leq r_0$, and suppose also that $\varphi(r) \equiv 0$ if $r \geq 2r_0$. In this case, the thermodynamic functions can be exactly computed and checked to be analytic: hence the equation of state cannot have any phase transition plateau. This is a special case of *van Hove’s theorem* establishing smoothness of the equation of state for interactions extending beyond the nearest neighbor and rapidly decreasing at infinity.

If the definition of phase transition based on the sensitivity of the thermodynamic limit to variations of boundary conditions is adopted then a more general, conceptually simple, argument can be given to show that in one-dimensional systems there cannot be any phase transition if the potential energy of mutual interaction between a

configuration \mathcal{Q} of particles to the left of a reference particle (located at the origin O , say) and a configuration \mathcal{Q}' to the right of the particle (with $\mathcal{Q} \cup O \cup \mathcal{Q}'$ compatible with the hard cores) is uniformly bounded below. Then a mathematical proof can be devised showing that the influence of boundary conditions disappears as the boundaries recede to infinity. One also says that no long-range order can be established in a one-dimensional case, in the sense that one loses any trace of the boundary conditions imposed.

The analysis fails if the space dimension is ≥ 2 : in this case, even if the interaction is short-ranged, the energy of interaction between two regions of space separated by a boundary is of the order of the boundary area. Hence, one cannot bound above and below the probability of any two configurations in two half-spaces by the product of the probabilities of the two configurations, each computed as if the other was not there. This is because such a bound would be proportional to the exponential of the surface of separation, which tends to ∞ when the surface grows large. This means that we cannot consider, at least not in general, the configurations in the two half-spaces as independently distributed.

Analytically, a condition on the potential sufficient to imply that the energy between a configuration to the left and one to the right of the origin is bounded below, if $d=1$, is simply expressed by

$$\int_{r'}^{\infty} r|\varphi(r)|dr < +\infty \quad \text{for } r' > r_0$$

Therefore, in order to have phase transitions in $d=1$, a potential is needed that is “so long range” that it has a divergent first moment. It can be shown by counterexamples that if the latter condition fails there can be phase transitions even in $d=1$ systems.

The results just quoted also apply to discrete models like lattice gases or lattice spin models that will be considered later in the article.

For more details, we refer the reader to Landau and Lifschitz (1967), Dyson (1969), Gallavotti (1999), and Gallavotti *et al.* (2004).

Continuous Symmetries: “No $d=2$ Crystal” Theorem

A second case in which it is possible to rule out existence of phase transitions or at least of certain kinds of transitions arises when the system under analysis enjoys large symmetry. By symmetry is meant a group of transformations acting on the configurations and transforming each of them into a

configuration which, at least for one boundary condition (e.g., periodic or open), has the same energy.

A symmetry is said to be “continuous” if the group of transformations is a continuous group. For instance, continuous systems have translational symmetry if considered in a container Ω with periodic boundary conditions. Systems with “too much symmetry” sometimes cannot show phase transitions. For instance, the continuous translation symmetry of a gas in a container Ω with periodic boundary conditions is sufficient to exclude the possibility of crystallization in dimension $d=2$.

To discuss this, which is a prototype of a proof which can be used to infer absence of many transitions in systems with continuous symmetries, consider the translational symmetry and a potential satisfying, besides the usual [14] and with the symbols used in [14], the further property that $|q|^2 |\partial_{ij}^2 \varphi(q)| < \bar{B} |q|^{-(d+\varepsilon_0)}$, with $\varepsilon_0 > 0$, for some \bar{B} holds for $r_0 < |q| \leq R$. This is a very mild extra requirement (and it allows for a hard-core interaction).

Consider an “ideal crystal” on a square lattice (for simplicity) of spacing a , exactly fitting in its container Ω of side L assumed with periodic boundary conditions: so that $N=(L/a)^d$ is the number of particles and a^{-d} is the density, which is supposed to be smaller than the close packing density if the interaction φ has a hard core. The probability distribution of the particles is rather trivial:

$$\bar{\mu} = \sum_p \prod_n \delta(q_{p(n)} - a\mathbf{n}) \frac{d\mathcal{Q}}{N!}$$

the sum running over the permutations $\mathbf{m} \rightarrow p(\mathbf{m})$ of the sites $\mathbf{m} \in \Omega$, $\mathbf{m} \in \mathbb{Z}^d$, $0 < m_i \leq La^{-1}$. The density at \mathbf{q} is

$$\hat{\rho}(\mathbf{q}) = \sum_n \delta(\mathbf{q} - a\mathbf{n}) \equiv \left\langle \sum_{j=1}^N \delta(\mathbf{q} - \mathbf{q}_j) \right\rangle$$

and its Fourier transform is proportional to

$$\rho(\mathbf{k}) \stackrel{\text{def}}{=} \frac{1}{N} \left\langle \sum_j e^{-i\mathbf{k} \cdot \mathbf{q}_j} \right\rangle, \quad \mathbf{k} = \frac{2\pi}{L} \mathbf{n}, \quad \mathbf{n} \in \mathbb{Z}^d$$

$\rho(\mathbf{k})$ has value 1 for all \mathbf{k} of the form $\mathbf{K} = (2\pi/a)\mathbf{n}$ and $(1/N)O(\max_{c=1,2} |e^{i\mathbf{k} \cdot \mathbf{c}a} - 1|^{-2})$ otherwise. In presence of interaction, it has to be expected that, in a crystal state, $\rho(\mathbf{k})$ has peaks near the values \mathbf{K} : but the value of $\rho(\mathbf{k})$ can depend on the boundary conditions.

Since the system is translation invariant a crystal state defined as a state with a distribution “close” to $\bar{\mu}$,

i.e., with $\hat{\rho}(q)$ with peaks at the ideal lattice points $q = na$, cannot be realized under periodic boundary conditions, even when the system state is crystalline. To realize such a state, a symmetry-breaking term is needed in the interaction.

This can be done in several ways, for example, by changing the boundary condition. Such a choice implies a discussion of how much the boundary conditions influence the positions of the peaks of $\rho(k)$: for instance, it is not obvious that a boundary condition will not generate a state with a period different from the one that *a priori* has been selected for disproval (a possibility which would imply a reciprocal lattice of K 's different from the one considered to begin with). Therefore, here the choice will be to imagine that an external weak force with potential $\varepsilon W(q)$ acts forcing a symmetry breaking that favors the occupation of regions around the points of the ideal lattice (which would mark the average positions of the particles in the crystal state that is being sought). The proof (*Mermin's theorem*) that no equilibrium state with particles distribution "close" to $\bar{\mu}$, i.e., with peaks in place of the delta functions (see below), is essentially reproduced below.

Take $W(q) = \sum_{na \in \Omega} \chi(q - na)$, where $\chi(q) \leq 0$ is smooth and zero everywhere except in a small vicinity of the lattice points around which it decreases to some negative minimum keeping a rotation symmetry around them. The potential W is invariant under translations by the lattice steps. By the choice of the boundary condition and εW , the density $\tilde{\rho}_\varepsilon(q)$ will be periodic with period a so that $\rho_\varepsilon(k)$ will, possibly, not have a vanishing limit as $N \rightarrow \infty$ only if k is a reciprocal vector $K = (2\pi/a)n$. If the potential is $\varphi + \varepsilon W$ and if there exists a crystal state in which particles have higher probability of being near the lattice points na , it should be expected that for small $\varepsilon > 0$ the system will be found in a state with Fourier transform of the density, $\rho_\varepsilon(k)$, satisfying, for some vector $K \neq 0$ in the reciprocal lattice,

$$\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} |\rho_\varepsilon(K)| = r > 0 \quad [30]$$

that is, the requirement is that uniformly in $\varepsilon \rightarrow 0$ the Fourier transform of the density has a peak at some $K \neq 0$. Note that if k is not in the reciprocal lattice $\rho_\varepsilon(k) \xrightarrow{N \rightarrow \infty} 0$, being bounded above by

$$\frac{1}{N} O\left(\max_{j=1,2} |e^{ik_j a} - 1|^{-2}\right)$$

because $(1/N)\tilde{\rho}_\varepsilon$ is periodic and its integral over q is equal to 1. Hence, excluding the existence of a

crystal will be identified with the impossibility of the [30]. Other criteria can be imagined, for example, considering crystals with a lattice different from simple cubic, which lead to the same result by following the same technique. Nevertheless, it is not mathematically excluded (but unlikely) that, with some weaker existence definition, a crystal state could be possible even in two dimensions.

The following inequalities hold under the present assumptions on the potential and in the canonical distribution with periodic boundary conditions and parameters (β, ρ) , $\rho = a^{-3}$ in a box Ω with side multiple of a (so that $N = (La^{-1})^d$) and potential of interaction $\varphi + \varepsilon W$. The further assumption that the lattice na is not a close-packed lattice is (of course) necessary when the interaction potential has a hard core. Then, for suitable $B_0, B, B_1, B_2 > 0$, independent of N , and ε and for $|\mathbf{k}| < \pi/a$ and for all Ω (if $K \neq 0$)

$$\begin{aligned} \frac{1}{N} \left\langle \left| \sum_{j=1}^N e^{-i(\mathbf{k}+K) \cdot q_j} \right|^2 \right\rangle &\geq B \frac{(\rho_\varepsilon(K) + \rho_\varepsilon(K + 2K))^2}{B_1 K^2 + \varepsilon B_2} \\ \frac{1}{N} \sum_{\mathbf{k}} \gamma(\mathbf{k}) \frac{d\mathbf{k}}{N} \left\langle \left| \sum_{j=1}^N e^{-i(\mathbf{k}+K) \cdot q_j} \right|^2 \right\rangle &\leq B_0 < \infty \quad [31] \end{aligned}$$

where the averages are in the canonical distribution (β, ρ) with periodic boundary conditions and a symmetry-breaking potential $\varepsilon W(q)$; $\gamma(\mathbf{k}) \geq 0$ is an (arbitrary) smooth function vanishing for $2|\mathbf{k}| \geq \delta$ with $\delta < 2\pi/a$ and B_0 depends on γ . See Appendix 3 for a derivation of [31].

Multiplying both sides of the first equation in [31] by $N^{-1}\gamma(\mathbf{k})$ and summing over \mathbf{k} , the crystallinity condition in the form [30] implies

$$B_0 \geq Br^2 a^d \int_{|\mathbf{k}| < \delta} \frac{\gamma(\mathbf{k}) d\mathbf{k}}{K^2 B_1 + \varepsilon B_2}$$

For $d = 1, 2$ the integral diverges, as $\varepsilon^{-1/2}$ or $\log \varepsilon^{-1}$, respectively, implying $|\rho_\varepsilon(K)| \xrightarrow{\varepsilon \rightarrow 0} r = 0$: the criterion of crystallinity, [30] cannot be satisfied if $d = 1, 2$.

The above inequality is an example of a general class of inequalities called *infrared inequalities* stemming from another inequality called *Bogoliubov's inequality* (see Appendix 3), which lead to the proof that certain kinds of ordered phases cannot exist if the dimension of the ambient space is $d = 2$ when a finite volume, under suitable boundary conditions (e.g., periodic), shows a continuous symmetry. The excluded phenomenon is, more precisely, the non-existence of equilibrium states exhibiting, in the thermodynamic limit, a symmetry lower than the continuous symmetry holding in a finite volume.

In general, existence of thermodynamic equilibrium states with symmetry lower than the

symmetry enjoyed by the system in finite volume and under suitable boundary conditions is called a “spontaneous symmetry breaking.” It is yet another manifestation of instability with respect to changes in boundary conditions, hence its occurrence reveals a phase transition. There is a large class of systems for which an infrared inequality implies absence of spontaneous symmetry breaking: in most of the one- or two-dimensional systems a continuous symmetry cannot be spontaneously broken.

The limitation to dimension $d \leq 2$ is a strong limitation to the generality of the applicability of infrared theorems to exclude phase transitions. More precisely, systems can be divided into classes each of which has a “critical dimension” below which too much symmetry implies absence of phase transitions (or of certain kinds of phase transitions).

It should be stressed that, at the critical dimension, the symmetry breaking is usually so weakly forbidden that one might need astronomically large containers to destroy small effects (due to boundary conditions or to very small fields) which break the symmetry. For example, in the crystallization just discussed, the Fourier transform peaks are only bounded by $O(1/\sqrt{\log \varepsilon^{-1}})$. Hence, from a practical point of view, it might still be possible to have some kind of order even in large containers.

The reader is referred to Mermin (1968), Hohenberg (1969), and Ruelle (1969).

High Temperature and Small Density

There is another class of systems in which no phase transitions take place. These are the systems with stable and tempered interactions φ (e.g., those satisfying [14]) in the high-temperature and low-density region. The property is obtained by showing that the equation of state is analytic in the variables (β, ρ) near the origin $(0, 0)$.

A simple algorithm (*Mayer's series*) yields the coefficients of the virial series

$$\beta p(\beta, \rho) = \rho + \sum_{k=2}^{\infty} c_k(\beta) \rho^k$$

It has the drawback that the k th order coefficient $c_k(\beta)$ is expressed as a sum of many terms (a number growing more than exponentially fast in the order k) and it is not so easy (but possible) to show combinatorially that their sum is bounded exponentially in k if β is small enough. A more efficient approach leads quickly to the desired solution. Denoting $\Phi(q_1, \dots, q_n) \stackrel{\text{def}}{=} \sum_{i < j} \varphi(q_i - q_j)$, consider the (“spatial or configurational”) correlation functions

defined, in the grand canonical distribution with parameters β, λ (and empty boundary conditions), by

$$\rho_{\Omega}(q_1, \dots, q_n) \stackrel{\text{def}}{=} \frac{1}{Z^{gc}(\beta, \lambda, V)} \sum_{m=0}^{\infty} z^{n+m} \times \int_{\Omega} e^{-\beta \Phi(q_1, \dots, q_n, y_1, \dots, y_m)} \frac{dy_1 \cdots dy_m}{m!} \quad [32]$$

This is the probability density for finding particles with any momentum in the volume element $dq_1 \cdots dq_n$ (irrespective of where other particles are), and $z = e^{\beta \lambda (\sqrt{2\pi m} \beta^{-1} \hbar^{-2})^d}$ accounts for the integration over the momenta variables and is called the activity: it has the dimension of a density (cf. [23]).

Assuming that the potential has a hard core (for simplicity) of radius R , the interaction energy $\Phi_{q_1}(q_2, \dots, q_n)$ of a particle at q_1 with any number of other particles at q_2, \dots, q_n with $|q_i - q_j| > R$ is bounded below by $-B$ for some $B \geq 0$ (related but not equal to the B in [14]). The functions ρ_{Ω} will be regarded as a sequence of functions “of one, two, . . . particle positions”: $\rho_{\Omega} = \{\rho_{\Omega}(q_1, \dots, q_n)\}_{n=1}^{\infty}$ vanishing for $q_j \notin \Omega$. Then, one checks that

$$\rho_{\Omega}(q_1, \dots, q_n) = z \delta_{n,1} \chi_{\Omega}(q_1) + K \rho_{\Omega}(q_1, \dots, q_n) \quad [33a]$$

with

$$K \rho_{\Omega}(q_1, \dots, q_n) \stackrel{\text{def}}{=} e^{-\beta \Phi_{q_1}(q_2, \dots, q_n)} (\rho_{\Omega}(q_2, \dots, q_n) \delta_{n>1} + \sum_{s=1}^{\infty} \int_{\Omega} \frac{dy_1 \cdots dy_s}{s!} \prod_{k=1}^s (e^{-\beta \varphi(q_1 - y_k)} - 1) \times \rho_{\Omega}(q_2, \dots, q_n, y_1, \dots, y_s)) \quad [33b]$$

where $\delta_{n,1}, \delta_{n>1}$ are Kronecker deltas and $\chi_{\Omega}(q)$ is the indicator function of Ω . Equation [33] is called the *Kirkwood–Salzburg equation* for the family of correlation functions in Ω . The kernel K of the equations is independent of Ω , but the domain of integration is Ω .

Calling α_{Ω} the sequence of functions $\alpha_{\Omega}(q_1, \dots, q_n) \equiv 0$ if $n \neq 1$ and $\alpha_{\Omega}(q) = z \chi_{\Omega}(q)$, a recursive expansion arises, namely

$$\rho_{\Omega} = z \alpha_{\Omega} + z^2 K \alpha_{\Omega} + z^3 K^2 \alpha_{\Omega} + z^4 K^3 \alpha_{\Omega} + \cdots \quad [34]$$

It gives the correlation functions, provided the series converges. The inequality

$$|K^p \alpha_{\Omega}(q_1, \dots, q_n)| \leq e^{(2\beta B + 1)p} \left(\int |e^{-\beta \varphi(q)} - 1| dq \right)^p \stackrel{\text{def}}{=} e^{(2\beta B + 1)p} r(\beta)^{3p} \quad [35]$$

shows that the series [34], called Mayer’s series, converges if $|z| < e^{-(2\beta B + 1)} r(\beta)^{-3}$. Convergence is uniform (as $\Omega \rightarrow \infty$) and $(K^p) \alpha_{\Omega}(q_1, \dots, q_n)$ tends to a limit as $V \rightarrow \infty$ at fixed q_1, \dots, q_n , and the limit is simply $(K^p \alpha)(q_1, \dots, q_n)$, if $\alpha(q_1, \dots, q_n) \equiv 0$ for $n \neq 1$, and $\alpha(q_1) \equiv 1$. This is because the kernel K contains

the factors $(e^{-\beta\varphi(q_1-y)} - 1)$ which decay rapidly or, if φ has finite range, will eventually even vanish. It is also clear that $(K^p\alpha)(q_1, \dots, q_n)$ is translation invariant.

Hence, if $|z|e^{2\beta B+1}r(\beta)^3 < 1$, the limits, as $\Omega \rightarrow \infty$, of the correlation functions exist and can be computed by a convergent power series in z ; the correlation functions will be translation invariant (in the thermodynamic limit).

In particular, the one-point correlation function $\rho = \rho(q)$ is $\rho = z(1 + O(zr(\beta)^3))$, which, to lowest order in z , just shows that activity and density essentially coincide when they are small enough. Furthermore, $\beta p_\Omega = (1/V) \log Z^{\text{gc}}(\beta, \lambda, V)$ is such that

$$z \partial_z \beta p_\Omega = \frac{1}{V} \int \rho_\Omega(q) dq$$

(from the definition of ρ_Ω in [32]). Therefore,

$$\begin{aligned} \beta p(\beta, z) &= \lim_{V \rightarrow \infty} \frac{1}{V} \log Z^{\text{gc}}(\beta, \lambda, V) \\ &= \int_0^z \frac{dz'}{z'} \rho(\beta, z') \end{aligned} \quad [36]$$

and, since the density ρ is analytic in z as well and $\rho \simeq z$ for z small, the grand canonical pressure is analytic in the density and $\beta p = \rho(1 + O(\rho^2))$, at small density. In other words, the equation of state is, to lowest order, essentially the equation of a perfect gas. All quantities that are conceivably of some interest turn out to be analytic functions of temperature and density. The system is essentially a free gas and it has no phase transitions in the sense of a discontinuity or of a singularity in the dependence of a thermodynamic function in terms of others. Furthermore, the system cannot show phase transitions in the sense of sensitive dependence on boundary conditions of fixed external particles. This also follows, with some extra work, from the Kirkwood–Salzburg equations.

The reader is referred to Ruelle (1969) and Gallavotti (1969) for more details.

Lattice Models

The problem of proving the existence of phase transitions in models of homogeneous gases with pair interactions is still open. Therefore, it makes sense to study the problem of phase transitions in simpler models, tractable to some extent but nontrivial, and which are of practical interest in their own right.

The simplest models are the so-called lattice models in which particles are constrained to points of a lattice: they cannot move in the ordinary sense of the word (but, of course, they could jump) and

therefore their configurations do not contain momentum variables.

The interaction energy is just the potential energy, and ensembles are defined as collections of probability distributions on the position coordinates of the particle configurations. Usually, the potential is a pair potential decaying fast at ∞ and, often, with a hard-core forbidding double or higher occupancy of the same lattice site. For instance, the *lattice gas* with potential φ , in a cubic box Ω with $|\Omega| = V = L^d$ sites of a square lattice with mesh $a > 0$, is defined by the potential energy attributed to the configuration X of occupied distinct sites, i.e., subsets $X \subset \Omega$:

$$H(X) = - \sum_{(x,y) \in X} \varphi(x-y) \quad [37]$$

where the sum is over pairs of distinct points in X . The canonical ensemble and the grand canonical ensemble are the collections of distributions, parametrized by (β, ρ) , ($\rho = N/V$), or, respectively, by (β, λ) , attributing to X the probability

$$p_{\beta, \rho}(X) = \frac{e^{-\beta H(X)}}{Z_p^c(\beta, N, \Omega)} \delta_{|X|, N} \quad [38a]$$

or

$$p_{\beta, \lambda}(X) = \frac{e^{\beta \lambda |X|} e^{-\beta H(X)}}{Z_p^{\text{gc}}(\beta, \lambda, \Omega)} \quad [38b]$$

where the denominators are normalization factors that can, respectively, be called, in analogy with the theory of continuous systems, canonical and grand canonical partition functions; the subscript p stands for particles.

A lattice gas in which in each site there can be at most one particle can be regarded as a model for the distribution of a family of spins on a lattice. Such models are quite common and useful (e.g., they arise in studying systems with magnetic properties). Simply identify an “occupied” site with a “spin up” or $+$ and an “empty” site with a “spin down” or $-$ (say). If $\sigma = \{\sigma_x\}_{x \in \Omega}$ is a spin configuration, the energy of the configuration “for potential φ and magnetic field h ” will be

$$H(\sigma) = - \sum_{(x,y) \in \Omega} \varphi(x-y) \sigma_x \sigma_y - h \sum_x \sigma_x \quad [39]$$

with the sum running over pairs $(x, y) \in \Omega$ of distinct sites. If $\varphi(x-y) \equiv J_{xy} \geq 0$, the model is called a *ferromagnetic Ising model*. As in the case of continuous systems, it will be assumed to have a finite range for φ : that is, $\varphi(x) = 0$ for $|x| > R$, for some R , unless explicitly stated otherwise.

The canonical and grand canonical ensembles in the box Ω with respective parameters (β, m) or (β, b) will be defined as the probability distributions on the spin configurations $\sigma = \{\sigma_x\}_{x \in \Omega}$ with $\sum_{x \in \Omega} \sigma_x = M = mV$ or without constraint on M , respectively; hence,

$$\begin{aligned} p_{\beta, m}(\sigma) &= \frac{\exp\left(-\beta \sum_{(x, y)} \varphi(x - y) \sigma_x \sigma_y\right)}{Z_s^c(\beta, M, \Omega)} \\ p_{\beta, b}(\sigma) &= \frac{\exp\left(-\beta h \sum \sigma_x - \beta \sum_{(x, y)} \varphi(x - y) \sigma_x \sigma_y\right)}{Z_s^{\text{gc}}(\beta, b, \Omega)} \end{aligned} \quad [40]$$

where the denominators are normalization factors again called, respectively, the canonical and grand canonical partition functions. As in the study of the previous continuous systems, canonical and grand canonical ensembles with “external fixed particle configurations” can be defined together with the corresponding ensembles with “external fixed spin configurations”; the subscript s stands for spins.

For each configuration $X \subset \Omega$ of a lattice gas, let $\{n_x\}$ be $n_x = 1$ if $x \in X$ and $n_x = 0$ if $x \notin X$. Then the transformation $\sigma_x = 2n_x - 1$ establishes a correspondence between lattice gas and spin distributions. In the correspondence, the potential $\varphi(x - y)$ of the lattice gas generates a potential $(1/4)\varphi(x - y)$ for the corresponding spin system and the chemical potential λ for the lattice gas is associated with a magnetic field h for the spin system with $h = (1/2)(\lambda + \sum_{x \neq 0} \varphi(x))$.

The correspondence between boundary conditions is natural: for instance, a boundary condition for the lattice gas in which all external sites are occupied becomes a boundary condition in which external sites contain a spin $+$. The close relation between lattice gas and spin systems permits switching from one to the other with little discussion.

In the case of spin systems, empty boundary conditions are often considered (no spins outside Ω). In lattice gases and spin systems (as well as in continuum systems), often periodic and semiperiodic boundary conditions are considered (i.e., periodic in one or more directions and with empty or fixed external particles or spins in the others).

Thermodynamic limits for the partition functions

$$\begin{aligned} -\beta f(\beta, \nu) &= \lim_{\substack{\Omega \rightarrow \infty \\ V/N = \nu}} \frac{1}{N} \log Z_p^c(\beta, N, \Omega) \\ \beta p(\beta, \lambda) &= \lim_{\Omega \rightarrow \infty} \frac{1}{V} \log Z_p^{\text{gc}}(\beta, \lambda, \Omega) \\ -\beta g(\beta, m) &= \lim_{\substack{\Omega \rightarrow \infty \\ M/V = m}} \frac{1}{V} \log Z_s^c(\beta, M, \Omega) \\ \beta f(\beta, h) &= \lim_{\Omega \rightarrow \infty} \frac{1}{V} \log Z_s^{\text{gc}}(\beta, h, \Omega) \end{aligned} \quad [41]$$

can be shown to exist by a method similar to the one discussed in Appendix 2. They have convexity and continuity properties as in the cases of the continuum systems. In the case of a lattice gas, the f, p functions are still interpreted as free energy and pressure, respectively. In the case of spin, $f(\beta, h)$ has the interpretation of magnetic free energy, while $g(\beta, m)$ does not have a special name in the thermodynamics of magnetic systems. As in the continuum systems, it is occasionally useful to define infinite-volume equilibrium states:

Definition An infinite-volume state with parameters (β, b) or (β, m) is a collection of average values $F \rightarrow \langle F \rangle$ obtained, respectively, as limits of finite-volume averages $\langle F \rangle_{\Omega_n}$ defined from canonical or grand canonical distributions in Ω_n with fixed parameters (β, b) or (β, m) , or (u, ν) and with general boundary condition of fixed external spins or empty sites, on sequences $\Omega_n \rightarrow \infty$ for which such limits exist simultaneously for all local observables F .

This is taken verbatim from the definition in the section “Phase transitions and boundary conditions.” In this way, it makes sense to define the spin correlation functions for $X = (\xi_1, \dots, \xi_n)$ as $\langle \sigma_X \rangle$ if $\sigma_X = \prod_j \sigma_{\xi_j}$. For instance, we shall call $\rho(\xi_1, \xi_2) \stackrel{\text{def}}{=} \langle \sigma_{\xi_1} \sigma_{\xi_2} \rangle$ and a pure phase can be defined as an infinite-volume state such that

$$\langle \sigma_X \sigma_{Y+\xi} \rangle - \langle \sigma_X \rangle \langle \sigma_{Y+\xi} \rangle \xrightarrow[\xi \rightarrow \infty]{} 0 \quad [42]$$

Again, for more details, we refer the reader to [Ruelle \(1969\)](#) and [Gallavotti \(1969\)](#).

Thermodynamic Limits and Inequalities

An interesting property of lattice systems is that it is possible to study delicate questions like the existence of infinite-volume states in some (moderate) generality. A typical tool is the use of inequalities. As the simplest example of a vast class of inequalities, consider the ferromagnetic Ising model with some finite (but arbitrary) range interaction $J_{xy} \geq 0$ in a field $h_x \geq 0$: J, h may even be not translationally invariant. Then the average of $\sigma_X \stackrel{\text{def}}{=} \sigma_{x_1} \sigma_{x_2} \cdots \sigma_{x_n}$, $X = (x_1, \dots, x_n)$, in a state with “empty boundary conditions” (i.e., no external spins) satisfies the inequalities

$$\langle \sigma_X \rangle, \partial_{h_x} \langle \sigma_X \rangle, \partial_{J_{xy}} \langle \sigma_X \rangle \geq 0 \quad X = (x_1, \dots, x_n)$$

More generally, let $H(\sigma)$ in [39] be replaced by $H(\sigma) = -\sum_X J_X \sigma_X$ with $J_X \geq 0$ and X can be any finite set; then, if $Y = (y_1, \dots, y_n)$, $X = (x_1, \dots, x_n)$, the following *Griffiths inequalities* hold:

$$\langle \sigma_X \rangle \geq 0, \quad \partial_{J_Y} \langle \sigma_X \rangle \equiv \langle \sigma_X \sigma_Y \rangle - \langle \sigma_X \rangle \langle \sigma_Y \rangle \geq 0 \quad [43]$$

The inequalities can be used to check, in ferromagnetic Ising models, [39], existence of infinite-volume states (cf. the sections “Phase transitions and boundary conditions” and “Lattice models”) obtained by fixing the boundary condition \mathcal{B} to be either “all external spins +” or “all external sites empty.” If $\langle F \rangle_{\mathcal{B}, \Omega}$ denotes the grand canonical average with boundary condition \mathcal{B} and any fixed $\beta, h > 0$, this means that for all local observables $F(\sigma_\Lambda)$ (i.e., for all F depending on the spin configuration in any fixed region Λ) all the following limits exist:

$$\lim_{\Omega \rightarrow \infty} \langle F \rangle_{\mathcal{B}, \Omega} = \langle F \rangle_{\mathcal{B}} \quad [44]$$

The reason is that the inequalities [43] imply that all averages $\langle \sigma_X \rangle_{\mathcal{B}, \Omega}$ are monotonic in Ω for all fixed $X \subset \Omega$: so the limit [44] exists for $F(\sigma) = \sigma_X$. Hence, it exists for all F 's depending only on finitely many spins, because any local function F “measurable in Λ ” can be expressed (uniquely) as a linear combination of functions σ_X with $X \subseteq \Lambda$.

Monotonicity with empty boundary conditions is seen by considering the sites outside Ω and in a region Ω' with side one unit larger than that of Ω and imagining that the couplings J_X with $X \subset \Omega'$ but $X \not\subset \Omega$ vanish. Then, $\langle \sigma_X \rangle_{\Omega'} \geq \langle \sigma_X \rangle_{\Omega}$, because $\langle \sigma_X \rangle_{\Omega'}$ is an average computed with a distribution corresponding to an energy with the couplings J_X with $X \not\subset \Omega$, but $X \subset \Omega'$, changed from 0 to $J_X \geq 0$.

Likewise, if the boundary condition is +, then enlarging the box from Ω to Ω' corresponds to decreasing an external field h acting on the external spins from $+\infty$ (which would force all external spins to be +) to a finite value $h \geq 0$: so, increasing the box Ω causes $\langle \sigma_X \rangle_{+, \Omega}$ to decrease. Therefore, as Ω increases, Ising ferromagnets spin correlations increase if the boundary condition is empty and decrease if it is +.

The inequalities can be used in similar ways to prove that the infinite-volume states obtained from + or empty boundary conditions are translation invariant; and that in zero external field, $h=0$, the + and – boundary conditions generate pure states if the interaction potential is only a pair ferromagnetic interaction.

There are many other important inequalities which can be used to prove several existence theorems along very simple paths. Unfortunately, their use is mostly restricted to lattice systems and requires very special assumptions on the energy (e.g., ferromagnetic interactions in the above example). The quoted examples were among the first discovered and provide a way to exhibit nontrivial thermodynamic limits and pure states.

For more details, see Ruelle (1969), Lebowitz (1974), Gallavotti (1999), Lieb and Thirring (2001), and Lieb (2002).

Symmetry-Breaking Phase Transitions

The simplest phase transitions (see the section “Phase transitions and boundary conditions”) are symmetry-breaking transitions in lattice systems: they take place when the energy of the system in a container Ω and with some special boundary condition (e.g., periodic, antiperiodic, or empty) is invariant with respect to the action of a group \mathcal{G} on phase space. This means that on the points x of phase space acts a group of transformations \mathcal{G} so that with each $\gamma \in \mathcal{G}$ is associated a map $x \rightarrow x\gamma$ which transforms x into $x\gamma$ respecting the composition law in \mathcal{G} , that is, $(x\gamma)\gamma' \equiv x(\gamma\gamma')$. If F is an observable, the action of the group on phase space induces an action on the observable F changing $F(x)$ into $F_\gamma(x) \stackrel{\text{def}}{=} F(x\gamma^{-1})$.

A symmetry-breaking transition occurs when, by fixing suitable boundary conditions and taking the thermodynamic limit, a state $F \rightarrow \langle F \rangle$ is obtained in which some local observable shows a nonsymmetric average $\langle F \rangle \neq \langle F_\gamma \rangle$ for some γ .

An example is provided by the “nearest-neighbor ferromagnetic Ising model” on a d-dimensional lattice with energy function given by [39] with $h=0$ and $\varphi(x-y) \equiv 0$ unless $|x-y|=1$, i.e., unless x, y are nearest neighbors, in which case $\varphi(x-y) = J > 0$. With periodic or empty boundary conditions, it exhibits a discrete “up–down” symmetry $\sigma \rightarrow -\sigma$.

Instability with respect to boundary conditions can be revealed by considering the two boundary conditions, denoted + or –, in which the lattice sites outside the container Ω are either occupied by spins + or by spins –. Consider also, for later reference, (1) the boundary conditions in which the boundary spins in the upper half of the boundary are + and the ones in the lower part are –: call this the \pm -boundary condition (see Figure 2); or (2) the boundary conditions in

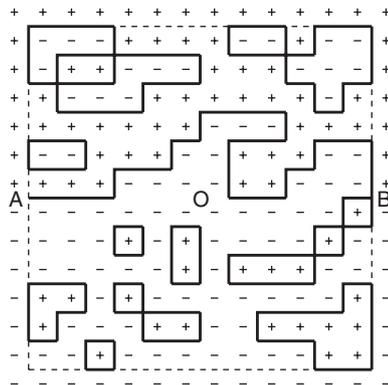


Figure 2 The dashed line is the boundary of Ω ; the outer spins correspond to the \pm boundary condition. The points A, B are points where an open “line” λ ends.

which some of the opposite sides of Ω are identified while $+$ or $-$ conditions are assigned on the remaining sides: call these “cylindrical or semiperiodic boundary conditions.”

A new description of the spin configurations is useful: given σ , draw a unit segment perpendicular to the center of each bond b having opposite spins at its extremes. An example of this construction is provided by **Figure 2** for the boundary condition \pm .

The set of segments can be grouped into lines separating regions where the spins are positive from regions where they are negative. If the boundary condition is $+$ or $-$, the lines form “closed polygons”, whereas, if the condition is \pm , there is also a single polygon λ_1 which is not closed (as in **Figure 2**). If the boundary condition is periodic or cylindrical, all polygons are closed but some may “go around” Ω . The polygons are also called “contours” and the length of a polygon γ will be denoted $|\gamma|$.

The correspondence $(\gamma_1, \gamma_2, \dots, \gamma_n, \lambda_1) \longleftrightarrow \sigma$, for the boundary condition \pm or, for the boundary condition $+$ (or $-$), $\sigma \longleftrightarrow (\gamma_1, \dots, \gamma_n)$ is one-to-one and, if $h=0$, the energy $H_\Omega(\sigma)$ of a configuration is higher than $-J \times (\text{number of bonds in } \Omega)$ by an amount $2J(|\lambda_1| + \sum_i |\gamma_i|)$ or, respectively, $2J \sum_i |\gamma_i|$. The grand canonical probability of each spin configuration is therefore proportional, if $h=0$, respectively, to

$$e^{-2\beta J(|\lambda_1| + \sum_i |\gamma_i|)} \quad \text{or} \quad e^{-2\beta J \sum_i |\gamma_i|} \quad [45]$$

and the “up–down” symmetry is clearly reflected by [45].

The average $\langle \sigma_x \rangle_{\Omega,+}$ of σ_+ with $+$ boundary conditions is given by $\langle \sigma_x \rangle_{\Omega,+} = 1 - 2P_{\Omega,+}(-)$, where $P_{\Omega,+}(-)$ is the probability that the spin σ_x is -1 . If the site x is occupied by a negative spin then the point x is inside some contour γ associated with the spin configuration σ under consideration. Hence, if $\rho(\gamma)$ is the probability that a given contour belongs to the set of contours describing a configuration σ , it is $P_{\Omega,+}(-) \leq \sum_{\gamma \ni x} \rho(\gamma)$ where $\gamma \ni x$ means that γ “surrounds” x .

If $\Gamma = (\gamma_1, \dots, \gamma_n)$ is a spin configuration and if the symbol $\Gamma \text{ comp } \gamma$ means that the contour γ is “disjoint” from $\gamma_1, \dots, \gamma_n$ (i.e., $\{\gamma \cup \Gamma\}$ is a new spin configuration), then

$$\begin{aligned} \rho(\gamma) &= \frac{\sum_{\Gamma \ni \gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}}{\sum_{\Gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}} \\ &\equiv e^{-2\beta J |\gamma|} \frac{\sum_{\Gamma \text{ comp } \gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}}{\sum_{\Gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}} \\ &\leq e^{-2\beta J |\gamma|} \end{aligned} \quad [46]$$

because the last ratio in [46] does not exceed 1. Note that there are $>3^p$ different shapes of γ with perimeter p and at most p^2 congruent γ 's containing x ; therefore, the probability that the spin at x is $-$ when the boundary condition is $+$ satisfies the inequality

$$P_{\Omega,+}(-) \leq \sum_{p=4}^{\infty} p^2 3^p e^{-2\beta J p} \xrightarrow{\beta \rightarrow \infty} 0$$

This probability can be made arbitrarily small so that $\langle \sigma_x \rangle_{\Omega,+}$ is estimated by a quantity which is as close to 1 as desired provided β is large enough and the closeness of $\langle \sigma_x \rangle_{\Omega,+}$ to 1 is estimated by a quantity which is both x and Ω independent.

A similar argument for the $(-)$ -boundary condition, or the remark that for $h=0$ it is $\langle \sigma_x \rangle_{\Omega,-} = -\langle \sigma_x \rangle_{\Omega,+}$, leads to conclude that, at large β , $\langle \sigma_x \rangle_{\Omega,-} \neq \langle \sigma_x \rangle_{\Omega,+}$ and the difference between the two quantities is positive uniformly in Ω . This is the proof (*Peierls' theorem*) of the fact that there is, if β is large, a strong instability, of the magnetization with respect to the boundary conditions, i.e., the nearest-neighbor Ising model in dimension 2 (or greater, by an identical argument) has a phase transition. If the dimension is 1, the argument clearly fails and no phase transition occurs (see the section “Absence of phase transitions: $d=1$ ”).

For more details, see [Gallavotti \(1999\)](#).

Finite-Volume Effects

The description in the last section of the phase transition in the nearest-neighbor Ising model can be made more precise both from physical and mathematical points of view giving insights into the nature of the phase transitions. Assume that the boundary condition is the $(+)$ -boundary condition and describe a spin configuration σ by means of the associated closed disjoint polygons $(\gamma_1, \dots, \gamma_n)$. Attribute to $\sigma = (\gamma_1, \dots, \gamma_n)$ a probability proportional to [45]. Then the following *Minlos–Sinai's theorem* holds:

Theorem *If β is large enough there exist $C > 0$, $\rho(\gamma) > 0$ with $\rho(\gamma) \leq e^{-2\beta J |\gamma|}$ and such that a spin configuration σ randomly chosen out of the grand canonical distribution with $+$ boundary conditions and $h=0$ will contain, with probability approaching 1 as $\Omega \rightarrow \infty$, a number $K_{(\gamma)}(\sigma)$ of contours congruent to γ such that*

$$|K_{(\gamma)}(\sigma) - \rho(\gamma)|\Omega| \leq C\sqrt{|\Omega|} e^{-\beta J |\gamma|} \quad [47]$$

and this relation holds simultaneously for all γ 's.

Thus, there are very few contours (and the larger they are the smaller is, in absolute and relative value, their number): a typical spin configuration in the grand canonical ensemble with (+)-boundary conditions is such that the large majority of the spins is “positive” and, in the “sea” of positive spins, there are a few negative spins distributed in small and rare regions (their number, however, is still of order of $|\Omega|$).

Another consequence of the analysis in the last section concerns the the approximate equation of state near the phase transition region at low temperatures and finite Ω . If Ω is finite, the graph of h versus $m_\Omega(\beta, h)$ will have a rather different behavior depending on the possible boundary conditions. For example, if the boundary condition is (+) or (−), one gets, respectively, the results depicted in **Figure 3a and 3b**, where $m^*(\beta)$ denotes the spontaneous magnetization (i.e., $m^*(\beta) \stackrel{\text{def}}{=} \lim_{h \rightarrow 0^+} \lim_{\Omega \rightarrow \infty} m_\Omega(\beta, h)$).

With periodic or empty boundary conditions, the diagram changes as in **Figure 4**. The thermodynamic limit $m(\beta, h) = \lim_{\Omega \rightarrow \infty} m_\Omega(\beta, h)$ exists for all $h \neq 0$ and the resulting graph is in **Figure 4b**, which shows that at $h=0$ the limit is discontinuous. It can be proved, if β is large enough, that $\infty > \lim_{h \rightarrow 0^+} \partial_h m(\beta, h) = \chi(\beta) > 0$ (i.e., the angle between the vertical part of the graph and the rest is sharp).

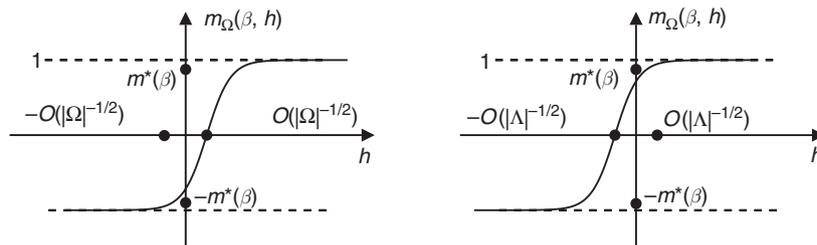
Furthermore, it can be proved that $m(\beta, h)$ is analytic in h for $h \neq 0$. If β is small enough,

analyticity holds at all h . For β large, the function $f(\beta, h)$ has an essential singularity at $h=0$: a result that can be interpreted as excluding a naive theory of metastability as a description of states governed by an equation of state obtained from an analytic continuation to negative values of h of $f(\beta, h)$.

The above considerations and results further clarify the meaning of a phase transition for a finite system. For more details, we refer the reader to [Gallavotti \(1999\)](#) and [Friedli and Pfister \(2004\)](#).

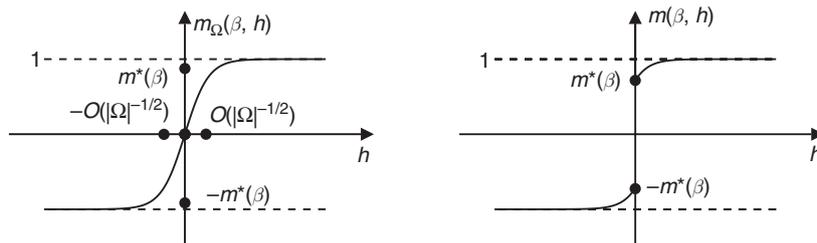
Beyond Low Temperatures (Ferromagnetic Ising Model)

A limitation of the results discussed above is the condition of low temperature (“ β large enough”). A natural problem is to go beyond the low-temperature region and to describe fully the phenomena in the region where boundary condition instability takes place and first develops. A number of interesting partial results are known, which considerably improve the picture emerging from the previous analysis. A striking list, but far from exhaustive, of such results follows and focuses on the properties of ferromagnetic Ising spin systems. The reason for restricting to such cases is that they are simple enough to allow a rather fine analysis, which sheds considerable light on the structure of statistical mechanics suggesting precise formulation



(a) (b)

Figure 3 The h vs $m_\Omega(\beta, h)$ graphs for Ω finite and (a) + and (b) − conditions.



(a) (b)

Figure 4 (a) The h vs $m_\Omega(\beta, h)$ graph for periodic or empty boundary conditions. (b) The discontinuity (at $h=0$) of the thermodynamic limit.

of the problems that it would be desirable to understand in more general systems.

1. Let $z \stackrel{\text{def}}{=} e^{\beta h}$ and consider that the product of z^V (V is the number of sites $|\Omega|$ of Ω) times the partition function with periodic or perfect-wall boundary conditions and with finite-range ferromagnetic interaction, not necessarily nearest-neighbor; a polynomial in z (of degree $2V$) is thus obtained. Its zeros lie on the unit circle $|z|=1$: this is *Lee–Yang’s theorem*. It implies that the only singularities of $f(\beta, h)$ in the region $0 < \beta < \infty, -\infty < h < +\infty$ can be found at $h=0$.

A singularity can appear only if the point $z=1$ is an accumulation point of the limiting distribution (as $\Omega \rightarrow \infty$) of the zeros on the unit circle: if the zeros are z_1, \dots, z_{2V} then

$$\begin{aligned} & \frac{1}{V} \log z^V Z(\beta, h, \Omega, \text{periodic}) \\ &= 2\beta J + \beta h + \frac{1}{V} \sum_{i=1}^{2V} \log(z - z_i) \end{aligned}$$

and if

$$V^{-1} \times (\text{number of zeros of the form } z_j = e^{i\theta_j}, \theta \leq \theta_j \leq \theta + d\theta) \xrightarrow{\Omega \rightarrow \infty} \frac{d\rho_\beta(\theta)}{2\pi}$$

it is

$$\beta f(\beta, h) = 2\beta J + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(z - e^{i\theta}) d\rho_\beta(\theta) \quad [48]$$

The existence of the measure $d\rho_\beta(\theta)$ follows from the existence of the thermodynamic limit: but $d\rho_\beta(\theta)$ is not necessarily $d\theta$ -continuous, i.e., not necessarily proportional to $d\theta$.

2. It can be shown that, with not necessarily a nearest-neighbor interaction, the zeros of the partition function do not move too much under small perturbations of the potential even if one perturbs the energy (at perfect-wall or periodic boundary conditions) into

$$\begin{aligned} H'_\Omega(\boldsymbol{\sigma}) &= H_\Omega(\boldsymbol{\sigma}) + (\delta H_\Omega)(\boldsymbol{\sigma}) \\ (\delta H_\Omega)(\boldsymbol{\sigma}) &= \sum_{X \subset \Omega} J'(X) \sigma_X \end{aligned} \quad [49]$$

where $J'(X)$ is very general and defined on subsets $X = (x_1, \dots, x_k) \subset \Omega$ such that the quantity $\|J'\| = \sup_{y \in Z^d} \sum_{y \in X} |J'(X)|$ is small enough. More precisely, with a ferromagnetic pair potential J fixed, suppose that one knows that, when $J'=0$, the partition function zeros in the variable $z = e^{\beta h}$ lie in a certain closed set N (of

the unit circle) in the z -plane. Then, if $J' \neq 0$, they lie in a closed set N^1 , Ω -independent and contained in a neighborhood of N of width shrinking to 0 when $\|J'\| \rightarrow 0$. This allows to establish various relations between analyticity properties and boundary condition instability as described in (3) below.

3. In the ferromagnetic Ising model, with not necessarily a nearest-neighbor interaction, one says that there is a gap around 0 if $d\rho_\beta(\theta) = 0$ near $\theta = 0$. It can be shown that if β is small enough there is a gap for all h of width *uniform* in h .
4. Another question is whether the boundary condition instability is always revealed by the one-spin correlation function (i.e., by the magnetization) or whether it might be shown only by some correlation functions of higher order. It can be proved that no boundary condition instability occurs for $h \neq 0$; at $h=0$ it is possible only if

$$\lim_{h \rightarrow 0^-} m(\beta, h) \neq \lim_{h \rightarrow 0^+} m(\beta, h) \quad [50]$$

5. A consequence of the Griffiths’ inequalities (cf. the section “Thermodynamic limits and inequalities”) is that if [50] is true for a given β_0 then it is true for all $\beta > \beta_0$. Therefore, item (4) leads to a natural definition of the critical temperature T_c as the least upper bound of the T ’s such that [50] holds ($k_B T = \beta^{-1}$).
6. If $d=2$ the free energy of the nearest-neighbor ferromagnetic Ising model has a singularity at β_c and the value of β_c is known exactly from the exact solutions of the model: $m(\beta, 0^+) \stackrel{\text{def}}{=} m^*(\beta) \equiv (1 - \sinh^4 2\beta J)^{1/8}$. The location and nature of the singularities of $f(\beta, 0)$ as a function of β remains an open question for $d=3$. In particular, the question whether there is a singularity of $f(\beta, 0)$ at $\beta = \beta_c$ is open.
7. For $\beta < \beta_c$ there is instability with respect to boundary conditions (see (6) above) and a natural question is: how many “pure” phases can exist in the ferromagnetic Ising model? (cf. the section “Phase transitions and boundary conditions,” eqn [22]). Intuition suggests that there should be only two phases: the positively magnetized and the negatively magnetized ones.

One has to distinguish between translation-invariant pure phases and non-translation-invariant ones. It can be proved that, in the case of the two-dimensional nearest-neighbor ferromagnetic Ising models, all infinite-volume states (cf. the section “Lattice models”) are translationally invariant. Furthermore, they can be obtained by

considering just the two boundary conditions + and -: the latter states are also pure states for models with non-nearest-neighbor ferromagnetic interaction. The solution of this problem has led to the introduction of many new ideas and techniques in statistical mechanics and probability theory.

8. In any dimension $d \geq 2$, for β large enough, it can be proved that the nearest-neighbor Ising model has only two translation-invariant phases. If the dimension is ≥ 3 and β is large, the + and - phases exhaust the set of translation-invariant pure phases but there exist non-translation-invariant phases. For β close to β_c , however, the question is much more difficult.

For more details, see Onsager (1944), Lee and Yang (1952), Ruelle (1971), Sinai (1991), Gallavotti (1999), Aizenman (1980), Higuchi (1981), and Friedli and Pfister (2004).

Geometry of Phase Coexistence

Intuition about the phenomena connected with the classical phase transitions is usually based on the properties of the liquid–gas phase transition; this transition is usually experimentally investigated in situations in which the total number of particles is fixed (canonical ensemble) and in presence of an external field (gravity).

The importance of such experimental conditions is obvious; the external field produces a nontranslationally invariant situation and the corresponding separation of the two phases. The fact that the number of particles is fixed determines, on the other hand, the fraction of volume occupied by each of the two phases.

Once more, consider the nearest-neighbor ferromagnetic Ising model: the results available for it can be used to obtain a clear picture of the solution to problems that one would like to solve but which in most other models are intractable with present-day techniques.

It will be convenient to discuss phase coexistence in the canonical ensemble distributions on configurations of fixed total magnetization $M = mV$ (see the section “Lattice models”; [40]). Let β be large enough to be in the two-phase region and, for a fixed $\alpha \in (0, 1)$, let

$$\begin{aligned} m &= \alpha m^*(\beta) + (1 - \alpha)(-m^*(\beta)) \\ &= (1 - 2\alpha)m^*(\beta) \end{aligned} \quad [51]$$

that is, m is in the vertical part of the diagram $m = m(\beta, h)$ at β fixed (see Figure 4).

Fixing m as in [51] does not yet determine the separation of the phases in two different regions; for this effect, it will be necessary to introduce some

external cause favoring the occupation of a part of the volume by a single phase. Such an asymmetry can be obtained in at least two ways: through a weak uniform external field (in complete analogy with the gravitational field in the liquid–vapor transition) or through an asymmetric field acting only on boundary spins. The latter should have the same qualitative effect as the former, because in a phase transition region a boundary perturbation produces volume effects (see sections “Phase transitions and inequalities” and “Symmetry-breaking phase transitions”). From a mathematical point of view, it is simpler to use a boundary asymmetry to produce phase separations and the simplest geometry is obtained by considering \pm -cylindrical or $++$ -cylindrical boundary conditions: this means $++$ or \pm boundary conditions periodic in one direction (e.g., in Figure 2 imagine the right and left boundary identified after removing the boundary spins on them).

Spins adjacent to the bases of Ω act as symmetry-breaking external fields. The $++$ -cylindrical boundary condition should favor the formation inside Ω of the positively magnetized phase; therefore, it will be natural to consider, in the canonical distribution, this boundary condition only when the total magnetization is fixed to be the spontaneous magnetization $m^*(\beta)$.

On the other hand, the \pm -boundary condition favors the separation of phases (positively magnetized phase near the top of Ω and negatively magnetized phase near the bottom). Therefore, it will be natural to consider the latter boundary condition in the case of a canonical distribution with magnetization $m = (1 - 2\alpha)m^*(\beta)$ with $0 < \alpha < 1$ ([51]). In the latter case, the positive phase can be expected to adhere to the top of Ω and to extend, in some sense to be discussed, up to a distance $O(L)$ from it; and then to change into the negatively magnetized pure phase.

To make the phenomenological description precise, consider the spin configurations σ through the associated sets of disjoint polygons (cf. the section “Symmetry-breaking phase transitions”). Fix the boundary conditions to be $++$ or \pm -cylindrical boundary conditions and note that polygons associated with a spin configuration σ are all closed and of two types: the ones of the first type, denoted $\gamma_1, \dots, \gamma_n$, are polygons which do not encircle Ω ; the second type of polygons, denoted by the symbols λ_α , are the ones which wind up, at least once, around Ω .

So, a spin configuration σ will be described by a set of polygons; the statistical weight of a configuration $\sigma = (\gamma_1, \dots, \gamma_n, \lambda_1, \dots, \lambda_b)$ is (cf. [45]):

$$e^{-2\beta J \left(\sum_i |\gamma_i| + \sum_i |\lambda_i| \right)} \quad [52]$$

The reason why the contours λ that go around the cylinder Ω are denoted by λ (rather than by γ) is that they “look like” open contours (see the section “Symmetry-breaking phase transitions”) if one forgets that the opposite sides of Ω have to be identified. In the case of the \pm -boundary conditions then the number of polygons of λ -type must be odd (hence $\neq 0$), while for the $++$ -boundary condition the number of λ -type polygons must be even (hence it could be 0).

For more details, the reader is referred to Sinai (1991) and Gallavotti (1999).

Separation and Coexistence of Phases

In the context of the geometric description of the spin configuration in the last section, consider the canonical distributions with $++$ -cylindrical or the \pm -cylindrical boundary conditions and zero field: they will be denoted briefly as $\mu_{\beta, ++}$, $\mu_{\beta, \pm}$, respectively. The following theorem (*Minlos–Sinai’s theorem*) provided the foundations of the microscopic theory of coexistence: it is formulated in dimension $d=2$ but, modulo obvious changes, it holds for $d \geq 2$.

Theorem For $0 < \alpha < 1$ fixed, let $m = (1 - 2\alpha)m^*(\beta)$; then for β large enough a spin configuration $\sigma = (\gamma_1, \dots, \gamma_n, \lambda_1, \dots, \lambda_{2b+1})$ randomly chosen with the distribution $\mu_{\beta, \pm}$ enjoys the properties (i)–(iv) below with a $\mu_{\beta, \pm}$ -probability approaching 1 as $\Omega \rightarrow \infty$:

- (i) σ contains only one contour of λ -type and

$$\|\lambda\| - (1 + \varepsilon(\beta))L < o(L) \quad [53]$$

where $\varepsilon(\beta) > 0$ is a suitable (α -independent) function of β tending to zero exponentially fast as $\beta \rightarrow \infty$.

- (ii) If Ω_λ^+ , Ω_λ^- denote respectively, the regions above and below λ , and $|\Omega| \equiv V$, $|\Omega^+|$, $|\Omega^-|$ are, respectively, the volumes of Ω , Ω^+ , Ω^- then

$$\begin{aligned} \left| |\Omega_\lambda^+| - \alpha V \right| &< \kappa(\beta) V^{3/4} \\ \left| |\Omega_\lambda^-| - (1 - \alpha)V \right| &< \kappa(\beta) V^{3/4} \end{aligned} \quad [54]$$

where $\kappa(\beta) \xrightarrow{\beta \rightarrow \infty} 0$ exponentially fast; the exponent $3/4$, here and below, is not optimal.

- (iii) If $M_\lambda^+ = \sum_{x \in \Omega_\lambda^+} \sigma_x$ and $M_\lambda^- = \sum_{x \in \Omega_\lambda^-} \sigma_x$, then

$$\begin{aligned} |M_\lambda^+ - \alpha m^*(\beta) V| &< \kappa(\beta) V^{3/4} \\ |M_\lambda^- - (1 - \alpha) m^*(\beta) V| &< \kappa(\beta) V^{3/4} \end{aligned} \quad [55]$$

- (iv) If $K_\gamma^\lambda(\sigma)$ denotes the number of contours congruent to a given γ and lying in Ω_λ^+ then, simultaneously for all the shapes of γ :

$$|K_\gamma^\lambda(\sigma) - \rho(\gamma)\alpha V| \leq C e^{-\beta J|\gamma|} V^{1/2}, \quad C > 0 \quad [56]$$

where $\rho(\gamma) \leq e^{-2\beta J|\gamma|}$ is the same quantity as already mentioned in the text of the theorem of “Finite-volume effects”. A similar result holds for the contours below λ (cf. the comments on [47]).

The above theorem not only provides a detailed and rather satisfactory description of the phase separation phenomenon, but it also furnishes a precise microscopic definition of the line of separation between the two phases, which should be naturally identified with the (random) line λ .

A similar result holds in the canonical distribution $\mu_{\beta, ++, m^*(\beta)}$ where (i) is replaced by: no λ -type polygon is present, while (ii), (iii) become superfluous, and (iv) is modified in the obvious way. In other words, a typical configuration for the distribution the $\mu_{\beta, ++, m^*(\beta)}$ has the same appearance as a typical configuration of the corresponding grand canonical ensemble with (+)-boundary condition (whose properties are described by the theorem given in the section “Beyond low temperatures (ferromagnetic Ising model”).

For more details, see Sinai (1991) and Gallavotti (1999).

Phase Separation Line and Surface Tension

Continuing to refer to the nearest-neighbor Ising ferromagnet, the theorem of the last section means that, if β is large enough, then the microscopic line λ , separating the two phases, is almost straight (since $\varepsilon(\beta)$ is small). The deviations of λ from a straight line are more conveniently studied in the grand canonical distributions μ_\pm^0 with boundary condition set to $+1$ in the upper half of $\partial\Omega$, vertical sites included, and to -1 in the lower half: this is illustrated in Figure 2 (see the section “Symmetry-breaking phase transitions”). The results can be converted into very similar results for grand canonical distributions with \pm -cylindrical boundary conditions of the last section.

Define λ to be *rigid* if the probability that λ passes through the center of the box Ω (i.e., 0) does not tend to 0 as $\Omega \rightarrow \infty$; otherwise, it is *not rigid*.

The notion of rigidity distinguishes between the possibilities for the line λ to be “straight.” The “excess” length $\varepsilon(\beta)L$ (see [53]) can be obtained in two ways: either the line λ is essentially straight (in the geometric sense) with a few “bumps” distributed with a density of order $\varepsilon(\beta)$ or, otherwise, it is only locally straight and with an important part of the excess length being gained through a small bending on a large length scale. In three dimensions a similar phenomenon is possible. Rigidity of λ , or its failure, can in principle be investigated by optical means;

there can be interference of coherent light scattered by macroscopically separated surface elements of λ only if λ is rigid in the above sense.

It has been rigorously proved that, the line λ is *not rigid* in dimension 2. And, at least at low temperature, the fluctuation of the middle point is of the order $O(\sqrt{L})$. In dimension 3 however, it has been shown that the surface λ is *rigid* at low enough temperature.

A deeper analysis is needed to study the shape of the separation surface under other conditions, for example, with $+$ boundary conditions in a canonical distribution with magnetization intermediate between $\pm m^*(\beta)$. It involves, as a prerequisite, the definition and many properties of the surface tension between the two phases. Here only the definition of surface tension in the case of \pm -boundary conditions in the two-dimensional case will be mentioned. If $Z^{++}(\Omega, m^*(\beta))$ and $Z^{+-}(\Omega, m)$ are, respectively, the canonical partition functions for the $++$ - and \pm -cylindrical boundary conditions the tension $\tau(\beta)$ is defined as

$$\beta\tau(\beta) = - \lim_{\Omega \rightarrow \infty} \frac{1}{L} \log \frac{Z^{+-}(\Omega, m)}{Z^{++}(\Omega, m^*(\beta))}$$

The limit can be shown to be α -independent for β large enough: the definition and its justification is based on the microscopic geometric description in the section “Geometry of phase co-existence.” The definition can be naturally extended to higher dimension (and to more general non-nearest-neighbor models). If $d=2$, the tension τ can be exactly computed at all temperatures below criticality and is $\beta\tau(\beta) = 2\beta J + \log \tanh \beta J$.

More remarkably, the definition can be extended to define the surface tension $\tau(\beta, \mathbf{n})$ in the “direction \mathbf{n} ,” that is, when the boundary conditions are such that the line of separation is in the average orthogonal to the unit vector \mathbf{n} . In this way, if $d=2$ and $\alpha \in (0, 1)$ is fixed, it can be proved that at low enough temperature the canonical distribution with $+$ boundary conditions and intermediate magnetization $m = (1 - 2\alpha)m^*(\beta)$ has typical configurations containing a spin $-$ region of area $\sim \alpha V$; furthermore, if the container is rescaled to size $L=1$, the region will have a limiting shape filling an area α bounded by a smooth curve whose form is determined by the classical macroscopic *Wulff’s theory* of the shape of crystals in terms of the surface tension $\tau(\mathbf{n})$.

An interesting question remains open in the three-dimensional case: it is conceivable that the surface, although rigid at low temperature, might become “loose” at a temperature \tilde{T}_c smaller than the critical

temperature T_c (the latter being defined as the highest temperature below which there are at least two pure phases). The temperature \tilde{T}_c , whose existence is rather well established in numerical experiments, would be called the “roughening transition” temperature. The rigidity of λ is connected with the existence of translationally non-invariant equilibrium states. The latter exist in dimension $d=3$, but not in dimension $d=2$, where the discussed nonrigidity of λ , established all the way to T_c , provides the intuitive reason for the absence of non-translation-invariant states. It has been shown that in $d=3$ the roughening temperature $\tilde{T}_c(\beta)$ necessarily cannot be smaller than the critical temperature of the two-dimensional Ising model with the same coupling.

Note that existence of translationally noninvariant equilibrium states is not necessary for the description of coexistence phenomena. The theory of the nearest-neighbor two-dimensional Ising model is a clear proof of this statement.

The reader is referred to [Onsager \(1944\)](#), [van Beyeren \(1975\)](#), [Sinai \(1991\)](#), [Miracle-Solé \(1995\)](#), [Pfister and Velenik \(1999\)](#), and [Gallavotti \(1999\)](#) for more details.

Critical Points

Correlation functions for a system with short-range interactions and in an equilibrium state (which is a pure phase) have cluster properties (see [22]): their physical meaning is that in a pure phase there is independence between fluctuations occurring in widely separated regions. The simplest cluster property concerns the “pair correlation function,” that is, the probability density $\rho(\mathbf{q}_1, \mathbf{q}_2)$ of finding particles at points $\mathbf{q}_1, \mathbf{q}_2$ independently of where the other particles may happen to be (see [23]). In the case of spin systems, the pair correlation $\rho(\mathbf{q}_1, \mathbf{q}_2) = \langle \sigma_{\mathbf{q}_1} \sigma_{\mathbf{q}_2} \rangle$ will be considered. The pair correlation of a translation-invariant equilibrium state has a cluster property ([22], [42]), if

$$|\rho(\mathbf{q}_1, \mathbf{q}_2) - \rho^2|_{|\mathbf{q}_1 - \mathbf{q}_2| \rightarrow \infty} \rightarrow 0 \quad [57]$$

where ρ is the probability density for finding a particle at \mathbf{q} (i.e., the physical density of the state) or $\rho = \langle \sigma_{\mathbf{q}} \rangle$ is the average of the value of the spin at \mathbf{q} (i.e., the magnetization of the state).

A general definition of critical point is a point c in the space of the parameters characterizing equilibrium states, for example, β, λ in grand canonical distributions, β, v in canonical distributions, or β, h in the case of lattice spin systems in a grand canonical

distribution. In systems with short-range interaction (i.e., with $\varphi(\mathbf{r})$ vanishing for $|\mathbf{r}|$ large enough) the point c is a critical point if the pair correlation tends to 0 (see [57]), slower than exponential (e.g., as a power of the distance $|\mathbf{r}| = |\mathbf{q}_1 - \mathbf{q}_2|$).

A typical example is the two-dimensional Ising model on a square lattice and with nearest-neighbor ferromagnetic interaction of size J . It has a single critical point at $\beta = \beta_c$, $h = 0$ with $\sinh 2\beta_c J = 1$. The cluster property is that $\langle \sigma_x \sigma_y \rangle - \langle \sigma_x \rangle \langle \sigma_y \rangle \xrightarrow{|x-y| \rightarrow \infty} 0$ as

$$\begin{aligned} A_+(\beta) \frac{e^{-\kappa(\beta)|x-y|}}{\sqrt{|x-y|}}, \quad A_-(\beta) \frac{e^{-\kappa(\beta)|x-y|}}{|x-y|^2} \\ A_c \frac{1}{|x-y|^{1/4}}, \end{aligned} \quad [58]$$

for $\beta < \beta_c$, $\beta > \beta_c$, or $\beta = \beta_c$, respectively, where $A_{\pm}(\beta)$, A_c , $\kappa(\beta) > 0$. The properties [58] stem from the exact solution of the model.

At the critical point, several interesting phenomena occur: the lack of exponential decay indicates lack of a length scale over which really distinct phenomena can take place, and properties of the system observed at different length scales are likely to be simply related by suitable scaling transformations. Many efforts have been dedicated at finding ways of understanding quantitatively the scaling properties pertaining to different observables. The result has been the development of the *renormalization group* approach to critical phenomena (cf. the section “Renormalization group”). The picture that emerges is that the closer the critical point is the larger becomes the maximal scale of length below which scaling properties are observed. For instance, in a lattice spin system in zero field the magnetization $M|\Lambda|^{-a}$ in a box $\Lambda \subset \Omega$ should have essentially the same distribution for all Λ 's with side $< l_0(\beta)$ and $l_0(\beta) \rightarrow \infty$ as $\beta \rightarrow \beta_c$, provided a is suitably chosen. The number a is called a *critical exponent*.

There are several other “critical exponents” that can be defined near a critical point. They can be associated with singularities of the thermodynamic function or with the behavior of the correlation functions involving joint densities at two or more than two points. As an example, consider a lattice spin system: then the “ $2n$ -spins correlation” $\langle \sigma_0 \sigma_{\xi_1} \dots \sigma_{\xi_{2n-1}} \rangle_c$ could behave proportionally to $\chi_{2n}(0, \xi_1, \dots, \xi_{2n-1})$, $n = 1, 2, 3, \dots$, for a suitable family of homogeneous functions χ_n , of some degree ω_{2n} , of the coordinates $(\xi_1, \dots, \xi_{2n-1})$ at least when the reciprocal distances are large but $< l_0(\beta)$ and

$$l_0(\beta) = \text{const.} (\beta - \beta_c)^{-\nu} \xrightarrow{\beta \rightarrow \beta_c} \infty$$

This means that if ξ_i are regarded as points in \mathbb{R}^d there are functions χ_{2n} such that

$$\begin{aligned} \chi_{2n} \left(0, \frac{\xi_1}{\lambda}, \dots, \frac{\xi_{2n-1}}{\lambda} \right) = \lambda^{\omega_{2n}} \chi_{2n}(0, \xi_1, \dots, \xi_{2n-1}) \\ 0 < \lambda \in \mathbb{R} \end{aligned} \quad [59]$$

and $\langle \sigma_0 \sigma_{\xi_1} \dots \sigma_{\xi_{2n-1}} \rangle \propto \chi_{2n}(0, \xi_1, \dots, \xi_{2n-1})$ if $1 \ll |x_i - x_j| \ll l_0(\beta)$. The numbers ω_{2n} define a sequence of critical exponents.

Other critical exponents can be associated with approaching the critical point along other directions (e.g., along $h \rightarrow 0$ at $\beta = \beta_c$). In this case, the length up to which there are scaling phenomena is $l_0(h) = \ell_0 h^{-\bar{\nu}}$. Further, the magnetization $m(h)$ tends to 0 as $h \rightarrow 0$ at fixed $\beta = \beta_c$ as $m(h) = m_0 h^{1/\delta}$ for $\delta > 0$.

None of the features of critical exponents is known rigorously, including their existence. An exception is the case of the two-dimensional nearest-neighbor Ising ferromagnet where some exponents are known exactly (e.g., $\omega_2 = 1/4$, $\omega_{2n} = n\omega_2$, or $\nu = 1$, while δ , $\bar{\nu}$ are not rigorously known). Nevertheless, for Ising ferromagnets (not even nearest-neighbor but, as always here, finite-range) in all dimensions, all of the exponents mentioned are conjectured to be the same as those of the nearest-neighbor Ising ferromagnet. A further exception is the derivation of rigorous relations between critical exponents and, in some cases, even their values under the assumption that they exist.

Remark Naively it could be expected that in a pure state in zero field with $\langle \sigma_x \rangle = 0$ the quantity $s = |\Lambda|^{-1/2} \sum_{x \in \Lambda} \sigma_x$, if Λ is a cubic box of side ℓ , should have a probability distribution which is Gaussian, with dispersion $\lim_{\Lambda \rightarrow \infty} \langle s^2 \rangle$. This is “usually true,” but not always. Properties [58] show that in the $d=2$ ferromagnetic nearest-neighbor Ising model, $\langle s^2 \rangle$ diverges proportionally to $\ell^{2-1/4}$ so that the variable s cannot have the above Gaussian distribution. The variable $S = |\Lambda|^{-7/8} \sum_{x \in \Lambda} \sigma_x$ will have a finite dispersion: however, there is no reason that it should be Gaussian. This makes clear the great interest of a fluctuation theory and its relevance for the critical point studies (see the next two sections).

For more details, the reader is referred to [Onsager \(1944\)](#), [Domb and Green \(1972\)](#), [McCoy and Wu \(1973\)](#), and [Aizenman \(1982\)](#).

Fluctuations

As it appears from the discussion in the last section, fluctuations of observables around their averages have interesting properties particularly at critical points. Of particular interest are observables that

are averages, over large volumes Λ , of local functions $F(x)$ on phase space: this is so because macroscopic observables often have this form. For instance, given a region Λ inside the system container Ω , $\Lambda \subset \Omega$, consider a configuration $x = (\mathbf{P}, \mathbf{Q})$ and the number of particles $N_\Lambda = \sum_{q \in \Lambda} 1$ in Λ , or the potential energy $\Phi_\Lambda = \sum_{(q, q') \in \Lambda} \varphi(q - q')$ or the kinetic energy $K_\Lambda = \sum_{q \in \Lambda} (1/2m)p^2$. In the case of lattice spin systems, consider a configuration σ and, for instance, the magnetization $M_\Lambda = \sum_{i \in \Lambda} \sigma_i$ in Λ . Label the above four examples by $\alpha = 1, \dots, 4$.

Let μ_α be the probability distribution describing the equilibrium state in which the quantities X_Λ are considered; let $x_\Lambda = \langle X_\Lambda / |\Lambda| \rangle_{\mu_\alpha}$ and $p \stackrel{\text{def}}{=} (X_\Lambda - x_\Lambda) / |\Lambda|$. Then typical properties of fluctuations that should be investigated are ($\alpha = 1, \dots, 4$):

1. for all $\delta > 0$ it is $\lim_{\Lambda \rightarrow \infty} \mu_\alpha(|p| > \delta) = 0$ (*law of large numbers*);
2. there is $D_\alpha > 0$ such that

$$\mu(p\sqrt{|\Lambda|} \in [a, b]) \xrightarrow{\Lambda \rightarrow \infty} \int_a^b \frac{dz}{\sqrt{2\pi D_\alpha}} e^{-z^2/2D_\alpha}$$

(*central limit law*); and

3. there is an interval $I_\alpha = (p_{\alpha,-}^*, p_{\alpha,+}^*)$ and a concave function $F_\alpha(p)$, $p \in I$, such that if $[a, b] \subset I$ then

$$\frac{1}{|\Lambda|} \log \mu(p \in [a, b]) \xrightarrow{\Lambda \rightarrow \infty} \max_{p \in [a, b]} F_\alpha(p)$$

(*large deviations law*).

The law of large numbers provides the certainty of the macroscopic values; the central limit law controls the small fluctuations (of order $\sqrt{|\Lambda|}$) of X_Λ around its average; and the large deviations law concerns the fluctuations of order $|\Lambda|$.

The relations (1)–(3) above are not always true: they can be proved under further general assumptions if the potential φ satisfies [14] in the case of particle systems or if $\sum_q |\varphi(q)| < \infty$ in the case of lattice spin systems. The function $F_\alpha(p)$ is defined in terms of the thermodynamic limits of suitable thermodynamic functions associated with the equilibrium state μ_α . The further assumption is, essentially in all cases, that a suitable thermodynamic function in terms of which $F_\alpha(p)$ will be expressed is smooth and has a nonvanishing second derivative.

For the purpose of a simple concrete example, consider a lattice spin system of Ising type with energy $-\sum_{x, y \in \Omega} \varphi(x - y)\sigma_x\sigma_y - \sum_x h\sigma_x$ and the fluctuations of the magnetization $M_\Lambda = \sum_{x \in \Lambda} \sigma_x$, $\Lambda \subset \Omega$, in the grand canonical equilibrium states $\mu_{h, \beta}$.

Let the free energy be $\beta f(\beta, h)$ (see [41]), let $m = m(h) \stackrel{\text{def}}{=} \langle M_\Lambda / |\Lambda| \rangle$ and let $h(m)$ be the inverse

function of $m(h)$. If $p = M_\Lambda / |\Lambda|$ the function $F(p)$ is given by

$$F(p) = \beta(f(\beta, h(p)) - f(\beta, h) - \partial_b f(\beta, h)(h(p) - h)) \quad [60]$$

then a quite general result is:

Theorem *The relations (1)–(3) hold if the potential satisfies $\sum_x |\varphi(x)| < \infty$ and if $F(p)$ [60] is smooth and $F''(p) \neq 0$ in open intervals around those in which p is considered, that is, around $p = 0$ for the law of large numbers and for the central limit law or in an open interval containing a, b for the case of the large deviations law.*

In the cases envisaged, the theory of equivalence of ensembles implies that the function F can also be computed via thermodynamic functions naturally associated with other equilibrium ensembles. For instance, instead of the grand canonical $f(\beta, h)$, one could consider the canonical $\beta g(\beta, m)$ (see [41]), then

$$F(p) = -\beta(g(\beta, p) - g(\beta, m) - \partial_m g(\beta, m)(p - m)) \quad [61]$$

It has to be remarked that there should be a strong relation between the central limit law and the law of large deviations. Setting aside stating the conditions for a precise mathematical theorem, the statement can be efficiently illustrated in the case of a ferromagnetic lattice spin system and with $\Lambda \equiv \Omega$, by showing that the law of large deviations in small intervals, around the average $m(h_0)$, at a value h_0 of the external field, is implied by the validity of the central limit law for all values of h near h_0 and vice versa (here β is fixed). Taking $h_0 = 0$ (for simplicity), the heuristic reasons are the following. Let $\mu_{h, \Omega}$ be the grand canonical distribution in external field h . Then:

1. The probability $\mu_{h, \Omega}(p \in dp)$ is proportional, by definition, to $\mu_{0, \Omega}(p \in dp)e^{-\beta hp|\Omega|}$. Hence, if the central limit law holds for all h near $h_0 = 0$, there will exist two functions $m(h)$ and $D(h) > 0$, defined for h near $h_0 = 0$, with $m(0) = 0$ and

$$\begin{aligned} \mu_0(p \in dp) e^{-\beta hp|\Omega|} \\ = \text{const. exp} \left(-|\Omega| \left(\frac{(p - m(h))^2}{2D(h)} + o(\Omega) \right) \right) dp \quad [62] \end{aligned}$$

2. There is a function $\zeta(m)$ such that $\partial_m \zeta(m(h)) = \beta h$ and $\partial_m^2 \zeta(m(h)) = D(h)^{-1}$. (This is obtained by noting that, given $D(h)$, the differential equation $\partial_m \beta h = D(h)^{-1}$ with the initial value $h(0) = 0$ determines the function $h(m)$; therefore, $\zeta(m)$ is determined by a second integration, from $\partial_m \zeta(m) = \beta h(m)$.)

It then follows, heuristically, that the probability of p in zero field has the form $\text{const. } e^{c(p)|\Omega|} dp$ so that the probability that $p \in [a, b]$ will be $\text{const} \exp(|\Omega| \max_{p \in [a, b]} \zeta(p))$.

Conversely, the large deviations law for p at $h=0$ implies the validity of the central limit law for the fluctuations of p in all small enough fields h : this simply arises from the function $F(p)$ having a negative second derivative.

This means that there is a “duality” between central limit law and large deviation law or that the law of large deviations is a “global version” of the central limit law, in the sense that:

1. if the central limit law holds for h in an interval around h_0 then the fluctuations of the magnetization at field h_0 satisfy a large deviation law in a small enough interval J around $m(h_0)$; and
2. if a large deviation law is satisfied in an interval around h_0 then the central limit law holds for the fluctuations of magnetization around its average in all fields h with $h - h_0$ small enough.

Going beyond the heuristic level in establishing the duality amounts to giving a precise meaning to “small enough” and to discuss which properties of $m(h)$ and $D(h)$, or $F(p)$ are needed to derive properties (1), (2).

For purposes of illustration consider the Ising model with ferromagnetic short range interaction φ : then the central limit law holds for all h if β is small enough and, under the same condition on β , the large deviations law holds for all h and all intervals $[a, b] \subset (-1, 1)$. If β is not small then the condition $h \neq 0$ has to be added. Hence, the conditions are fairly weak and the apparent exceptions concern the value $h=0$ and β not small where the statements may become invalid because of possible phase transitions.

In presence of phase transitions, the law of large numbers, the central limit law, and law of large deviations should be reformulated. Basically, one has to add the requirement that fluctuations are considered in pure phases and change, in a natural way, the formulation of the laws. For instance, the large fluctuations of magnetization in a pure phase of the Ising model in zero field and large β (i.e., in a state obtained as limit of finite-volume states with $+$ or $-$ boundary conditions) in intervals $[a, b]$ which do not contain the average magnetization m^* are not necessarily exponentially small with the size of $|\Lambda|$: if $[a, b] \subset [-m^*, m^*]$ they are exponentially small but only with the size of the surface of Λ (i.e., with $|\Lambda|^{(d-1)/d}$) while they are exponentially small with the volume if $[a, b] \cap [-m^*, m^*] = \emptyset$.

The discussion of the last section shows that at the critical point the nature of the large fluctuations is also expected to change: no central limit law is expected to hold in general because of the example of [58] with the divergence of the average of the normal second moment of the magnetization in a box as the side tends to ∞ .

For more details the reader is referred to Olla (1987).

Renormalization Group

The theory of fluctuations just discussed concerns only fluctuations of a single quantity. The problem of joint fluctuations of several quantities is also interesting and in fact led to really new developments in the 1970s. It is necessary to restrict attention to rather special cases in order to illustrate some ideas and the philosophy behind the approach. Consider, therefore, the equilibrium distribution μ_0 associated with one of the classical equilibrium ensembles. To fix the ideas we consider the equilibrium distribution of an Ising energy function βH_0 , having included the temperature factor in the energy: the inclusion is done because the discussion will deal with the properties of μ_0 as a function of β . It will also be assumed that the average of each spin is zero (“no magnetic field,” see [39] with $h=0$). Keeping in mind a concrete case, imagine that βH_0 is the energy function of the nearest-neighbor Ising ferromagnet in zero field.

Imagine that the volume Ω of the container has periodic boundary conditions and is very large, ideally infinite. Define the family of blocks $k\xi$, parametrized by $\xi \in \mathbb{Z}^d$ and with k an integer, consisting of the lattice sites $x = \{k\xi_i \leq x_i < (k+1)\xi_i\}$. This is a lattice of cubic blocks with side size k that will be called the “ k -rescaled lattice.”

Given α , the quantities $m_\xi = k^{-\alpha d} \sum_{x \in k\xi} \sigma_x$ are called the block spins and define the map $R_{\alpha, k}^* \mu_0 = \mu_k$ transforming the initial distribution on the original spins into the distribution of the block spins. Note that if the initial spins have only two values $\sigma_x = \pm 1$, the block spins take values between $-k^d/k^{\alpha d}$ and $k^d/k^{\alpha d}$ at steps of size $2/k^{\alpha d}$. Furthermore, the map $R_{\alpha, k}^*$ makes sense independently of how many values the initial spins can assume, and even if they assume a continuum of values $S_x \in \mathbb{R}$.

Taking $\alpha=1$ means, for k large, looking at the probability distribution of the joint large fluctuations in the blocks $k\xi$. Taking $\alpha=1/2$ corresponds to studying a joint central limit property for the block variables.

Considering a one-parameter family of initial distributions μ_0 parametrized by a parameter β

(that will be identified with the inverse temperature), typically there will be a unique value $\alpha(\beta)$ of α such that the joint fluctuations of the block variables admit a limiting distribution,

$$\begin{aligned} & \text{prob}_k(m_\xi \in [a_\xi, b_\xi], \sigma \in \Lambda) \\ & \xrightarrow{k \rightarrow \infty} \int_{\{a_\xi\}}^{\{b_\xi\}} g_\Lambda((S_\xi)_{\xi \in \Lambda}) \prod_{\xi \in \Lambda} dS_\xi \end{aligned} \quad [63]$$

for some distribution $g_\Lambda(\mathbf{z})$ on \mathbb{R}^Λ .

If $\alpha > \alpha(\beta)$, the limit will then be $\prod_{\xi \in \Lambda} \delta(S_\xi) dS_\xi$, or if $\alpha < \alpha(\beta)$ the limit will not exist (because the block variables will be too large, with a dispersion diverging as $k \rightarrow \infty$).

It is convenient to choose as sequence of $k \rightarrow \infty$ the sequence $k = 2^n$ with $n = 0, 1, \dots$ because in this way it is $R_{\alpha, k}^* \equiv R_{\alpha, 1}^{*n}$ and the limits $k \rightarrow \infty$ along the sequence $k = 2^n$ can be regarded as limits on a sequence of iterations of a map $R_{\alpha, 1}^*$ acting on the probability distributions of generic spins S_x on the lattice \mathbb{Z}^d (the sequence 3^n would be equally suited).

It is even more convenient to consider probability distributions that are expressed in terms of energy functions H which generate, in the thermodynamic limit, a distribution μ : then $R_{\alpha, 1}^*$ defines an action R_α on the energy functions so that $R_\alpha H = H'$ if H generates μ , H' generates μ' and $R_{\alpha, 1}^* \mu = \mu'$. Of course, the energy function will be more general than [39] and at least a form like δU in [49] has to be admitted.

In other words, R_α gives the result of the action of $R_{\alpha, 1}^*$ expressed as a map acting on the energy functions. Its iterates also define a semigroup which is called the block spin renormalization group.

While the map $R_{\alpha, 1}^*$ is certainly well defined as a map of probability distributions into probability distributions, it is by no means clear that R_α is well defined as a map on the energy functions. Because, if μ is given by an energy function, it is not clear that $R_{\alpha, 1}^* \mu$ is such.

A remarkable theorem can be (easily) proved when $R_{\alpha, 1}^*$ and its iterates act on initial μ_0 's which are equilibrium states of a spin system with short-range interactions and at high temperature (β small). In this case, if $\alpha = 1/2$, the sequence of distributions $R_{1/2, 1}^{*n} \mu_0(\beta)$ admits a limit which is given by a product of independent Gaussians:

$$\begin{aligned} & \text{prob}_k(m_\xi \in [a_\xi, b_\xi], \sigma \in \Lambda) \\ & \xrightarrow{k \rightarrow \infty} \int_{\{a_\xi\}}^{\{b_\xi\}} \prod_{\xi \in \Lambda} \exp\left(-\frac{1}{2D(\beta)} S_\xi^2\right) \prod_{\xi \in \Lambda} \frac{dS_\xi}{\sqrt{2\pi D(\beta)}} \end{aligned} \quad [64]$$

Note that this theorem is stated without even mentioning the renormalization maps $R_{1/2}^n$: it can nevertheless be interpreted as stating that

$$R_{1/2}^n \beta H_0 \xrightarrow{n \rightarrow \infty} \sum_{\xi \in \mathbb{Z}^d} \frac{1}{2D(\beta)} S_\xi^2 \quad [65]$$

but the interpretation is not rigorous because [64] does not state require that $R_{1/2}^n H_0(\beta)$ makes sense for $n \geq 1$. It states that at high temperature block spins have normal independent fluctuations: it is therefore an extension of the central limit law.

There are a few cases in which the map R_α can be rigorously shown to be well defined at least when acting on special equilibrium states like the high-temperature lattice spin systems: but these are exceptional cases of relatively little interest.

Nevertheless, there is a vast literature dealing with approximate representations of the map R_α . The reason is that, assuming not only its existence but also that it has the properties that one would normally expect to hold for a map acting on a finite dimensional space, it follows that a number of consequences can be drawn; quite nontrivial ones as they led to the first theory of the critical point that goes beyond the van der Waals theory described in the section ‘‘van der Waals theory.’’

The argument proceeds essentially as follows. At the critical point, the fluctuations are expected to be anomalous (cf. the last remark in the section ‘‘Critical points’’) in the sense that $\langle (\sum_{x \in \Lambda} \sigma_x / \sqrt{|\Lambda|})^2 \rangle$ will tend to ∞ , because $\alpha = 1/2$ does not correspond to the right fluctuation scale of $\sum_{\xi \in \Lambda} \sigma_\xi$, signaling that $R_{1/2, 1}^{*n} \mu_0(\beta_c)$ will not have a limit but, possibly, there is $\alpha_c > 1/2$ such that $R_{\alpha_c, 1}^{*n} \mu_0(\beta_c)$ converges to a limit in the sense of [63]. In the case of the critical nearest-neighbor Ising ferromagnetic $\alpha_c = 7/8$ (see ending remark in the section ‘‘Critical points’’). Therefore, if the map $R_{\alpha_c, 1}^*$ is considered as acting on $\mu_0(\beta)$, it will happen that for all $\beta < \beta_c$, $R_{\alpha_c, 1}^{*n} \mu_0(\beta_c)$ will converge to a trivial limit $\prod_{\xi \in \Lambda} \delta(S_\xi) dS_\xi$ because the value α_c is greater than $1/2$ while normal fluctuations are expected.

If the map R_{α_c} can be considered as a map on the energy functions, this says that $\prod_{\xi \in \Lambda} \delta(S_\xi) dS_\xi$ is a ‘‘(trivial) fixed point of the renormalization group’’ which ‘‘attracts’’ the energy functions βH_0 corresponding to the high-temperature phases.

The existence of the critical β_c can be associated with the existence of a *nontrivial fixed point* H^* for R_{α_c} which is hyperbolic with just one Lyapunov exponent $\lambda > 1$; hence, it has a stable manifold of codimension 1. Call μ^* the probability distribution corresponding to H^* .

The migration towards the trivial fixed point for $\beta < \beta_c$ can be explained simply by the fact that for

such values of β the initial energy function βH_0 is outside the stable manifold of the nontrivial fixed point and under application of the renormalization transformation $R_{\alpha_c}^n$, βH_0 migrates toward the trivial fixed point, which is attractive in all directions.

By increasing β , it may happen that, for $\beta = \beta_c$, βH_0 crosses the stable manifold of the nontrivial fixed point H^* for R_{α_c} . Then $R_{\alpha_c}^n \beta_c H_0$ will no longer tend to the trivial fixed point but it will tend to H^* : this means that the block spin variables will exhibit a completely different fluctuation behavior. If β is close to β_c , the iterations of R_{α_c} will bring $R_{\alpha_c}^n \beta H_0$ close to H^* , only to be eventually repelled along the unstable direction reaching a distance from it increasing as $\lambda^n |\beta - \beta_c|$.

This means that up to a scale length $O(2^{n(\beta)})$ lattice units with $\lambda^{n(\beta)} |\beta - \beta_c| = 1$ (i.e., up to a scale $O(|\beta - \beta_c|^{-1/\log_2 \lambda})$), the fluctuations will be close to those of the fixed point distribution μ^* , but beyond that scale they will come close to those of the trivial fixed point: to see them the block spins would have to be normalized with index $\alpha = 1/2$ and they would appear as uncorrelated Gaussian fluctuations (cf. [64], [65]).

The next question concerns finding the nontrivial fixed points, which means finding the energy functions H^* and the corresponding α_c which are fixed points of R_{α_c} . If the above picture is correct, the distributions μ^* corresponding to the H^* would describe the critical fluctuations and, if there was only one choice, or a limited number of choices, of α_c and H^* this would open the way to a universality theory of the critical point hinted already by the “primitive” results of van der Waals’ theory.

The initial hope was, perhaps, that there would be a very small number of critical values α_c and H^* possible: but it rapidly faded away leaving, however, the possibility that the critical fluctuations could be classified into universality classes. Each class would contain many energy functions which, upon iterated actions of R_{α_c} , would evolve under the control of the trivial fixed point (always existing) for β small while, for $\beta = \beta_c$, they would be controlled, instead, by a nontrivial fixed point H^* for R_{α_c} with the same α_c and the same H^* . For $\beta < \beta_c$, a “resolution” of the approach to the trivial fixed point would be seen by considering the map $R_{1/2}$ rather than R_{α_c} whose iterates would, however, lead to a Gaussian distribution like [64] (and to a limit energy function like [65]).

The picture is highly hypothetical: but it is the first suggestion of a mechanism leading to critical points with the character of universality and with exponents different from those of the van der Waals theory or, for ferromagnets on a lattice, from those of its lattice version (the *Curie–Weiss theory*). Furthermore, accepting the approximations

(e.g., the Wilson–Fisher ε -expansion) that allow one to pass from the well-defined $R_{\alpha_c,1}^*$ to the action of R_{α_c} on the energy functions, it is possible to obtain quite unambiguously values for α_c and expressions for H^* which are associated with the action of R_{α_c} on various classes of models.

For instance, it can lead to conclude that the critical behavior of all ferromagnetic finite-range lattice spin systems (with energy functions given by [39]) have critical points controlled by the same α_c and the same nontrivial fixed point: this property is far from being mathematically proved, but it is considered a major success of the theory. One has to compare it with van der Waals’ critical point theory: *for the first time*, an approximation scheme has led, even though under approximations not fully controllable, to computable critical exponents which are not equal to those of the van der Waals theory.

The renormalization group approach to critical phenomena has many variants, depending on which kind of fluctuations are considered and on the models to which it is applied. In statistical mechanics, there are a few mathematically complete applications: certain results in higher dimensions, theory of dipole gas in $d=2$, hierarchical models, some problems in condensed matter and in statistical mechanics of lattice spins, and a few others. Its main mathematical successes have occurred in various related fields where not only the philosophy described above can be applied but it leads to renormalization transformations that can be defined precisely and studied in detail: for example, constructive field theory, KAM theory of quasiperiodic motions, and various problems in dynamical systems.

However, the applications always concern special cases and in each of them the general picture of the trivial–nontrivial fixed point dichotomy appears realized but without being accompanied, except in rare cases (like the hierarchical models or the universality theory of maps of the interval), by the full description of stable manifold, unstable direction, and action of the renormalization transformation on objects other than the one of immediate interest (a generality which looks often an intractable problem, but which also turns out not to be necessary).

In the renormalization group context, mathematical physics has played an important role also by providing clear evidence that universality classes could not be too few: this was shown by the numerous exact solutions after Onsager’s solution of the nearest-neighbor Ising ferromagnet: there are in fact several lattice models in $d=2$ that exhibit critical points with some critical exponents exactly computable and that depend continuously on the models parameters.

For more details, we refer the reader to McCoy and Wu (1973), Baxter (1982), Bleher and Sinai (1975), Wilson and Fisher (1972), Gawedzky and Kupiainen (1983, 1985), Benfatto and Gallavotti (1995), and Mastropietro (2004).

Quantum Statistics

Statistical mechanics is extended to assemblies of quantum particles rather straightforwardly. In the case of N identical particles, the observables are operators O on the Hilbert space

$$\mathcal{H}_N = L_2(\Omega)_\alpha^N \quad \text{or} \quad \mathcal{H}_N = (L_2(\Omega) \otimes \mathcal{C}^2)_\alpha^N$$

where $\alpha = +, -$, of the symmetric ($\alpha = +$, bosonic particles) or antisymmetric ($\alpha = -$, fermionic particles) functions $\psi(\mathbf{Q})$, $\mathbf{Q} = (q_1, \dots, q_N)$, of the position coordinates of the particles or of the position and spin coordinates $\psi(\mathbf{Q}, \boldsymbol{\sigma})$, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$, normalized so that

$$\int |\psi(\mathbf{Q})|^2 d\mathbf{Q} = 1 \quad \text{or} \quad \sum_{\boldsymbol{\sigma}} \int |\psi(\mathbf{Q}, \boldsymbol{\sigma})|^2 d\mathbf{Q} = 1$$

here only $\sigma_j = \pm 1$ is considered. As in classical mechanics, a state is defined by the average values $\langle O \rangle$ that it attributes to the observables.

Microcanonical, canonical, and grand canonical ensembles can be defined quite easily. For instance, consider a system described by the Hamiltonian ($\hbar = \text{Planck's constant}$)

$$H_N = -\frac{\hbar^2}{2m} \sum_{j=1}^N \Delta_{q_j} + \sum_{j < j'} \varphi(q_j - q_{j'}) + \sum_j w(q_j) \stackrel{\text{def}}{=} K + \Phi \quad [66]$$

where periodic boundary conditions are imagined on Ω and $w(q)$ is periodic, smooth potential (the side of Ω is supposed to be a multiple of the periodic potential period if $w \neq 0$). Then a canonical equilibrium state with inverse temperature β and specific volume $v = V/N$ attributes to the observable O the average value

$$\langle O \rangle \stackrel{\text{def}}{=} \frac{\text{tr} e^{-\beta H_N} O}{\text{tr} e^{-\beta H_N}} \quad [67]$$

Similar definitions can be given for the grand canonical equilibrium states.

Remarkably, the ensembles are orthodic and a “heat theorem” (see the section “Heat theorem and ergodic hypothesis”) can be proved. However, “equipartition” does not hold: that is, $\langle K \rangle \neq (d/2)N\beta^{-1}$, although β^{-1} is still the integrating factor of $dU + p dV$ in the heat theorem; hence, β^{-1} continues to be proportional to temperature.

Lack of equipartition is important, as it solves paradoxes that arise in classical statistical mechanics applied to systems with infinitely many degrees of freedom, like crystals (modeled by lattices of coupled oscillators) or fields (e.g., the electromagnetic field important in the study of black body radiation). However, although this has been the first surprise of quantum statistics (and in fact responsible for the very discovery of quanta), it is by no means the last.

At low temperatures, new unexpected (i.e., with no analogs in classical statistical mechanics) phenomena occur: Bose–Einstein condensation (superfluidity), Fermi surface instability (superconductivity), and appearance of off-diagonal long-range order (ODLRO) will be selected to illustrate the deeply different kinds of problems of quantum statistical mechanics. Largely not yet understood, such phenomena pose very interesting problems not only from the physical point of view but also from the mathematical point of view and may pose challenges even at the level of a definition. However, it should be kept in mind that in the interesting cases (i.e., three-dimensional systems and even most two- and one-dimensional systems) there is no proof that the objects defined below really exist for the systems like [66] (see, however, the final comment for an important exception).

Bose–Einstein Condensation

In a canonical state with parameters β, v , a definition of the occurrence of Bose condensation is in terms of the eigenvalues $\nu_j(\Omega, N)$ of the kernel $\rho(q, q')$ on $L_2(\Omega)$, called the *one-particle reduced density matrix*, defined by

$$N \sum_{n=1}^{\infty} \frac{e^{-\beta E_n(\Omega, N)}}{\text{tr} e^{-\beta H_N}} \int \bar{\psi}_n(q, q_1, \dots, q_{N-1}) \times \psi_n(q', q_1, \dots, q_{N-1}) dq_1 \dots dq_{N-1} \quad [68]$$

where $E_n(\Omega, N)$ are the eigenvalues of H_N and $\psi_n(q_1, \dots, q_N)$ are the corresponding eigenfunctions. If ν_j are ordered by increasing value, the state with parameters β, v is said to contain a *Bose–Einstein condensate* if $\nu_1(\Omega, N) \geq bN > 0$ for all large Ω at $v = V/N, \beta$ fixed. This receives the interpretation that there are more than bN particles with equal momentum. The free Bose gas exhibits a Bose condensation phenomenon at fixed density and small temperature.

Fermi Surface

The wave functions $\psi_n(q_1, \sigma_1, \dots, q_N, \sigma_N) \equiv \psi_n(\mathbf{Q}, \boldsymbol{\sigma})$ are now antisymmetric in the permutations of the pairs (q_i, σ_i) . Let $\psi(\mathbf{Q}, \boldsymbol{\sigma}; N, n)$ denote the n th

eigenfunction of the N -particle energy H_N in [66] with eigenvalue $E(N, n)$ (labeled by $n=0, 1, \dots$ and non-decreasingly ordered). Setting $\mathbf{Q}'' = (q_1'', \dots, q_{N-p}'')$, $\boldsymbol{\sigma}'' = (\sigma_1'', \dots, \sigma_{N-p}'')$, introduce the kernels $\rho_p^{H_N}(\mathbf{Q}, \boldsymbol{\sigma}; \mathbf{Q}', \boldsymbol{\sigma}')$ by

$$\rho_p(\mathbf{Q}, \boldsymbol{\sigma}; \mathbf{Q}', \boldsymbol{\sigma}') \stackrel{\text{def}}{=} p! \binom{N}{p} \int \sum_{\boldsymbol{\sigma}''} d^{N-p} \mathbf{Q}'' \sum_{n=0}^{\infty} \frac{e^{-\beta E(N, n)}}{\text{tr } e^{-\beta H_N}} \times \bar{\psi}(\mathbf{Q}, \boldsymbol{\sigma}; \mathbf{Q}'', \boldsymbol{\sigma}'') \psi(\mathbf{Q}', \boldsymbol{\sigma}'; \mathbf{Q}'', \boldsymbol{\sigma}'') \quad [69]$$

which are called p -particle reduced density matrices (extending the corresponding one-particle reduced density matrix [68]). Denote $\rho(q_1 - q_2) \stackrel{\text{def}}{=} \sum_{\sigma} \rho_1(q_1, \sigma, q_2, \sigma)$. It is also useful to consider spinless fermionic systems: the corresponding definitions are obtained simply by suppressing the spin labels and will not be repeated.

Let $r_1(\mathbf{k})$ be the Fourier transform of $\rho_1(\mathbf{q} - \mathbf{q}')$: the Fermi surface can be defined as the locus of the \mathbf{k} 's in the neighborhood of which $\partial_{\mathbf{k}} r_1(\mathbf{k})$ is unbounded as $\Omega \rightarrow \infty$, $\beta \rightarrow \infty$. The limit as $\beta \rightarrow \infty$ is important because the notion of a Fermi surface is, possibly, precise only at zero temperature, that is at $\beta = \infty$.

So far, existence of Fermi surface (i.e., the smoothness of $r_1(\mathbf{k})$ except on a smooth surface in \mathbf{k} -space) has been proved in free Fermi systems ($\varphi = 0$) and

1. certain exactly soluble one-dimensional spinless systems and
2. in rather general one-dimensional spinless systems or systems with spin and repulsive pair interaction, possibly in an external periodic potential.

The spinning case in a periodic potential and dimension $d \geq 2$ is the most interesting case to study for its relevance in the theory of conduction in crystals. Essentially no mathematical results are available as the above-mentioned ones do not concern any case in dimension > 1 : this is a rather deceiving aspect of the theory and a challenge.

In dimension 2 or higher, for fermionic systems with Hamiltonian [66], not only there are no results available, even without spin, but it is not even clear that a Fermi surface can exist in presence of interesting interactions.

Cooper Pairs

The superconductivity theory has been phenomenologically related to the existence of Cooper pairs. Consider the Hamiltonian [66] and define (cf. [69])

$$\rho(\mathbf{x} - \mathbf{y}, \sigma; \mathbf{x}' - \mathbf{y}', \sigma'; \mathbf{x} - \mathbf{x}') \stackrel{\text{def}}{=} \rho_2(\mathbf{x}, \sigma, \mathbf{y}, -\sigma; \mathbf{x}', \sigma', \mathbf{y}', -\sigma')$$

The system is said to contain *Cooper pairs* with spins $\sigma, -\sigma$ ($\sigma = +$ or $\sigma = -$) if there exist functions $g^{\alpha}(\mathbf{q}, \sigma) \neq 0$ with

$$\int \bar{g}^{\alpha}(\mathbf{q}, \sigma) g^{\alpha'}(\mathbf{q}, \sigma) d\mathbf{q} = 0 \quad \text{if } \alpha \neq \alpha'$$

such that

$$\lim_{V \rightarrow \infty} \rho(\mathbf{x} - \mathbf{y}, \sigma, \mathbf{x}' - \mathbf{y}', \sigma'; \mathbf{x} - \mathbf{x}') \xrightarrow{\mathbf{x} - \mathbf{x}' \rightarrow \infty} \sum_{\alpha} g^{\alpha}(\mathbf{x} - \mathbf{y}, \sigma) \bar{g}^{\alpha}(\mathbf{x}' - \mathbf{y}', \sigma') \quad [70]$$

In this case, $g^{\alpha}(\mathbf{x} - \mathbf{y}, \sigma)$ with largest L_2 norm can be called, after normalize, the wave function of the paired state of lowest energy: this is the analog of the plane wave for a free particle (and, like it, it is manifestly not normalizable, i.e., it is not square integrable as a function of \mathbf{x}, \mathbf{y}). If the system contains Cooper pairs and the nonleading terms in the limit [70] vanish quickly enough the two-particle reduced density matrix [70] regarded as a kernel operator has an eigenvalue of order V as $V \rightarrow \infty$: that is, the state of lowest energy is “macroscopically occupied,” quite like the free Bose condensation in the ground state.

Cooper pairs instability might destroy the Fermi surface in the sense that $r_1(\mathbf{k})$ becomes analytic in \mathbf{k} ; but it is also possible that, even in the presence of them, there remains a surface which is the locus of the singularities of the function $r_1(\mathbf{k})$. In the first case, there should remain a trace of it as a very steep gradient of $r_1(\mathbf{k})$ of the order of an exponential in the inverse of the coupling strength; this is what happens in the *BCS model* for superconductivity. The model is, however, a mean-field model and this particular regularity aspect might be one of its peculiarities. In any event, a smooth singularity surface is very likely to exist for some interesting density matrix (e.g., in the BCS model with “gap parameter γ ” the wave function

$$g(\mathbf{x} - \mathbf{y}, \sigma) \equiv \frac{1}{(2\pi)^d} \int_{\varepsilon(\mathbf{k}) > 0} e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{y})} \frac{\gamma}{\sqrt{\varepsilon(\mathbf{k})^2 + \gamma^2}} d\mathbf{k}$$

of the lowest energy level of the Cooper pairs is singular on a surface coinciding with the Fermi surface of the free system).

ODLRO

Consider the k -fermion reduced density matrix $\rho_k(\mathbf{Q}, \boldsymbol{\sigma}; \mathbf{Q}', \boldsymbol{\sigma}')$ as kernel operators O_k on $L_2((\Omega \times \mathcal{C}^2)^k)_-$. Suppose k is even, then if O_k has a (generalized) eigenvalue of order $N^{k/2}$ as $N \rightarrow \infty$, $N/V = \rho$, the system is said to exhibit *off-diagonal long-range order* of order k . For k odd, ODLRO is defined to exist if O_k has an eigenvalue of order $N^{(k-1)/2}$ and $k \geq 3$ (if $k = 1$ the largest eigenvalue of O_1 is necessarily ≤ 1).

For bosons, consider the reduced density matrix $\rho_k(\mathbf{Q}; \mathbf{Q}')$ regarding it as a kernel operator O_k on $L_2(\Omega)_+^k$ and define ODLRO of order k to be present if $O(k)$ has a (generalized) eigenvalue of order N^k as $N \rightarrow \infty, N/V = \rho$.

ODLRO can be regarded as a unification of the notions of Bose condensation and of the existence of Cooper pairs, because Bose condensation could be said to correspond to the kernel operator $\rho_1(\mathbf{q}_1 - \mathbf{q}_2)$ in [68] having a (generalized) eigenvalue of order N , and to be a case of ODLRO of order 1. If the state is pure in the sense that it has a cluster property (see the sections “Phase transitions and boundary conditions” and “Lattice models”), then the existence of ODLRO, Bose condensation, and Cooper pairs implies that the system shows a spontaneously broken symmetry: conservation of particle number and clustering imply that the off-diagonal elements of (all) reduced density matrices vanish at infinite separation in states obtained as limits of states with periodic boundary conditions and Hamiltonian [66], and this is incompatible with ODLRO.

The free Fermi gas has no ODLRO, the BCS model of superconductivity has Cooper pairs and ODLRO with $k=2$, but no Fermi surface in the above sense (possibly too strict). Fermionic systems cannot have ODLRO of order 1 (because the reduced density matrix of order 1 is bounded by 1).

The contribution of mathematical physics has been particularly effective in providing exactly soluble models: however, the soluble models deal with one-dimensional systems and it can be shown that in dimensions 1, 2 no ODLRO can take place. A major advance is the recent proof of ODLRO and Bose condensation in the case of a lattice version of [66] at a special density value (and $d \geq 3$).

In no case, for the Hamiltonian [66] with $\varphi \neq 0$, existence of Cooper pairs has been proved nor existence of a Fermi surface for $d > 1$. Nevertheless, both Bose condensation and Cooper pairs formation can be proved to occur rigorously in certain limiting situations. There are also a variety of phenomena (e.g., simple spectral properties of the Hamiltonians) which are believed to occur once some of the above-mentioned ones do occur and several of them can be proved to exist in concrete models.

If $d=1,2$, ODLRO can be proved to be impossible at $T > 0$ through the use of Bogoliubov’s inequality (used in the “no $d=2$ crystal theorem,” see the section “Continuous symmetries: ‘no $d=2$ crystal’ theorem”).

For more details, the reader is referred to Penrose and Onsager (1956), Yang (1962), Ruelle (1969), Hohenberg (1967), Gallavotti (1999), and Aizenman *et al.* (2004).

Appendix 1: The Physical Meaning of the Stability Conditions

It is useful to see what would happen if the conditions of stability and temperedness (see [14]) are violated. The analysis also illustrates some of the typical methods of statistical mechanics.

Coalescence Catastrophe due to Short-Distance Attraction

The simplest violation of the first condition in [14] occurs when the potential φ is smooth and negative at the origin.

Let $\delta > 0$ be so small that the potential at distances $\leq 2\delta$ is $\leq -b < 0$. Consider the canonical distribution with parameters β, N in a (cubic) box Ω of volume V . The probability P_{collapse} that all the N particles are located in a little sphere of radius δ around the center of the box (or around any prefixed point of the box) is estimated from below by remarking that

$$\Phi \leq -b \binom{N}{2} \sim -\frac{b}{2} N^2$$

so that

$$\begin{aligned} P_{\text{collapse}} &= \frac{\int_{\mathcal{C}} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}N!} e^{-\beta(K(\mathbf{p}) + \Phi(\mathbf{q}))}}{\int \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}N!} e^{-\beta(K(\mathbf{p}) + \Phi(\mathbf{q}))}} \\ &\geq \frac{\left(\frac{4\pi\sqrt{2m\beta^{-1}}}{3h^3}\right)^N \frac{\delta^{3N}}{N!} e^{\beta b(1/2)N(N-1)}}{\int \frac{d\mathbf{q}}{h^{3N}N!} e^{-\beta\Phi(\mathbf{q})}} \end{aligned} \quad [71]$$

The phase space is extremely small: nevertheless, such configurations are far more probable than the configurations which “look macroscopically correct,” that is, configurations with particles more or less spaced by the average particle distance expected in a macroscopically homogeneous configuration, namely $(N/V)^{-1/3} = \rho^{-1/3}$. Their energy $\Phi(\mathbf{q})$ is of the order of uN for some u , so that their probability will be bounded above by

$$\begin{aligned} P_{\text{regular}} &\leq \frac{\int \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}N!} e^{-\beta(K(\mathbf{p}) + uN)}}{\int \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}N!} e^{-\beta(K(\mathbf{p}) + \Phi(\mathbf{q}))}} \\ &= \frac{V^N \sqrt{2m\beta^{-1}}^3 e^{-\beta uN}}{\int \frac{d\mathbf{q}}{h^{3N}N!} e^{-\beta\Phi(\mathbf{q})}} \end{aligned} \quad [72]$$

However, no matter how small δ is, the ratio $P_{\text{regular}}/P_{\text{collapse}}$ will approach 0 as $V \rightarrow \infty$, $N/V \rightarrow \nu^{-1}$; this occurs extremely rapidly because $e^{\beta b N^2/2}$ eventually dominates over $V^N \sim e^{N \log N}$.

Thus, it is far more probable to find the system in a microscopic volume of size δ rather than in a configuration in which the energy has some macroscopic value proportional to N . This catastrophe can be called an ultraviolet catastrophe (as it is due to the behavior at very short distances) and it causes the collapse of the particles into configurations concentrated in regions as small as we please as $V \rightarrow \infty$.

Coalescence Catastrophe due to Long-Range Attraction

It occurs when the potential is too attractive near ∞ . For simplicity, suppose that the potential has a hard core, i.e., it is $+\infty$ for $r < r_0$, so that the above-discussed coalescence cannot occur and the system density bounded above by a certain quantity $\rho_{\text{cp}} < \infty$ (*close-packing density*).

The catastrophe occurs if $\varphi(q) \sim -g|q|^{-3+\varepsilon}$, $g, \varepsilon > 0$, for $|q|$ large. For instance, this is the case for matter interacting gravitationally; if k is the gravitational constant, m is the particle mass, then $g = km^2$ and $\varepsilon = 2$.

The probability P_{regular} of “regular configurations,” where particles are at distances of order $\rho^{-1/3}$ from their close neighbors, is compared with the probability P_{collapse} of “catastrophic configurations,” with the particles at distances r_0 from their close neighbors to form a configuration of density $\rho_{\text{cp}}/(1 + \delta)^3$ almost in close packing (so that r_0 is equal to the hard-core radius times $1 + \delta$). In the latter case, the system does not fill the available volume and leaves empty a region whose volume is a fraction $\sim ((\rho_{\text{cp}} - \rho)/\rho_{\text{cp}})V$ of V . Further, it can be checked that the ratio $P_{\text{regular}}/P_{\text{collapse}}$ tends to 0 at a rate $O(\exp(g\frac{1}{2}N(\rho_{\text{cp}}(1 + \delta)^{-3} - \rho)))$ if δ is small enough (and $\rho < \rho_{\text{cp}}$).

A system which is too attractive at infinity will not occupy the available volume but will stay confined in a close-packed configuration even in empty space.

This is important in the theory of stars: stars cannot be expected to obey “regular thermodynamics” and in particular will not “evaporate” because their particles interact via the gravitational force at large distances. Stars do not occupy the whole volume given to them (i.e., the universe); they do not collapse to a point only because the interaction has a strongly repulsive core (even when they are burnt out and the radiation pressure is no longer able to keep them at a reasonable size).

Evaporation Catastrophe

This is another *infrared catastrophe*, that is, a catastrophe due to the long-range structure of the

interactions in the above subsection; it occurs when the potential is too repulsive at ∞ , that is,

$$\varphi(q) \sim +g|q|^{-3+\varepsilon} \quad \text{as } q \rightarrow \infty$$

so that the temperedness condition is again violated.

In addition, in this case, the system does not occupy the whole volume: it will generate a layer of particles sticking, in close-packed configuration, to the walls of the container. Therefore, if the density is lower than the close-packing density, $\rho < \rho_{\text{cp}}$, the system will leave a region around the center of the container Ω empty; and the volume of the empty region will still be of the order of the total volume of the box (i.e., its diameter will be a fraction of the box side L). The proof is completely analogous to the one of the previous case; except that now the configuration with lowest energy will be the one sticking to the wall and close packed there, rather than the one close packed at the center.

Also this catastrophe is important as it is realized in systems of charged particles bearing the same charge: the charges adhere to the boundary in close-packing configuration, and dispose themselves so that the electrostatic potential energy is minimal. Therefore, charges deposited on a metal will not occupy the whole volume: they will rather form a surface layer minimizing the potential energy (i.e., so that the Coulomb potential in the interior is constant). In general, charges in excess of neutrality do not behave thermodynamically: for instance, besides not occupying the whole volume given to them, they will not contribute normally to the specific heat.

Neutral systems of charges behave thermodynamically if they have hard cores, so that the ultraviolet catastrophe cannot occur or if they obey quantum-mechanical laws and consist of fermionic particles (plus possibly bosonic particles with charges of only one sign).

For more details, we refer the reader to Lieb and Lebowitz (1972) and Lieb and Thirring (2001).

Appendix 2: The Subadditivity Method

A simple consequence of the assumptions is that the exponential in (5.2) can be bounded above by $e^{\beta B N} \exp(-\frac{\beta}{2m} \sum_{i=1}^N P_i^2)$ so that

$$\begin{aligned} 1 \leq Z_{\text{gc}}(\beta, \lambda, V) &\leq \exp\left(V e^{\beta \lambda} e^{\beta B} \sqrt{2m\beta^{-1}d}\right) \\ \Rightarrow 0 &\leq \frac{1}{V} \log Z_{\text{gc}}(\beta, \lambda, V) \leq e^{\beta \lambda} e^{\beta B} \sqrt{2m\beta^{-1}d} \quad [73] \end{aligned}$$

Consider, for simplicity, the case of a hard-core interaction with finite range (cf. [14]). Consider a

sequence of boxes Ω_n with sides $2^n L_0$, where $L_0 > 0$ is arbitrarily fixed to be $> 2R$. The partition function $Z_{gc}(\beta, z)$ relative to the volume Ω_n is

$$Z_n = \sum_{N=0}^{\infty} \frac{z^N}{N!} \int_{\Omega_n} d\mathbf{Q} e^{-\beta\Phi(\mathbf{Q})}$$

because the integral over the \mathbf{P} variables can be explicitly performed and included in z^N if z is defined as $z = e^{\beta\lambda(2m\beta^{-1})^{d/2}}$.

Then the box Ω_n contains 2^d boxes Ω_{n-1} for $n \geq 1$ and

$$1 \leq Z_n \leq Z_{n-1}^{2^d} \exp(\beta B 2^d (L_{n-1}/R)^{d-1} 2^{2d}) \quad [74]$$

because the corridor of width $2R$ around the boundaries of the 2^d cubes Ω_{n-1} filling Ω_n has volume $2RL_{n-1}2^d$ and contains at most $(L_{n-1}/R)^{d-1}2^d$ particles, each of which interacts with at most 2^d other particles. Therefore,

$$\beta p_n \stackrel{\text{def}}{=} L_n^d \log Z_n \\ \leq L_{n-1}^d \log Z_{n-1} + \beta B \gamma_d 2^{-n} (L_0/R)^{d-1}$$

for some $\gamma_d > 0$. Hence, $0 \leq \beta p_n \leq \beta p_{n-1} + \Gamma_d 2^{-n}$ for some $\Gamma_d > 0$ and p_n is bounded above and below uniformly in n . So, the limit [13] exists on the sequence $L_n = L_0 2^n$ and defines a function $\beta p_\infty(\beta, \lambda)$.

A box of arbitrary size L can be filled with about $(L/L_{\bar{n}})^d$ boxes of side $L_{\bar{n}}$ with \bar{n} so large that, prefixed $\delta > 0$, $|p_\infty - p_n| < \delta$ for all $n \geq \bar{n}$. Likewise, a box of size L_n can be filled by about $(L_n/L)^d$ boxes of size L if n is large. The latter remarks lead us to conclude, by standard inequalities, that the limit in [13] exists and coincides with p_∞ .

The subadditivity method just demonstrated for finite-range potentials with hard core can be extended to the potentials satisfying just stability and temperedness (cf. the section ‘‘Thermodynamic limit’’).

For more details, the reader is referred to Ruelle (1969) and Gallavotti (1999).

Appendix 3: An Infrared Inequality

The infrared inequalities stem from *Bogoliubov’s inequality*. Consider as an example the problem of crystallization discussed in the section ‘‘Continuous symmetries: ‘no $d=2$ crystal’ theorem’’. Let $\langle \cdot \rangle$ denote average over a canonical equilibrium state with Hamiltonian

$$H = \sum_{j=1}^N \frac{\mathbf{p}_j^2}{2} + U(\mathbf{Q}) + \varepsilon W(\mathbf{Q})$$

with given temperature and density parameters $\beta, \rho, \rho = a^{-3}$. Let $\{X, Y\} = \sum_i (\partial_{p_i} X \partial_{q_i} Y - \partial_{q_i} X \partial_{p_i} Y)$

be the Poisson bracket. Integration by parts, with periodic boundary conditions, yields

$$\langle A^* \{C, H\} \rangle \equiv - \frac{\int A^* \{C, e^{-\beta H}\} d\mathbf{P} d\mathbf{Q}}{\beta Z_c(\beta, \rho, N)} \\ \equiv -\beta^{-1} \langle \{A^*, C\} \rangle \quad [75]$$

as a general identity. The latter identity implies, for $A = \{C, H\}$, that

$$\langle \{H, C\}^* \{H, C\} \rangle = -\beta^{-1} \langle \{C, \{H, C^*\}\} \rangle \quad [76]$$

Hence, the Schwartz inequality $\langle A^* A \rangle \langle \{H, C\}^* \{H, C\} \rangle \geq |\langle \{A^*, C\} \rangle|^2$ combined with the two relations in [75], [76] yields Bogoliubov’s inequality:

$$\langle A^* A \rangle \geq \beta^{-1} \frac{|\langle \{A^*, C\} \rangle|^2}{\langle \{C, \{C^*, H\}\} \rangle} \quad [77]$$

Let g, h be arbitrary complex (differentiable) functions and $\partial_i = \partial_{q_i}$

$$A(\mathbf{Q}) \stackrel{\text{def}}{=} \sum_{j=1}^N g(\mathbf{q}_j), \quad C(\mathbf{P}, \mathbf{Q}) \stackrel{\text{def}}{=} \sum_{j=1}^N \mathbf{p}_j h(\mathbf{q}_j) \quad [78]$$

Then $H = \sum \frac{1}{2} \mathbf{p}_j^2 + \Phi(\mathbf{q}_1, \dots, \mathbf{q}_N)$, if

$$\Phi(\mathbf{q}_1, \dots, \mathbf{q}_N) = \frac{1}{2} \sum_{i \neq j} \varphi(|\mathbf{q}_i - \mathbf{q}_j|) + \varepsilon \sum_j W(\mathbf{q}_j)$$

so that, via algebra,

$$\{C, H\} \equiv \sum_j (h_j \partial_j \Phi - \mathbf{p}_j \cdot \partial_j h_j)$$

with $h_j \stackrel{\text{def}}{=} h(\mathbf{q}_j)$. If h is real valued, $\langle \{C, \{C^*, H\}\} \rangle$ becomes, again via algebra,

$$\left\langle \sum_{j \neq i} h_j h_i \partial_j \cdot \partial_i \Phi(\mathbf{Q}) \right\rangle \\ + \left\langle \varepsilon \sum_j h_j^2 \Delta W(\mathbf{q}_j) + \frac{4}{\beta} \sum_j (\partial_j h_j)^2 \right\rangle$$

(integrals on \mathbf{p}_j just replace \mathbf{p}_j^2 by $2\beta^{-1}$ and $\langle (\mathbf{p}_j)_i (\mathbf{p}_j)_{i'} \rangle = \beta^{-1} \delta_{i, i'}$). Therefore, the average $\langle \{C, \{C^*, H\}\} \rangle$ becomes

$$\left\langle \frac{1}{2} \sum_{j \neq i} (h_j - h_{j'})^2 \Delta \varphi(|\mathbf{q}_j - \mathbf{q}_{j'}|) \right. \\ \left. + \varepsilon \sum_j h_j^2 \Delta W(\mathbf{q}_j) + 4\beta^{-1} \sum_j (\partial_j h_j)^2 \right\rangle \quad [79]$$

Choose $g(\mathbf{q}) \equiv e^{-i(\boldsymbol{\kappa} + \mathbf{K}) \cdot \mathbf{q}}$, $h(\mathbf{q}) = \cos \mathbf{q} \cdot \boldsymbol{\kappa}$ and bound $(h_j - h_{j'})^2$ by $\boldsymbol{\kappa}^2 (\mathbf{q}_j - \mathbf{q}_{j'})^2$, $(\partial_j h_j)^2$ by $\boldsymbol{\kappa}^2$ and

b_j^2 by 1. Hence [79] is bounded above by $ND(\boldsymbol{\kappa})$ with

$$D(\boldsymbol{\kappa}) \stackrel{\text{def}}{=} \left\langle \kappa^2 \left(4\beta^{-1} + \frac{1}{2N} \sum_{i \neq j} (q_i - q_j)^2 |\Delta\varphi(q_i - q_j)| \right) + \varepsilon \frac{1}{N} \sum_j |\Delta W(q_j)| \right\rangle \quad [80]$$

This can be used to estimate the denominator in [77]. For the LHS remark that

$$\langle A^*, A \rangle = \left| \sum_{j=1}^N e^{-iq \cdot (\boldsymbol{\kappa} + \mathbf{K})} \right|^2$$

and

$$\begin{aligned} |\langle \{A^*, C\} \rangle|^2 &= \left\langle \left| \sum_j b_j \partial g_j \right|^2 \right\rangle \\ &= |\mathbf{K} + \boldsymbol{\kappa}|^2 N^2 (\rho_\varepsilon(\mathbf{K}) + \rho_\varepsilon(\mathbf{K} + 2\boldsymbol{\kappa}))^2 \end{aligned}$$

hence [77] becomes, after multiplying both sides by the auxiliary function $\gamma(\boldsymbol{\kappa})$ (assumed even and vanishing for $|\boldsymbol{\kappa}| > \pi/a$) and summing over $\boldsymbol{\kappa}$,

$$\begin{aligned} D_1 &\stackrel{\text{def}}{=} \frac{1}{N} \sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa}) \left\langle \frac{1}{N} \left| \sum_{j=1}^N e^{-i(\mathbf{K} + \boldsymbol{\kappa}) \cdot q_j} \right|^2 \right\rangle \\ &\geq \frac{1}{N} \sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa}) \\ &\quad \times \frac{|\mathbf{K}|^2 (\rho_\varepsilon(\mathbf{K}) + \rho_\varepsilon(\mathbf{K} + 2\boldsymbol{\kappa}))^2}{4\beta D(\boldsymbol{\kappa})} \quad [81] \end{aligned}$$

To apply [77] the averages in [80], [81] have to be bounded above: this is a technical point that is discussed here, as it illustrates a general method of using the results on the thermodynamic limits and their convexity properties to obtain estimates.

Note that $\langle (1/N) \sum_{\mathbf{k}} \gamma(\mathbf{k}) d^d \mathbf{k} \left| \sum_{j=1}^N e^{-i\mathbf{k} \cdot q_j} \right|^2 \rangle$ is identically $\tilde{\varphi}(0) + (2/N) \langle \sum_{j < j'} \tilde{\varphi}(q_j - q_{j'}) \rangle$ with $\tilde{\varphi}(q) \stackrel{\text{def}}{=} (1/N) \sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa}) e^{i\boldsymbol{\kappa} \cdot q}$.

Let $\varphi_{\lambda, \zeta}(q) \stackrel{\text{def}}{=} \varphi(q) + \lambda q^2 |\Delta\varphi(q)| + \eta \tilde{\varphi}(q)$ and let $F_V(\lambda, \eta, \zeta) \stackrel{\text{def}}{=} (1/N) \log Z^c(\lambda, \eta, \zeta)$ with Z^c the partition function in the volume Ω computed with energy $U' = \sum_{j \neq j'} \varphi_{\lambda, \zeta}(q_j - q_{j'}) + \varepsilon \sum_j W(q_j) + \eta \varepsilon \sum |\Delta W(q_j)|$. Then $F_V(\lambda, \eta, \zeta)$ is convex in λ, η and it is uniformly bounded above and below if $|\eta|, |\varepsilon|, |\zeta| \leq 1$ (say) and $|\lambda| \leq \lambda_0$: here $\lambda_0 > 0$ exists if $r^2 |\Delta\varphi(r)|$ satisfies the assumption set at the beginning of the section ‘‘Continuous symmetries: ‘no $d = 2$ crystal’ theorem’’ and the density is smaller than a close packing (this is because the potential U' will still satisfy conditions similar to [14] uniformly in $|\varepsilon|, |\eta| < 1$ and $|\lambda|$ small enough).

Convexity and boundedness above and below in an interval imply bounds on the derivatives in

the interior points, in this case on the derivatives of F_V with respect to λ, η, ζ at 0. The latter are identical to the averages in [80], [81]. In this way, the constants B_1, B_2, B_0 such that $D(\boldsymbol{\kappa}) \leq \kappa^2 B_1 + \varepsilon B_2$ and $B_0 > D_1$ are found.

For more details, the reader is referred to Mermin (1968).

Further Reading

- Aizenman M (1980) Translation invariance and instability of phase coexistence in the two dimensional Ising system. *Communications in Mathematical Physics* 73: 83–94.
- Aizenman M (1982) *Geometric analysis of φ^4 fields and Ising models*. 86: 1–48.
- Aizenman M, Lieb EH, Seiringer R, Solovej JP, and Yngvason J (2004) Bose–Einstein condensation as a quantum phase transition in a optical lattice, *Physical Review A* 70: 023612.
- Baxter R (1982) *Exactly Solved Models*. London: Academic Press.
- Benfatto G and Gallavotti G (1995) *Renormalization group*. Princeton: Princeton University Press.
- Bleher P and Sinai Y (1975) Critical indices for Dyson’s asymptotically hierarchical models. *Communications in Mathematical Physics* 45: 247–278.
- Boltzmann L (1968a) Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie. In: Hasenöhr F (ed.) *Wissenschaftliche Abhandlungen*, vol. I, pp. 9–33. New York: Chelsea.
- Boltzmann L (1968b) Über die Eigenschaften monzyklischer und anderer damit verwandter Systeme. In: Hasenöhr FP (ed.) *Wissenschaftliche Abhandlungen*, vol. III, pp. 122–152. New York: Chelsea.
- Dobrushin RL (1968) Gibbsian random fields for lattice systems with pairwise interactions. *Functional Analysis and Applications* 2: 31–43.
- Domb C and Green MS (1972) *Phase Transitions and Critical Points*. New York: Wiley.
- Dyson F (1969) Existence of a phase transition in a one–dimensional Ising ferromagnet. *Communications in Mathematical Physics* 12: 91–107.
- Dyson F and Lenard A (1967, 1968) Stability of matter. *Journal of Mathematical Physics* 8: 423–434, 9: 698–711.
- Friedli S and Pfister C (2004) On the singularity of the free energy at a first order phase transition. *Communications in Mathematical Physics* 245: 69–103.
- Gallavotti G (1999) *Statistical Mechanics*. Berlin: Springer.
- Gallavotti G, Bonetto F and Gentile G (2004) *Aspects of the Ergodic, Qualitative and Statistical Properties of Motion*. Berlin: Springer.
- Gawedzky K and Kupiainen A (1983) Block spin renormalization group for dipole gas and $(\partial\phi)^4$. *Annals of Physics* 147: 198–243.
- Gawedzky K and Kupiainen A (1985) Massless lattice ϕ_4^4 theory: rigorous control of a renormalizable asymptotically free model. *Communications in Mathematical Physics* 99: 197–252.
- Gibbs JW (1981) *Elementary Principles in Statistical Mechanics*. Woodbridge (Connecticut): Ox Bow Press (reprint of the 1902 edition).
- Higuchi Y (1981) On the absence of non translationally invariant Gibbs states for the two dimensional Ising system. In: Fritz J, Lebowitz JL, and Szaz D (eds.) *Random Folds*. Amsterdam: North-Holland.
- Hohenberg PC (1967) Existence of long range order in one and two dimensions. *Physical Review* 158: 383–386.

- Landau L and Lifschitz LE (1967) *Physique Statistique*. Moscow: MIR.
- Lanford O and Ruelle D (1969) Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics* 13: 194–215.
- Lebowitz JL (1974) GHS and other inequalities. *Communications in Mathematical Physics* 28: 313–321.
- Lebowitz JL and Penrose O (1979) Towards a rigorous molecular theory of metastability. In: Montroll EW and Lebowitz JL (eds.) *Fluctuation Phenomena*. Amsterdam: North-Holland.
- Lee TD and Yang CN (1952) Statistical theory of equations of state and phase transitions, II. Lattice gas and Ising model. *Physical Review* 87: 410–419.
- Lieb EH (2002) *Inequalities*. Berlin: Springer.
- Lieb EH and Lebowitz JL (1972) Lectures on the Thermodynamic Limit for Coulomb Systems, In: Lenard A (ed.) *Springer Lecture Notes in Physics*, vol. 20, pp. 135–161. Berlin: Springer.
- Lieb EH and Thirring WE (2001) *Stability of Matter from Atoms to Stars*. Berlin: Springer.
- Mastropietro V (2004) Ising models with four spin interaction at criticality. *Communications in Mathematical Physics* 244: 595–642.
- McCoy BM and Wu TT (1973) *The two Dimensional Ising Model*. Cambridge: Harvard University Press.
- Mermin ND (1968) Crystalline order in two dimensions. *Physical Review* 176: 250–254.
- Miracle-Solé S (1995) Surface tension, step free energy and facets in the equilibrium crystal shape. *Journal Statistical Physics* 79: 183–214.
- Olla S (1987) Large deviations for Gibbs random fields. *Probability Theory and Related Fields* 77: 343–357.
- Onsager L (1944) Crystal statistics. I. A two dimensional Ising model with an order–disorder transition. *Physical Review* 65: 117–149.
- Penrose O and Onsager L (1956) Bose–Einstein condensation and liquid helium. *Physical Review* 104: 576–584.
- Pfister C and Velenik Y (1999) Interface, surface tension and Reentrant pinning transition in the 2D Ising model. *Communications in Mathematical Physics* 204: 269–312.
- Ruelle D (1969) *Statistical Mechanics*. New York: Benjamin.
- Ruelle D (1971) Extension of the Lee–Yang circle theorem. *Physical Review Letters* 26: 303–304.
- Sinai Ya G (1991) *Mathematical Problems of Statistical Mechanics*. Singapore: World Scientific.
- van Beyeren H (1975) Interphase sharpness in the Ising model. *Communications in Mathematical Physics* 40: 1–6.
- Wilson KG and Fisher ME (1972) Critical exponents in 3.99 dimensions. *Physical Review Letters* 28: 240–243.
- Yang CN (1962) Concept of off-diagonal long-range order and the quantum phases of liquid He and of superconductors. *Reviews of Modern Physics* 34: 694–704.

Introductory Article: Functional Analysis

S Paycha, Université Blaise Pascal, Aubière, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Functional analysis is concerned with the study of functions and function spaces, combining techniques borrowed from classical analysis with algebraic techniques. Modern functional analysis developed around the problem of solving equations with solutions given by functions. After the differential and partial differential equations, which were studied in the eighteenth century, came the integral equations and other types of functional equations investigated in the nineteenth century, at the end of which arose the need to develop a new analysis, with functions of an infinite number of variables instead of the usual functions. In 1887, Volterra, inspired by the calculus of variations, suggested a new infinitesimal calculus where usual functions are replaced by functionals, that is, by maps from a function space to \mathbb{R} or \mathbb{C} , but he and his followers were still missing some algebraic and topological tools to be developed later. Modern analysis was born with the development of an “algebra of the infinite” closely related to classical linear algebra which by 1890 had (up to the concept of duality,

which was developed later) settled on firm ground. Strongly inspired by algebraic methods, Fredholm’s work at the turn of the nineteenth century, in which emerged the concept of kernel of an operator, became a founding stone for the modern theory of integral equations. Hilbert developed further Fredholm’s methods for symmetric kernels, exploiting analogies with the theory of real quadratic forms and thereby making clear the importance of the notion of square-integrable functions. With Hilbert’s *Grundzüge einer allgemeinen Theorie der Integralgleichung*, a further step was made from the “algebra of the infinite” to the “geometry of the infinite.” The contribution of Fréchet, who introduced the abstract notion of a space endowed with a distance, made it possible to transfer Euclidean geometry to the framework of what have since then been called Hilbert spaces, a basic concept in mathematics and quantum physics.

The usefulness of functional analysis in the study of quantum systems became clear in the 1950s when Kato proved the self-adjointness of atomic Hamiltonians, and Garding and Wightman formulated axioms for quantum field theory. Ever since functional analysis lies at the very heart of many approaches to quantum field theory. Applications of functional analysis stretch out to many branches of mathematics, among which are numerical

analysis, global analysis, the theory of pseudodifferential operators, differential geometry, operator algebras, noncommutative geometry, etc.

Topological Vector Spaces

Most topological spaces one comes across in practice are metric spaces. A metric on a topological space E is a map $d: E \times E \rightarrow [0, +\infty[$ which is symmetric, such that $d(u, v) = 0 \Leftrightarrow u = v$ and which verifies the triangle inequality $d(u, w) \leq d(u, v) + d(v, w)$ for all vectors u, v, w . A topological space E is metrizable if there is a metric d on E compatible with the topology on E , in which case the balls with radius $1/n$ centered at any point $x \in E$ form a local base at x – that is, a collection of neighborhoods of x such that every neighborhood of x contains a member of this collection. A sequence (u_n) in E then converges to $u \in E$ if and only if $d(u_n, u)$ converges to 0.

The Banach fixed-point theorem on a complete metric space (E, d) is a useful tool in nonlinear functional analysis: it states that a (strict) contraction on E , that is, a map $T: E \rightarrow E$ such that $d(Tu, Tv) \leq k d(u, v)$ for all $u \neq v \in E$ and fixed $0 < k < 1$, has a unique fixed point $Tu_0 = u_0$. In particular, it provides local existence and uniqueness of solutions of differential equations $dy/dt = F(y, t)$ with initial condition $y(0) = y_0$, where F is Lipschitz continuous.

Linear functional analysis starts from topological vector spaces, that is, vector spaces equipped with a topology for which the operations are continuous. A topological vector space equipped with a local base whose members are convex is said to be locally convex. Examples of locally convex spaces are normed linear spaces, namely vector spaces equipped with a norm, a concept that first arose in the work of Fréchet. A seminorm on a vector space V is a map $\rho: V \rightarrow [0, \infty[$ which obeys the triangle identity $\rho(u + v) \leq \rho(u) + \rho(v)$ for any vectors u, v and such that $\rho(\lambda u) = |\lambda| \rho(u)$ for any scalar λ and any vector u ; if $\rho(u) = 0 \Rightarrow u = 0$, it is a norm, often denoted by $\|\cdot\|$. A norm on a vector space E gives rise to a translation-invariant distance function $d(u, v) = \|u - v\|$ making it a metric space.

Historically, one of the first examples of normed spaces is the space $C([0, 1])$ investigated by Riesz of (real- or complex-valued) continuous functions on the interval $[0, 1]$ equipped with the supremum norm $\|f\|_\infty := \sup_{x \in [0, 1]} |f(x)|$. In the 1920s, the general definition of Banach space arose in connection with the works of Hahn and Banach. A normed linear space is a Banach space if it is complete as a metric space for the induced metric, $C([0, 1])$ being a prototype of a Banach space. More generally, for

any non-negative integer k , the space $C^k([0, 1])$ of functions on $[0, 1]$ of class C^k equipped with the norm $\|f\|_k = \sum_{i=0}^k \|f^{(i)}\|_\infty$ expressed in terms of a finite number of seminorms $\|f^{(i)}\|_\infty = \sup_{x \in [0, 1]} |f^{(i)}(x)|, i = 0, \dots, k$, is also a Banach space.

The space $C^\infty([0, 1])$ of smooth functions on the interval $[0, 1]$ is not anymore a Banach space since its topology is described by a countable family of seminorms $\|f\|_k$ with k varying in the positive integers. The metric

$$d(f, g) = \sum_{k=1}^{\infty} 2^{-k} \frac{\|f - g\|_k}{1 + \|f - g\|_k}$$

turns it into a Fréchet space, that is, a locally convex complete metric space. The space $\mathcal{S}(\mathbb{R}^n)$ of rapidly decreasing functions, which are smooth functions f on \mathbb{R}^n for which

$$\|f\|_{\alpha, \beta} := \sup_{x \in \mathbb{R}^n} |x^\alpha D_x^\beta f(x)|$$

is finite for any multiindices α and β , is also a Fréchet space with the topology given by the seminorms $\|\cdot\|_{\alpha, \beta}$. Further examples of Fréchet spaces are the space $C_0^\infty(K)$ of smooth functions with support in a fixed compact subset $K \subset \mathbb{R}^n$ equipped with the countable family of seminorms

$$\|D^\alpha f\|_{\infty, K} = \sup_{x \in K} |D_x^\alpha f(x)|, \quad \alpha \in \mathbb{N}^n$$

and the space $C^\infty(M, E)$ of smooth sections of a vector bundle E over a closed manifold M equipped with a similar countable family of seminorms. Given an open subset $\Omega = \cup_{p \in \mathbb{N}} K_p$ with $K_p, p \in \mathbb{N}$ compact subsets of \mathbb{R}^n , the space $\mathcal{D}(\Omega) = \cup_{p \in \mathbb{N}} C_0^\infty(K_p)$ equipped with the inductive limit topology – for which a sequence (f_n) in $\mathcal{D}(\Omega)$ converges to $f \in \mathcal{D}(\Omega)$ if each f_n has support in some fixed compact subset K and $(D^\alpha f_n)$ converges uniformly to $D^\alpha f$ on K for each multiindex α – is a locally convex space.

Among Banach spaces are Hilbert spaces which have properties very similar to those of finite-dimensional spaces and are historically the first type of infinite-dimensional space to appear with the works of Hilbert at the beginning of the twentieth century. A Hilbert space is a Banach space equipped with a norm $\|\cdot\|$ that derives from an inner product, that is, $\|u\|^2 = \langle u, u \rangle$ with $\langle \cdot, \cdot \rangle$ a positive-definite bilinear (or sesquilinear according to whether the base space is real or complex) form. Hilbert spaces are fundamental building blocks in quantum mechanics; using (closed) tensor products, from a Hilbert space H one builds the Fock space $\mathcal{F}(H) = \sum_{k=0}^{\infty} \otimes^k H$ and from there the bosonic Fock space $\mathcal{F}(H) = \sum_{k=0}^{\infty} \otimes_s^k H$ (where \otimes_s stands for the (closed) symmetrized tensor product) as well

as the fermionic Fock space $\mathcal{F}(H) = \sum_{k=0}^{\infty} \Lambda^k H$ (where Λ^k stands for the antisymmetrized (closed) tensor product).

A prototype of Hilbert space is the space $l_2(\mathbb{Z})$ of complex-valued sequences $(u_n)_{n \in \mathbb{Z}}$ such that $\sum_{n \in \mathbb{Z}} |u_n|^2$ is finite, which is already implicit in Hilbert's Grundzügen. Shortly afterwards, Riesz and Fischer, with the help of the integration tool introduced by Lebesgue, showed that the space $L^2(]0, 1[)$ (first introduced by Riesz) of square-summable functions on the interval $]0, 1[$, that is, functions f such that

$$\|f\|_{L^2} = \left(\int_0^1 |f(x)|^2 dx \right)^{1/2}$$

is finite, provides an example of Hilbert space. These were then further generalized to spaces $L^p(]0, 1[)$ of p -summable ($1 \leq p < \infty$) functionals on $]0, 1[$ (i.e., functions f such that

$$\|f\|_{L^p} = \left(\int_0^1 |f(x)|^p dx \right)^{1/p}$$

is finite), which are not Hilbert unless $p = 2$ but which provide further examples of Banach spaces, the space $L^\infty(]0, 1[)$ of functions on $]0, 1[$ bounded almost everywhere with respect to the Lebesgue measure, offering yet another example of Banach space.

In 1936, Sobolev gave a generalization of the notion of function and their derivatives through integration by parts, which led to the so-called Sobolev spaces $W^{k,p}(]0, 1[)$ of functions $f \in L^p(]0, 1[)$ with derivatives up to order k lying in $L^p(]0, 1[)$, obtained as the closure of $C^\infty(]0, 1[)$ for the norm

$$f \mapsto \|f\|_{W^{k,p}} = \left(\sum_{j=1}^k \|\partial^j f\|_{L^p}^p \right)^{1/p}$$

(for $p = 2$, $W^{k,p}(]0, 1[)$ is a Hilbert space often denoted by $H^k(]0, 1[)$). They differ from the Sobolev spaces $W_0^{k,p}(]0, 1[)$, which correspond to the closure of the set $\mathcal{D}(]0, 1[)$ for the norm $f \mapsto \|f\|_{W^{k,p}}$; for example, an element $u \in W^{1,p}(]0, 1[)$ lies in $W_0^{1,p}(]0, 1[)$ if and only if it vanishes at 0 and 1, that is, if and only if it satisfies Dirichlet-type boundary conditions on the boundary of the interval. Similarly, one defines Sobolev spaces $W_0^{k,p}(\mathbb{R}) = W^{k,p}(\mathbb{R})$ on \mathbb{R} , Sobolev spaces $W^{k,p}(\Omega)$ and $W_0^{k,p}(\Omega)$ on open subsets $\Omega \subset \mathbb{R}^n$ and using a partition of unity on a closed manifold M , Sobolev spaces $H^k(M, E) = W^{k,2}(M, E)$ of sections of vector bundles E over M . Using the Fourier transform (discussed later), one can drop the assumption that k be an integer and extend the notion of Sobolev space

to define $W^{s,p}(\Omega)$ and $H^s(M, E)$ with s any real number.

Sobolev spaces arise in many areas of mathematics; one central example in probability theory is the Cameron–Martin space $H^1([0, t])$ embedded in the Wiener space $C([0, t])$. This embedding is a particular case of more general Sobolev embedding theorems, which embed (possibly continuously, sometimes even compactly (the notion of compact operator is discussed in a later section)) $W^{k,p}$ -Sobolev spaces in L^q -spaces with $q > p$ such as the continuous inclusion $W^{k,p}(\mathbb{R}^n) \subset L^q(\mathbb{R}^n)$ with $1/q = 1/p - k/n$, or in C^l -spaces with $l \leq k$ such as, for a bounded open and regular enough subset Ω of \mathbb{R}^n and for any $s \geq l + n/p$ with $p > n$, the continuous inclusion $W^{s,p}(\Omega) \subset C^l(\bar{\Omega})$ (the set of functions in $C^l(\Omega)$ such that $D^\alpha u$ can be continuously extended to the closure $\bar{\Omega}$ for all $|\alpha| \leq l$). Sobolev embeddings have important applications for the regularity of solutions of partial differential equations, when showing that weak solutions one constructs are in fact smooth. In particular, on an n -dimensional closed manifold M for $s > l + n/2$, the Sobolev space $H^s(M, E)$ can be continuously embedded in the space $C^l(M, E)$ of sections of E of class C^l , which in particular implies that the solutions of a hypoelliptic partial differential equation $Au = v$ with $v \in L^2(M, E)$ are smooth, as for example in the case of solutions of the Seiberg–Witten equations.

Duality

The concept of duality (in a topological sense) was initiated at the beginning of the twentieth century by Hadamard, who was looking for continuous linear functionals on the Banach space $C(I)$ of continuous functions on a compact interval I equipped with a uniform topology. It is implicit in Hilbert's theory and plays a central part in Riesz' work, who managed to express such continuous functionals as Stieltjes integrals, one of the starting points for the modern theory of integration.

The topological dual of a topological vector space E is the space E^* of continuous linear forms on E which, when E is a normed space, can be equipped with the dual norm $\|L\|_{E^*} = \sup_{u \in E, \|u\| \leq 1} |L(u)|$.

Dual spaces often provide a receptacle for singular objects; any of the functions $f \in L^p(\mathbb{R}^n)$ ($p \geq 1$) and the delta-function at point $x \in \mathbb{R}^n$, $\delta_x : f \mapsto f(x)$, all lie in the space $\mathcal{S}'(\mathbb{R}^n)$ dual to $\mathcal{S}(\mathbb{R}^n)$ of tempered distributions on \mathbb{R}^n , which is itself contained in the space $\mathcal{D}'(\mathbb{R}^n)$ of distributions dual to $\mathcal{D}(\mathbb{R}^n)$. Furthermore, the topological dual E^* of a nuclear space E contains the support of a probability

measure with characteristic function (see the next section) given by a continuous positive-definite function on E . Among nuclear spaces are projective limits $E = \bigcap_{p \in \mathbb{N}} H_p$ (a sequence $(u_n) \in E$ converges to $u \in E$ whenever it converges to u in each H_p) of countably many nested Hilbert spaces $\cdots \subset H_p \subset H_{p-1} \subset \cdots \subset H_0$ such that the embedding $H_p \subset H_{p-1}$ is a trace-class operator (see the section “Operator algebras”). If H_p is the closure of E for the norm $\|\cdot\|_p$, the topological dual E' of E for the norm $\|\cdot\|_0$ is an inductive limit $E' = \bigcup_{p \in \mathbb{N}_0} H_{-p}$, where H_{-p} are the dual (with respect to $\|\cdot\|_0$) Hilbert spaces with norm $\|\cdot\|_{-p}$ (a sequence $(u_n) \in E'$ converges to $u \in E'$ whenever it lies in some H_{-p} and converges to u for the topology of H_{-p}) and we have

$$\begin{aligned} E &\subset \cdots \subset H_p \subset H_{p-1} \subset \cdots \subset H_0 \\ &= H'_0 \subset H_{-1} \subset \cdots \subset H_{-p} \subset \cdots \subset E' \end{aligned}$$

As a result of the theory of elliptic operators on a closed manifold, the Fréchet space $C^\infty(M, E)$ of smooth sections of a vector bundle over a closed manifold M is nuclear as the inductive limit of countably many Sobolev spaces $H^p(M, E)$ with L^2 -dual given by the projective limit of countably many Sobolev spaces $H^{-p}(M, E)$.

The existence of nontrivial continuous linear forms on a normed linear space E is ensured by the Hahn–Banach theorem, which asserts that for any closed linear subspace F of E , there is a nonvanishing continuous linear form that vanishes on F . When the space is a Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$, it follows from the Riesz–Fréchet theorem that any continuous linear form L on H is represented in a unique way by a vector $v \in H$ such that $L(u) = \langle v, u \rangle_H$ for all $u \in H$, thus relating the dual pairing on the left with the Hilbert inner product on the right and identifying the topological dual H^* with H .

The strong topology induced by the norm $\|\cdot\|$ on a normed vector space E – that is, the topology in which a sequence (u_n) converges to u whenever $\|u_n - u\| \rightarrow 0$ – is too refined to have compact sets when E is infinite dimensional since the compactness of the unit ball in E for the strong topology characterizes finite-dimensional spaces. Since compact sets are useful for existence theorems, one is inclined to weaken the topology: the weak topology on E – which coincides with the strong topology when E is finite dimensional and for which a sequence (u_n) converges to u if and only if $L(u_n) \rightarrow L(u) \forall L \in E^*$ – has compact unit ball if and only if E is reflexive or, in other words, if E can be canonically identified with its double dual $(E^*)^*$. For $1 < p < \infty$, given an open subset $\Omega \subset \mathbb{R}^n$, the topological dual of

$L^p(\Omega)$ can be identified via the Riesz representation with $L^{p^*}(\Omega)$ with p^* conjugate to p , that is, $1/p + 1/p^* = 1$ and $L^p(\Omega)$ is reflexive, whereas the topological duals of $W^{s,p}(\Omega)$ and $W_0^{s,p}(\Omega)$ both coincide with $W_0^{-s,p^*}(\Omega)$ so that only $W_0^{s,p}(\Omega)$ is reflexive. Neither $L^1(\Omega)$ nor its topological dual $L^\infty(\Omega)$ is reflexive since $L^1(\Omega)$ is strictly contained in the topological dual of $L^\infty(\Omega)$ for there are continuous linear forms L on $L^\infty(\Omega)$ that are not of the form

$$L(u) = \int_{\Omega} uv \quad \forall u \in L^\infty(\Omega) \quad \text{with } v \in L^1(\Omega)$$

Similarly, the topological dual E^* of a normed linear space E can be equipped with the topology induced by the dual norm $\|\cdot\|_{E^*}$ and the the weak *-topology, namely the weakest one for which the maps $L \mapsto L(u), u \in E$, are continuous, and the unit ball in E^* is indeed compact for this topology (Banach–Alaoglu theorem).

Duality does not always preserve separability – a topological vector space is separable if it has a countable dense subspace – since $L^\infty(\Omega)$, which is not separable, is the topological dual of $L^1(\Omega)$, which is separable. However, as a consequence of the Hahn–Banach theorem, if the topological dual of a Banach space is separable then so is the original space and one has equivalence when adding the reflexivity assumption; a Banach space is reflexive and separable whenever its topological dual is. For $1 \leq p < \infty$, $L^p(\Omega)$ and $W_0^{s,p}(\Omega)$ are separable and moreover reflexive if $p \neq 1$.

Fourier Transform

In the middle of the eighteenth century, oscillations of a vibrating string were interpreted by Bernoulli as a limit case for the oscillation of n -point masses when n tends the infinity, and Bernoulli introduced the novel idea of the superposition principle by which the general oscillation of the string should decompose in a superposition of “proper oscillations.” This point of view triggered off a discussion as to whether or not an arbitrary function can be expanded as a trigonometric series. Other examples of expansions in “orthogonal functions” (this terminology actually only appears with Hilbert) had been found in the mean time in relation to oscillation problems and investigations on heat theory, but it was only in the nineteenth century, with the works of Fourier and Dirichlet, that the superposition problem was solved.

Separable Hilbert spaces can be equipped with a countable orthonormal system $\{e_n\}_{n \in \mathbb{Z}}$ ($\langle e_n, e_m \rangle_H = \delta_{nm}$ with $\langle \cdot, \cdot \rangle_H$ the scalar product on H) which is

complete, that is, any vector $u \in H$ can be expanded in this system in a unique way $u = \sum_{n \in \mathbb{Z}} \hat{u}_n e_n$ with Fourier coefficients $\hat{u}_n = \langle u, e_n \rangle$. The latter obey Parseval's relation $\sum_{n \in \mathbb{Z}} |\hat{u}_n|^2 = \|u\|^2$ (where $\|\cdot\|$ is the norm associated with $\langle \cdot, \cdot \rangle$), and the Fourier transform $u \mapsto (\hat{u}(n))_{n \in \mathbb{Z}}$ gives rise to an isometric isomorphism between the separable Hilbert space H and the Hilbert space $l^2(\mathbb{Z})$ of square-summable sequences of complex numbers. In particular, the space $L^2(S^1)$ of L^2 -functions on the unit circle $S^1 = \mathbb{R}/\mathbb{Z}$ with its usual Haar measure dt is separable with complete orthonormal system $t \mapsto e_n(t) = e^{2i\pi nt}$, $n \in \mathbb{Z}$ and the Fourier transform

$$u \mapsto \left(t \mapsto \hat{u}(n) = \int_0^1 e^{-2i\pi nt} u(t) dt \right)_{n \in \mathbb{Z}}$$

identifies it with the space $l^2(\mathbb{Z})$. Under this identification, the Hilbert subspace $l^2(\mathbb{N})$ obtained as the range in $l^2(\mathbb{Z})$ of the projection $p_+ : (u)_{n \in \mathbb{Z}} \mapsto (u_n)_{n \in \mathbb{N}}$ corresponds to the Hardy space $\mathcal{H}^2(S^1)$.

The Fourier transform extends to the space $\mathcal{S}(\mathbb{R}^n)$, sending a function $f \in \mathcal{S}(\mathbb{R}^n)$ to the map

$$\xi \mapsto \hat{f}(\xi) = \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} e^{-i\xi \cdot x} f(x) dx$$

and maps $\mathcal{S}(\mathbb{R}^n)$ onto itself linearly and continuously with continuous inverse $f \mapsto \hat{f}(-\xi)$. When $n = 1$, the Poisson formula relates $f \in \mathcal{S}(\mathbb{R})$ with its Fourier transform \hat{f} by $\sum_{n=-\infty}^{\infty} f(2\pi n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$.

Since Fourier transformation turns (up to a constant multiplicative factor) differentiation D_ξ^α for a multiindex $\alpha = (\alpha_1, \dots, \alpha_n)$ into multiplication by $\xi^\alpha = \xi_1^{\alpha_1} \dots \xi_n^{\alpha_n}$, it can be used to define $W^{s,p}$ -Sobolev spaces with s a real number as the space of L^p -functions with finite Sobolev norms $\|u\|_{W^{s,p}} = (\int |(1 + |\xi|)^s \hat{u}(\xi)|^p)^{1/p}$ (which coincide with the ones defined previously when $s = k$ is a non-negative integer).

Fourier transforms are also used to describe a linear pseudodifferential operator A (see next two sections where the notions of bounded and unbounded linear operator are discussed) of order a acting on smooth functions on an open subset U of \mathbb{R}^n in terms of its symbol σ_A – a smooth map σ on $U \times \mathbb{R}^n$ with compact support in x such that for any multi-indices $\alpha, \beta \in \mathbb{N}_0^n$, there is a constant $C_{\alpha,\beta}$ with

$$|D_x^\alpha D_\xi^\beta \sigma(x, \xi)| \leq C_{\alpha,\beta} (1 + |\xi|)^{a-|\beta|}$$

for any $\xi \in \mathbb{R}^n - \text{by}$

$$(Af)(x) = \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} e^{-ix \cdot \xi} \sigma_A(x, \xi) \hat{f}(\xi) d\xi$$

Fourier transform maps a Gaussian function $x \mapsto e^{-(1/2)\lambda|x|^2}$ on \mathbb{R}^n , where λ is a nonzero scalar, to another Gaussian function $\xi \mapsto e^{-(1/2)\lambda^{-1}|\xi|^2}$ (up to a nonzero multiplicative factor), a starting point for T -duality in string theory. More generally, the characteristic function

$$\hat{\mu}(\xi) := \int_H e^{i\langle x, \xi \rangle_H} \mu(dx)$$

of a Gaussian probability measure μ with covariance C on a Hilbert space H is the function $\xi \mapsto e^{-(1/2)\langle \xi, C\xi \rangle_H}$. Such probability measures typically arise in Euclidean quantum field theory; in axiomatic quantum field theory, the analyticity properties of n -point functions can be derived from the Wightman axioms using Fourier transforms. Thus, Fourier transformation underlies many different aspects of quantum field theory.

Fredholm operators

A complex-valued continuous function K on $[0, 1] \times [0, 1]$ gives rise to an integral operator

$$A : f \rightarrow \int_0^1 K(x, y) f(y) dy$$

on complex-valued continuous functions on $[0, 1]$ (equipped with the supremum norm $\|\cdot\|_\infty$) with the following upper bound property:

$$\|Af\|_\infty \leq \text{Sup}_{[0,1] \times [0,1]} |K(x, y)| \|f\|_\infty$$

In other words, A is a bounded linear operator with norm bounded from above by $\sup_{[0,1] \times [0,1]} |K(x, y)|$; a linear operator $A : E \rightarrow F$ from a normed linear space $(E, \|\cdot\|_E)$ to a normed linear space $(F, \|\cdot\|_F)$ is bounded (or continuous) if and only if its (operator) norm $\|A\| := \sup_{\|u\|_E \leq 1} \|Au\|_F$ is bounded.

An integral operator

$$A : f \rightarrow \int_0^1 K(x, y) f(y) dy$$

defined by a continuous kernel K is, moreover, compact; a compact operator is a bounded operator of normed spaces that maps bounded sets to a precompact sets, that is, to sets whose closure is compact. Other examples of compact operators on normed spaces are finite-rank operators, operators with finite-dimensional range. In fact, any compact operator on a separable Hilbert space can be approximated in the topology induced by the operator norm $\|\cdot\|$ by a sequence of finite-rank operators.

Inspired by the work of Volterra, who, in the case of the integral operator defined above, produced

continuous solutions $\phi = (I - A)^{-1}f$ of the equation $f = (I - A)\phi$ for $f \in C([0, 1])$, Fredholm in 1900 (*Sur une classe d'équations fonctionnelles*) studied the equation $f = (I - \lambda A)\phi$, introducing a complex parameter λ . He proved what is since then called the Fredholm alternative, which states that either the equation $f = (I - \lambda A)\phi$ has a unique solution for every $f \in C([0, 1])$ or the corresponding homogeneous equation $(I - \lambda A)\phi = 0$ has nontrivial solutions. In modern language, it means that the resolvent $R(A, \mu) = (A - \mu I)^{-1}$ of a compact linear operator A is surjective if and only if it is injective. The Fredholm alternative is a powerful tool to solve partial differential equations among which the Dirichlet problem, the solutions of which are harmonic functions u (i.e., $\Delta u = 0$, where $\Delta = -\sum_{i=1}^n \partial^2 u / \partial x_i^2$) on some domain $\Omega \in \mathbb{R}^n$ with Dirichlet boundary conditions $u|_{\partial\Omega} = f$, where f is a continuous function on the boundary $\partial\Omega$. The Dirichlet problem has geometric applications, in particular to the nonlinear Plateau problem, which minimizes the area of a surface in \mathbb{R}^d with given boundary curves and which reduces to a (linear) Dirichlet problem.

The operator $B = I - A$ built from the compact operator A is a particular Fredholm operator, namely a bounded linear operator $B : E \rightarrow F$ which is invertible “up to compact operators,” that is, such that there is a bounded linear operator $C : F \rightarrow E$ with both $BC - I_F$ and $CB - I_E$ compact. A Fredholm operator B has a finite-dimensional kernel $\text{Ker } B$ and when $(E, \langle \cdot, \cdot \rangle_E)$ and $(F, \langle \cdot, \cdot \rangle_F)$ are Hilbert spaces its cokernel $\text{Ker } B^*$, where B^* is the adjoint of B defined by

$$\langle B u, v \rangle_F = \langle u, B^* v \rangle_E \quad \forall u \in E, \forall v \in F$$

is also finite dimensional, so that it has a well-defined index $\text{ind}(B) = \dim(\text{Ker } B) - \dim(\text{Ker } B^*)$, a starting point for index theory. Töplitz operators T_ϕ , where ϕ is a continuous function on the unit circle S^1 , provide first examples of Fredholm operators; they act on the Hardy space $\mathcal{H}^2(S^1)$ by

$$T_{e^{-n}} \left(\sum_{m \geq 0} a_m e_m \right) = \sum_{m \geq 0} a_{m+n} e_m$$

under the identification $\mathcal{H}^2(S^1) \simeq l^2(\mathbb{N}) \subset l^2(\mathbb{Z})$, with $l^2(\mathbb{Z})$ equipped with the canonical complete orthonormal basis $(e_n, n \in \mathbb{Z})$. The Fredholm index $\text{ind}(T_{e^{-n}})$ is exactly the integer n so that the index of its adjoint is $-n$, as a consequence of which the index map from Fredholm operators to integers is onto.

One-Parameter (Semi) groups

Unlike in the finite-dimensional situation, a linear operator $A : E \rightarrow F$ between two normed linear spaces $(E, \| \cdot \|_E)$ and $(F, \| \cdot \|_F)$ is not expected to be

bounded. Unbounded operators arise in partial differential equations that involve differential operators such as the Laplacian Δ on an open subset $\Omega \subset \mathbb{R}^n$. The following equations provide fundamental examples of partial differential equations which arose over time from the study of various problems in mathematical physics with the works of Poisson, Fourier, and Cauchy:

$$\begin{aligned} \Delta u &= 0 && \text{Laplace equation} \\ \frac{\partial^2 u}{\partial t^2} + \Delta u &= 0 && \text{wave equation} \\ \frac{\partial u}{\partial t} + \Delta u &= 0 && \text{heat equation} \end{aligned}$$

and later the Schrödinger equation in quantum mechanics:

$$i \frac{\partial u}{\partial t} = \Delta u$$

where t is a time parameter.

An unbounded linear operator on an infinite-dimensional normed space is usually defined on a domain $D(A)$ which is strictly contained in E . The Laplacian Δ is defined on the dense domain $D(\Delta) = H^2(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$; it defines a bounded operator from $H^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$ but does not extend to a bounded operator on $L^2(\mathbb{R}^n)$. Like this operator, most unbounded operators $A : E \rightarrow F$ one comes across have dense domain $D(A)$ in E and are closed, that is, their graph $\{(u, Au), u \in D(A)\}$ is closed as a subset of the normed linear space $E \times F$. When not actually closed, they can be closable, that is, they can have a closed extension called the closure of the operator. By the closed-graph theorem, when E and F are Banach spaces, a linear operator $A : E \rightarrow F$ is continuous whenever its graph is closed, as a consequence of which a closed linear operator $A : E \rightarrow F$ defined on a dense domain is bounded provided its domain coincides with the whole space.

For a closed operator $A : E \rightarrow F$ with dense domain $D(A)$, when E and F are Hilbert spaces equipped with inner products $\langle \cdot, \cdot \rangle_E$ and $\langle \cdot, \cdot \rangle_F$, the adjoint A^* of A is defined on its domain $D(A^*)$ by

$$\langle Au, v \rangle_F = \langle u, A^* v \rangle_E \quad \forall (u, v) \in D(A) \times D(A^*)$$

A self-adjoint operator A with domain $D(A)$ is one for which $D(A) = D(A^*)$ and $A = A^*$; the Laplacian Δ on \mathbb{R}^n is self-adjoint on the Sobolev space $H^2(\mathbb{R}^n)$ but it is only essentially self-adjoint on the dense domain $\mathcal{D}(\mathbb{R}^n)$, the latter meaning that its closure is self-adjoint.

Unbounded self-adjoint operators can arise as generators of one-parameter semigroups of bounded

operators. A one-parameter family of bounded operators $T_t, t \geq 0$ ($T_t, t \in \mathbb{R}$) on a Hilbert space H is a semigroup (resp. group) if $T_s T_t = T_{t+s} \forall t, s \geq 0$ (resp. $\forall t, s \in \mathbb{R}$) and it is strongly continuous (or simply continuous) if $\lim_{t \rightarrow t_0} T_t u = T_{t_0} u$ at any $t_0 \geq 0$ (resp. $t_0 \in \mathbb{R}$) and for any $u \in H$.

Stones' theorem sets up a one-to-one correspondence between continuous one-parameter unitary ($U_t^* U_t = U_t U_t^* = I$) groups $U_t, t \in \mathbb{R}$ on a Hilbert space such that $U_0 = \text{Id}$ and self-adjoint operators A obtained as infinitesimal generators, that is, as the strong limit

$$Au = \lim_{t \rightarrow 0} \frac{U_t u - u}{t}, \quad u \in H$$

of $U_t, t \in \mathbb{R}$, which in a compact form reads $U_t = e^{itA}$. An important example in quantum mechanics is $U_t = e^{itH} U_0, t \in \mathbb{R}$ with H a self-adjoint Hamiltonian, which solves the Schrödinger equation $d/dtu = iHu$. The Lie–Trotter formula, which has important applications for Feynman path integrals, expresses the unitary semigroup generated by $A + B$, where A, B , and $A + B$ are self-adjoint on their respective domains as a strong limit

$$e^{it(A+B)} = \lim_{t \rightarrow \infty} \left(e^{\frac{iA}{n}} e^{\frac{iB}{n}} \right)^n$$

On the other hand, positive operators on a Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ – that is, A self-adjoint and such that $\langle Au, u \rangle_H \geq 0 \quad \forall u \in D(A)$ – generate one-parameter semigroups $T_t = e^{-tA}, t \geq 0$. Hille and Yosida proved that on a Hilbert space, strongly continuous contraction (i.e., $\|T_t\| \leq 1 \quad \forall t > 0$) semigroups such that $T_0 = \text{Id}$ are in one-to-one correspondence with densely defined positive operators $A : D(A) \subset H \rightarrow H$ that are maximal (i.e., $I + A$ is onto), obtained as (minus the) infinitesimal generators

$$-Au = \lim_{t \rightarrow 0} \frac{T_t u - u}{t}, \quad u \in H$$

of the corresponding semigroups. Similarly, a positive densely defined self-adjoint operator A on a Hilbert space H gives rise to a densely defined closed symmetric sesquilinear form $(u, v) \mapsto \langle \sqrt{A}u, \sqrt{A}v \rangle_H$ (see next section for a definition of $\sqrt{A}; \langle \cdot, \cdot \rangle_H$ is the scalar product on H) and this map yields a one-to-one correspondence between operators and sesquilinear forms on H with the aforementioned properties, one of the starting points for the theory of Dirichlet forms. To a probability measure μ on a separable Banach space E , one can associate a densely defined closed symmetric sesquilinear form (it is in fact a Dirichlet form) on a Hilbert space H

such that $E^* \subset H^* = H \subset E$, which in the particular case of the standard Wiener measure μ on the Wiener space $E = C([0, t])$ and with Hilbert space given by the Cameron–Martin space $H = H^1([0, t])$, is the bilinear form

$$(u, v) \mapsto \int \langle \bar{\nabla} u, \bar{\nabla} v \rangle_H$$

with $\bar{\nabla}$ the (closed) gradient of Malliavin calculus.

The operator $-\Delta$, where Δ is the Laplacian on \mathbb{R}^n , generates the heat-operator semigroup $e^{-\Delta t}, t \geq 0$. It has a smooth kernel $K_t \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n)$ defined by

$$(e^{-\Delta t} f)(x) = \int_{\mathbb{R}^n} K_t(x, y) f(y) dy \quad \forall f \in C_0^\infty(\mathbb{R}^n)$$

and defines a smoothing operator, an operator that maps Sobolev function to smooth function. In general, a pseudodifferential operators A on an open subset U of \mathbb{R}^n with symbol σ_A only has a distribution kernel

$$K_A(x, y) = \int_{\mathbb{R}^n} e^{i(x-y, \xi)} \sigma(\xi) d\xi$$

The kernel of the inverse Laplacian $(\Delta + m^2)^{-1}$ on \mathbb{R}^n (the non-negative real number m^2 stands for the mass) called Green's function on \mathbb{R}^n , plays an essential role in the theory of Feynman graphs.

Spectral Theory

Spectral theory is the study of the distribution of the values of the complex parameter λ for which, given a linear operator A on a normed space E , the operator $A - \lambda I$ has an inverse and of the properties of this inverse when it exists, the resolvent $R(A, \lambda) = (A - \lambda I)^{-1}$ of A . The resolvent $\rho(A)$ of A is the set of complex numbers λ for which $A - \lambda I$ is invertible with densely defined bounded inverse. The spectrum $\text{Sp}(A)$ of A is the complement in \mathbb{C} of the resolvent; it consists of a union of three disjoint sets: the set of all complex numbers λ for which $A - \lambda I$ is not injective, called the point spectrum – such a λ is an eigenvalue of A with associated eigenfunction any $u \in D(A)$ such that $Au = \lambda u$; the set of points λ for which $A - \lambda I$ has a densely defined unbounded inverse $R(A, \lambda)$ called the continuous spectrum; and the set of points λ for which $A - \lambda I$ has a well-defined unbounded but not densely defined inverse $R(A, \lambda)$ called the residual spectrum.

A bounded operator has bounded spectrum and a self-adjoint operator A acting on a Hilbert space has real spectrum and no residual spectrum since the range of $A - \lambda I$ is dense. As a consequence of the

Fredholm alternative, the spectrum of a compact operator consists only of point spectrum; it is countable with accumulation point at 0. A Hamiltonian of a quantum mechanical system can have both point and continuous spectra, but its point spectrum is of special interest because the corresponding eigenfunctions are stationary states of the system. As was first pointed out by Kac (“Can you hear the shape of a drum?”), the spectrum of an operator acting on functions can reflect the geometry of the space these functions are defined on, a starting point for many interesting and far-reaching questions in differential geometry.

A self-adjoint linear operator on a Hilbert space can be described in terms of a family of projections E_λ , $\lambda \in \mathbb{R}$ via the spectral representation

$$A = \int_{\text{Sp}(A)} \lambda dE_\lambda$$

Given a Borel real-valued function f on \mathbb{R} , the operator

$$f(A) = \int_{\text{Sp}(A)} f(\lambda) dE_\lambda$$

yields another self-adjoint operator. A positive operator A on a dense domain $D(A)$ of some Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ has non-negative spectrum and for any positive real number t , the map $\lambda \mapsto e^{-t\lambda}$ gives the associated bounded heat-operator

$$e^{-tA} = \int_{\text{Sp}(A)} e^{-t\lambda} dE_\lambda$$

while the map $\lambda \mapsto \sqrt{\lambda}$ gives rise to a positive operator \sqrt{A} such that $\sqrt{A}^2 = A$.

The resolvent can also be used to define new operators

$$f(A) = \frac{1}{2i\pi} \int_C f(\lambda) R(A, \lambda) d\lambda$$

from a linear operator via a Cauchy-type integral along a contour C around the spectrum; this way one defines complex powers A^{-z} of (essentially self-adjoint) positive elliptic pseudodifferential operators which enter the definition of the zeta-function, $z \mapsto \zeta(A, z)$, of the operator A . The ζ -function is a useful tool to extend the ordinary determinant to ζ -determinants of self-adjoint elliptic operators, thereby providing an ansatz to give a meaning to partition functions in the path integral approach to quantum field theory.

Operator Algebras

Bounded linear operators on a Hilbert space H form an algebra $\mathcal{L}(H)$ closed for the operator norm

with involution given by the adjoint operation $A \mapsto A^*$; it is a C^* -algebra, that is, an algebra over \mathbb{C} with a norm $\|\cdot\|$ and an involution $*$ such that A is closed for this norm and such that $\|ab\| \leq \|a\|\|b\|$ and $\|a^*a\| = \|a\|^2$ for all $a, b \in A$ and by the Gelfand–Naimark theorem, every C^* -algebra is isomorphic to a sub- C^* -algebra of some $\mathcal{L}(H)$. The notion of spectrum extends from bounded operators to C^* -algebras; the spectrum $\text{sp}(a)$ of an element a in a C^* -algebra A is a (compact) set of complex numbers such that $a - \lambda \cdot 1$ is not invertible. The notion of self-adjointness also extends ($a = a^*$), and just as a self-adjoint operator $B \in \mathcal{L}(H)$ is non-negative (in which case its spectrum lies in \mathbb{R}^+) if and only if $B = A^*A$ for some bounded operator A , an element $b \in A$ is said to be non-negative if and only if $b = a^*a$ for some $a \in A$, in which case $\text{sp}(a) \subset \mathbb{R}_0^+$.

The algebra $C(X)$ of continuous functions $f : X \rightarrow \mathbb{C}$ vanishing at infinity on some locally compact Hausdorff space X equipped with the supremum norm and the conjugation $f \mapsto \bar{f}$ is also a C^* -algebra and a prototype for abelian C^* -algebras, since Gelfand showed that every abelian C^* -algebra is isometrically isomorphic to $C(X)$, with X compact if the algebra is unital. To a C^* -algebra A , one can associate an abelian group $K_0(A)$ which is dual to the Grothendieck group $K^0(X)$ of isomorphism classes of vector bundles over a compact Hausdorff space X .

Compact operators on a Hilbert space H form the only proper two-sided ideal $\mathcal{K}(H)$ of the C^* -algebra $\mathcal{L}(H)$ which is closed for the operator norm topology on $\mathcal{L}(H)$. The quotient $\mathcal{L}(H)/\mathcal{K}(H)$ is called the Calkin space, after Calkin, who classified all two-sided ideals in $\mathcal{L}(H)$ for a separable Hilbert space H ; one can set up a one-to-one correspondence between such ideals and certain sequence spaces. Corresponding to the Banach space $l^1(\mathbb{Z})$ of complex-valued sequences (u_n) such that $\sum_{n \in \mathbb{N}} |u_n| < \infty$, is the $*$ -ideal $\mathcal{I}_1(H)$ of trace-class operators. The trace $\text{tr}(A) = \sum_{n \in \mathbb{Z}} \langle A e_n, e_n \rangle_H$ of a negative operator $A \in \mathcal{L}(H)$ lies in $[0, +\infty)$ and is independent of the choice of the complete orthonormal basis $\{e_n, n \in \mathbb{Z}\}$ of H equipped with the inner product $\langle \cdot, \cdot \rangle_H$. $\mathcal{I}_1(H)$ is the Banach space of bounded linear operators on H such that $\|A\|_1 = \text{tr}(|A|)$ is bounded. Given an (essentially self-adjoint) positive differential operator D of order d acting on smooth functions on a closed n -dimensional Riemannian manifold M , its complex power D^{-z} is a trace class on the space of L^2 -functions on M provided $\text{Re}(z) > n/d$ and the corresponding trace $\text{tr}(D^{-z})$ extends to a meromorphic function on the whole plane, the ζ -function $\zeta(D, z)$ which is holomorphic at 0.

More generally, Banach spaces $l^p(\mathbb{Z})$, $1 \leq p < \infty$, of complex-valued sequences $(u_n)_{n \in \mathbb{Z}}$ such that $\sum_{n \in \mathbb{Z}} |u_n|^p < \infty$ relate to Schatten ideals $\mathcal{I}_p(H)$, $1 \leq p < \infty$, where $\mathcal{I}_p(H)$ is the Banach space of bounded linear operators on H such that $\|A\|_p = (\text{tr}(|A|^p))^{1/p}$ is bounded. Just as all l^p -sequences converge to 0, the Schatten ideals $\mathcal{I}_p(H)$ all lie in $\mathcal{K}(H)$ and we have $\cdots \subset \mathcal{I}_{p+1}(H) \subset \mathcal{I}_p(H) \subset \cdots \subset \mathcal{K}(H)$.

Compact operators and Schatten ideals are useful to extend index theory to a noncommutative context; a Fredholm module (H, F) over an involutive algebra A is given by an involutive representation π of A in a Hilbert space H and a self-adjoint bounded linear operator F on H such that $F^2 = \text{Id}_H$ and the operator brackets $[F, \pi(a)]$ are compact for all $a \in A$. To a p -summable Fredholm module (H, F) , that is, $[F, \pi(a)] \in \mathcal{I}_p(H)$ for all $a \in A$, one associates a representative τ of the Chern character $\text{ch}^*(H, F)$ given by a cyclic cocycle on A , which pairs up with K -theory to build an integer-valued index map τ on K -theory.

Schatten ideals are also useful to investigate the geometry of infinite-dimensional spaces such as loop groups, for which the Hilbert–Schmidt operators (operators in $\mathcal{I}_2(H)$) are also called Hilbert–Schmidt

operators) are particularly useful. A Hölder-type inequality shows that the product of two Hilbert–Schmidt operators is trace-class. Moreover, for any two Hilbert–Schmidt operators A and B , the “cyclicity property” that $\text{tr}(AB) = \text{tr}(BA)$ holds, and the sesquilinear form $(A, B) \mapsto \text{tr}(AB^*)$ makes $\mathcal{L}_2(H)$ a Hilbert space.

Further Reading

- Adams R (1975) *Sobolev Spaces*. London: Academic Press.
 Dunford N and Schwartz J (1971) *Linear Operators*. Part I. General Theory. Part II. Spectral Theory. Part III. Spectral Operators. New York: Wiley.
 Hille E (1972) *Methods in Classical and Functional Analysis*. London: Academic Press and Addison-Wesley.
 Kato T (1982) *A Short Introduction to Perturbation Theory for Linear Operators*. New York–Berlin: Springer.
 Reed M and Simon B (1980) *Methods of Modern Mathematical Physics* vols. I–IV, 2nd edn. New York: Academic Press.
 Riesz F and SZ-Nagy B (1968) *Leçons d'analyse fonctionnelle*. Paris: Gauthier–Villars; Budapest Akademiai Kiado.
 Rudin W (1994) *Functional Analysis*, 2nd edn. New York: International Series in Pure and Applied Mathematics.
 Yosida K (1980) *Functional Analysis*, 6th edn. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen Band vol. 132. Berlin–New York: Springer.

Introductory Article: Minkowski Spacetime and Special Relativity

G L Naber, Drexel University, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Minkowski spacetime is generally regarded as the appropriate mathematical context within which to formulate those laws of physics that do not refer specifically to gravitational phenomena. Here we shall describe this context in rigorous terms, postulate what experience has shown to be its correct physical interpretation, and illustrate by means of examples its appropriateness for the formulation of physical laws.

Minkowski Spacetime and the Lorentz Group

Minkowski spacetime \mathcal{M} is a four-dimensional real vector space on which is defined a bilinear form $\mathbf{g} : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ that is symmetric ($\mathbf{g}(v, w) = \mathbf{g}(w, v)$ for all $v, w \in \mathcal{M}$) and nondegenerate ($\mathbf{g}(v, v) = 0$

for all $w \in \mathcal{M}$ implies $v = 0$). Further, \mathbf{g} has index 1, that is, there exists a basis $\{e_1, e_2, e_3, e_4\}$ for \mathcal{M} with

$$\mathbf{g}(e_a, e_b) = \eta_{ab} = \begin{cases} 1 & \text{if } a = b = 1, 2, 3 \\ -1 & \text{if } a = b = 4 \\ 0 & \text{if } a \neq b \end{cases}$$

\mathbf{g} is called a Lorentz inner product for \mathcal{M} and any basis of the type just described is an orthonormal basis for \mathcal{M} . We shall often write $v \cdot w$ for the value $\mathbf{g}(v, w)$ of \mathbf{g} on $(v, w) \in \mathcal{M} \times \mathcal{M}$. A vector $v \in \mathcal{M}$ is said to be spacelike, timelike, or null if $v \cdot v$ is positive, negative, or zero, respectively, and the set C_N of all null vectors is called the null cone in \mathcal{M} . If $\{e_1, e_2, e_3, e_4\}$ is an orthonormal basis and if we write $v = v^1 e_1 + v^2 e_2 + v^3 e_3 + v^4 e_4 = v^a e_a$ (using the Einstein summation convention, according to which a repeated index, one subscript and one superscript, is summed over its possible values) and $w = w^b e_b$, then

$$\begin{aligned} v \cdot w &= v^1 w^1 + v^2 w^2 + v^3 w^3 - v^4 w^4 \\ &= \eta_{ab} v^a w^b \end{aligned}$$

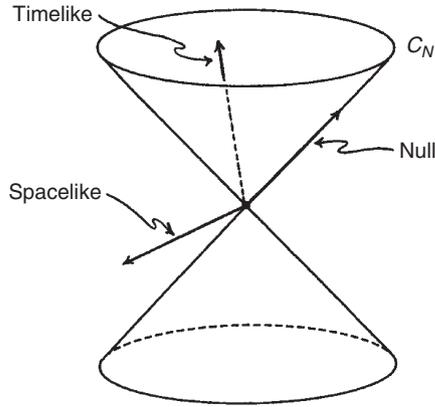


Figure 1 Spacelike, timelike and null vectors.

In particular, v is null if and only if

$$(v^4)^2 = (v^1)^2 + (v^2)^2 + (v^3)^2$$

(hence the name null “cone” for C_N). Timelike vectors are “inside” the null cone and spacelike vectors are “outside” (see [Figure 1](#)).

We select some orientation for the vector space \mathcal{M} and will henceforth consider only oriented, orthonormal bases for \mathcal{M} . From the Schwartz inequality for \mathbb{R}^3 , one can show ([Naber 1992](#), theorem 1.3.1) that, if v is timelike and w is either timelike or null and nonzero, then $v \cdot w < 0$ if and only if $v^4 w^4 > 0$ in any orthonormal basis. In particular, one can define an equivalence relation on the set of all timelike vectors by decreeing that two such, v and w , are equivalent if and only if $v \cdot w < 0$. For reasons that will emerge shortly we then say that v and w have the same time orientation. There are precisely two equivalence classes, one of which we select and designate future directed. Timelike vectors in the other class are then called past directed. One can show ([Naber 1992](#), section 1.3 and corollary 1.4.5) that this classification can be extended to nonzero null vectors as well (but not to spacelike vectors). We will call an oriented, orthonormal basis time oriented if its timelike vector e_4 is future directed and will consider only these in what follows. An oriented, time-oriented, orthonormal basis for \mathcal{M} will be called an admissible basis. If $\{e_1, e_2, e_3, e_4\}$ and $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ are two such bases and if we write

$$\begin{aligned} e_b &= \Lambda^1_b \hat{e}_1 + \Lambda^2_b \hat{e}_2 + \Lambda^3_b \hat{e}_3 + \Lambda^4_b \hat{e}_4 \\ &= \Lambda^a_b \hat{e}_a, \quad b = 1, 2, 3, 4 \end{aligned} \quad [1]$$

then the matrix $\Lambda = (\Lambda^a_b)$ (a =row index, b =column index) can be shown to satisfy the following three conditions ([Naber 1992](#), section 1.3):

- (orthogonality) $\Lambda^T \eta \Lambda = \eta$,
where T means transpose and

$$\eta = (\eta_{ab}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

- (orientability) $\det \Lambda = 1$, and
- (time orientability) $\Lambda^4_4 \geq 1$.

We shall refer to any 4×4 matrix $\Lambda = (\Lambda^a_b)$ satisfying these three conditions as a Lorentz transformation (although one often sees the adjectives “proper” and “orthochronous” appended to emphasize conditions (2) and (3), respectively). The set \mathcal{L} of all such matrices forms a group under matrix multiplication that we call simply the Lorentz group. It is a simple matter to show ([Naber 1992](#), lemma 1.3.4) from the orthogonality condition (1) that, if $\Lambda^4_4 = 1$, then Λ must be of the form

$$\begin{pmatrix} & & & 0 \\ & (R^i_j) & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where (R^i_j) is an element of $SO(3)$, that is, a 3×3 orthogonal matrix with determinant 1. The set \mathcal{R} of all matrices of this form is a subgroup of \mathcal{L} called the rotation subgroup. Although it will play no role in what we do here, it should be pointed out that in many applications (e.g., in particle physics) it is necessary to consider the larger group of transformations of \mathcal{M} generated by the Lorentz group and spacetime translations ($x^a \rightarrow x^a + \Lambda^a$, for some constants $\Lambda^a, a=1, 2, 3, 4$). This is called the inhomogeneous Lorentz group, or Poincaré group.

Physical Interpretation

For the purpose of describing how one is to think of Minkowski spacetime and the Lorentz group physically it will be convenient to distinguish (intuitively and terminologically, if not mathematically) between a “vector” in \mathcal{M} and a “point” in \mathcal{M} (the “tip” of a vector). The points in \mathcal{M} are called events and are to be thought of as actual physical occurrences, albeit idealized as “point events” which have no spatial extension and no duration. One might picture, for example, an instantaneous collision, or explosion, or an “instant” in the history of some point material particle or photon (“particle of light”).

Events are observed and identified by the assignment of coordinates. We will be interested in coordinates assigned in a very particular way by a

very particular type of observer. Specifically, our admissible observers preside over three-dimensional, right-handed, Cartesian spatial coordinate systems, relative to which photons always move along straight lines in any direction. With a single clock located at the origin, such an observer can determine the speed, c , of light *in vacuo* by the so-called Fizeau procedure (emit a photon from the origin when the clock there reads t_1 , bounce it back from a mirror located at (x^1, x^2, x^3) , receive the photon at the origin again when the clock there reads t_2 and set $c = 2\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}/(t_2 - t_1)$). Now place an identical clock at each spatial point and synchronize them by emitting from the origin a spherical electromagnetic wave (photons in all directions) and setting the clock whose location is (x^1, x^2, x^3) to read $\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}/c$ at the instant the wave arrives. An observer now assigns to an event the three spatial coordinates of the location at which it occurred in his coordinate system as well as the time reading on the clock at that location at the instant the event occurred. We shall assume also that our admissible observers are inertial in the sense of Newtonian mechanics (the trajectory of a particle on which no forces act, when described in terms of the coordinates just introduced, is a point or a straight line traversed at constant speed). It is an experimental fact (and quite a remarkable one) that all of these admissible observers (whether or not they are in relative motion) agree on the numerical value of the speed of light *in vacuo* ($c \approx 3.00 \times 10^{10} \text{ cm s}^{-1}$). We shall exploit this fact at the outset to have all of our admissible observers measure time in units of distance by simply multiplying their time coordinates t by c . The resulting time coordinate is denoted $x^4 = ct$. In these units all speeds are dimensionless and the speed of light *in vacuo* is 1.

In our mathematical model \mathcal{M} of the world of events, this very subtle and complex notion of an admissible observer is fully identified with the conceptually very simple notion of an admissible basis $\{e_1, e_2, e_3, e_4\}$. If $x \in \mathcal{M}$ is an event and if we write $x = x^a e_a$, then (x^1, x^2, x^3) are the spatial and x^4 is the time coordinate supplied for x by the corresponding observer. If $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ is another basis/observer related to $\{e_1, e_2, e_3, e_4\}$ by [1] and if we write $x = \hat{x}^a \hat{e}_a$, then

$$\hat{x}^a = \Lambda^a_b x^b, \quad a = 1, 2, 3, 4 \quad [2]$$

Thus, Lorentz transformations relate the space and time coordinates supplied for any given event by two admissible observers. If $(\Lambda^a_b) \in \mathcal{R}$, then the two observers differ only in the orientation of their spatial

coordinate axes. On the other hand, for any real number θ one can define an element $L(\theta)$ of \mathcal{L} by

$$L(\theta) = \begin{pmatrix} \cosh \theta & 0 & 0 & -\sinh \theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh \theta & 0 & 0 & \cosh \theta \end{pmatrix} \quad [3]$$

and, if two admissible bases are related by this Lorentz transformation, then the coordinate transformation [2] becomes

$$\begin{aligned} \hat{x}^1 &= (\cosh \theta) x^1 - (\sinh \theta) x^4 \\ \hat{x}^2 &= x^2 \\ \hat{x}^3 &= x^3 \\ \hat{x}^4 &= -(\sinh \theta) x^1 + (\cosh \theta) x^4 \end{aligned} \quad [4]$$

Letting $\beta = \tanh \theta$ (so that $-1 < \beta < 1$) and suppressing $\hat{x}^2 = x^2$ and $\hat{x}^3 = x^3$, one obtains

$$\begin{aligned} \hat{x}^1 &= \frac{1}{\sqrt{1-\beta^2}} x^1 - \frac{\beta}{\sqrt{1-\beta^2}} x^4 \\ \hat{x}^4 &= -\frac{\beta}{\sqrt{1-\beta^2}} x^1 + \frac{1}{\sqrt{1-\beta^2}} x^4 \end{aligned} \quad [5]$$

This corresponds to two observers whose spatial axes are oriented as shown in Figure 2 with the hatted coordinate system moving along the common x^1 -, \hat{x}^1 -axis with speed $|\beta|$, to the right if $\beta > 0$ and to the left if $\beta < 0$.

We remark that, reverting to traditional time units, $\beta = v/c$, where $|v|$ is the relative speed of the two coordinate systems, and [5] becomes what is generally referred to as a ‘‘Lorentz transformation’’ in elementary expositions of special relativity, that is,

$$\begin{aligned} \hat{x}^1 &= \frac{x^1 - vt}{\sqrt{1 - v^2/c^2}} \\ \hat{t} &= \frac{t - (v/c^2)x^1}{\sqrt{1 - v^2/c^2}} \end{aligned} \quad [6]$$

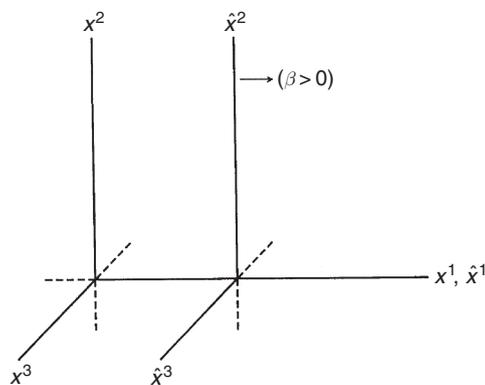


Figure 2 Observers in standard configuration.

There is a sense in which, to understand the kinematic effects of special relativity, it is enough to restrict one's attention to the so-called special Lorentz transformations $L(\theta)$. Specifically, one can show (Naber 1992, theorem 1.3.5) that if $\Lambda \in \mathcal{L}$ is any Lorentz transformation, then there exists a real number θ and two rotations $R_1, R_2 \in \mathcal{R}$ such that $\Lambda = R_1 L(\theta) R_2$. Since R_1 and R_2 involve no relative motion, all of the kinematics is contained in $L(\theta)$. We shall explore these kinematic effects in more detail shortly.

Now suppose that x and x_0 are two distinct events in \mathcal{M} and consider the displacement vector $x - x_0$ from x_0 to x . If $\{e_1, e_2, e_3, e_4\}$ is an admissible basis and if we write $x = x^a e_a$ and $x_0 = x_0^a e_a$, then $x - x_0 = (x^a - x_0^a) e_a = \Delta x^a e_a$. If $x - x_0$ is null, then

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 = (\Delta x^4)^2$$

so the spatial separation of the two events is equal to the distance light would travel during the time lapse between the events. The same must be true in any other admissible basis since Lorentz transformations are the matrices of linear maps that preserve the Lorentz inner product. Consequently, all admissible observers agree that x_0 and x are “connectible by a photon.” They even agree as to which of the two events is to be regarded as the “emission” of the photon and which is to be regarded as its “reception” since one can show (Naber 1992, theorem 1.3.3) that, when a vector is either timelike or null and nonzero, the sign of its fourth coordinate is the same in every admissible basis (because $\Lambda^4_4 \geq 1$). Thus, $x^4 - x_0^4$ is either positive for all admissible observers (x_0 occurred before x) or negative for all admissible observers (x_0 occurred after x). Since photons move along straight lines in admissible coordinate systems we adopt the following terminology. If $x_0, x \in \mathcal{M}$ are such that $x - x_0$ is null, then the straight line in \mathcal{M} containing x_0 and x is called the world line of a photon in \mathcal{M} and is to be thought of as the set of all events in the history of some particle of light that “experiences” both x_0 and x .

Let us now suppose instead that $x - x_0$ is timelike. Then, in any admissible basis,

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 < (\Delta x^4)^2$$

so the spatial separation of x_0 and x is less than the distance light would travel during the time lapse between the events. In this case, one can prove (Naber 1992, section 1.4) that there exists an admissible basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ in which $\Delta \hat{x}^1 = \Delta \hat{x}^2 = \Delta \hat{x}^3 = 0$, that is, there is an admissible observer for whom the two events occur at the same spatial location, one after the other. Thinking of this location as occupied by some

material object (e.g., the observer's clock situated at that point) we find that the events x_0 and x are both “experienced” by this material particle and that, moreover, $\sqrt{|\mathbf{g}(x - x_0, x - x_0)|}$ is just the time lapse between the events recorded by a clock carried along by this material particle. To any other admissible observer this material particle appears “free” (not subject to forces) because it moves on a straight line with constant speed. This leads us to the following definitions. If $x_0, x \in \mathcal{M}$ are such that $x - x_0$ is timelike, then the straight line in \mathcal{M} containing x_0 and x is called the world line of a free material particle in \mathcal{M} and $\sqrt{|\mathbf{g}(x - x_0, x - x_0)|}$, usually written $\tau(x - x_0)$, or simply $\Delta\tau$, is the proper time separation of x_0 and x . One can think of $\tau(x - x_0)$ as a sort of “length” for $x - x_0$ measured, however, by a clock carried along by a free material particle that experiences both x_0 and x . It is an odd sort of length, however, since it satisfies not the usual triangle inequality, but the following “reversed” version.

Reversed triangle inequality (Naber 1992, theorem 1.4.2) *Let x_0, x and y be events in \mathcal{M} for which $y - x_0$ and $x - x_0$ are timelike with the same time orientation. Then $y - x_0 = (y - x) + (x - x_0)$ is timelike and*

$$\tau(y - x_0) \geq \tau(y - x) + \tau(x - x_0) \quad [7]$$

with equality holding if and only if $y - x$ and $x - x_0$ are linearly dependent.

The sense of the inequality in [7] has interesting consequences about which we will have more to say shortly.

Finally, let us suppose that $x - x_0$ is spacelike. Then, in any admissible basis

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 > (\Delta x^4)^2$$

so the spatial separation of x_0 and x is greater than the distance light could travel during the time lapse that separates them. There is clearly no admissible observer for whom the events occur at the same location. No free material particle (or even photon) can experience both x_0 and x . However, one can show (Naber 1992, section 1.5) that, given any real number T (positive, negative, or zero), one can find an admissible basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ in which $\Delta \hat{x}^4 = T$. Some admissible observers will judge the events simultaneous, some will assert that x_0 occurred before x , and others will reverse the order. Temporal order, cause and effect, have no meaning for such pairs of events. For those admissible observers for whom the events are simultaneous ($\Delta \hat{x}^4 = 0$), the quantity $\sqrt{\mathbf{g}(x - x_0, x - x_0)}$ is the distance between them and for this reason this quantity is called the proper spatial separation of x_0 and x (whenever $x - x_0$ is spacelike).

For any two events $x_0, x \in \mathcal{M}$, $g(x - x_0, x - x_0)$ is given in any admissible basis by $(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 - (\Delta x^4)^2$ and is called the interval separating x_0 and x . It is the closest analog in Minkowskian geometry to the (squared) length in Euclidean geometry. It can, however, assume any real value depending on the physical relationship between the events x_0 and x . Historically, of course, it was the various physical interpretations of this interval that we have just described which led Minkowski (Einstein *et al.* 1958) to the introduction of the structure that bears his name.

Kinematic Effects

All of the well-known kinematic effects of special relativity (the addition of velocities formula, the relativity of simultaneity, time dilation, and length contraction) follow easily from what we have done. Because it eases visualization and because, as we mentioned earlier, it suffices to do so, we will limit our discussion to the special Lorentz transformations.

Let θ_1 and θ_2 be two real numbers and consider the corresponding elements $L(\theta_1)$ and $L(\theta_2)$ of \mathcal{L} defined by [3]. Sum formulas for $\sinh \theta$ and $\cosh \theta$ imply that $L(\theta_1)L(\theta_2) = L(\theta_1 + \theta_2)$. Defining $\beta_i = \tanh \theta_i, i = 1, 2$, and $\beta = \tanh(\theta_1 + \theta_2)$, the sum formula for $\tanh \theta$ then gives

$$\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2} \tag{8}$$

The physical interpretation is simple. One has three admissible observers whose spatial axes are related in the manner shown in Figure 2. If the speed of the second relative to the first is β_1 and the speed of the third relative to the second is β_2 , then the speed of the third relative to the first is not $\beta_1 + \beta_2$ as a Newtonian predisposition would lead one to expect, but rather β , given by [8]. This is the relativistic addition of velocities formula.

We have seen already that, when the interval between x_0 and x is spacelike, the events will be judged simultaneous by some admissible observers, but not by others. Indeed, if $\Delta x^4 = 0$ and the observers are related by [5], then $\Delta \hat{x}^4 = -(\beta/\sqrt{1 - \beta^2})\Delta x^1 = -\beta\Delta \hat{x}^1$, which will not be zero unless $\beta = 0$ and so there is no relative motion ($\Delta \hat{x}^1$ cannot be zero since then $\Delta \hat{x}^a = 0$ for $a = 1, 2, 3, 4$ and $x = x_0$). This phenomenon is called the relativity of simultaneity and we now construct a simple geometrical representation of it.

Select two perpendicular lines in the plane to represent the x^1 - and x^4 -axes (the Euclidean orthogonality of the lines has no physical significance and

is unnecessary, but makes the pictures easier to draw). The \hat{x}^1 -axis will be represented by the straight line $\hat{x}^4 = 0$ which, from [5], is given by $x^4 = \beta x^1$ (in Figure 3 we have assumed that $\beta > 0$). Similarly, the \hat{x}^4 -axis is identified with the line $x^4 = (1/\beta)x^1$. Since Lorentz transformations leave the Lorentz inner product invariant, the hyperbolas $(x^1)^2 - (x^4)^2 = k$ coincide with $(\hat{x}^1)^2 - (\hat{x}^4)^2 = k$ and we calibrate the axes accordingly, for example, the branch of $(x^1)^2 - (x^4)^2 = 1$ with $x^1 > 0$ intersects the x^1 -axis at the point $(x^1, x^4) = (1, 0)$ and intersects the \hat{x}^1 -axis at the point $(\hat{x}^1, \hat{x}^4) = (1, 0)$. This necessitates a different scale on the hatted and unhatted axes, but one can show (Naber 1992, section 1.3) that, with this calibration, all coordinates can be obtained geometrically by projecting parallel to the opposite axis (e.g., the x^4 - and \hat{x}^4 -coordinates of an event result from projecting parallel to the x^1 - and \hat{x}^1 -axes, respectively).

Thus, a line of simultaneity in the hatted (respectively, unhatted) coordinates is parallel to the \hat{x}^1 - (respectively, x^1 -) axis so that, in general, a pair of events lying on one will not lie on the other (note, however, that these lines are “really” three-dimensional hyperplanes so what appears to be a point of intersection is actually a two-dimensional “plane of agreement”, any two events in which are judged simultaneous by both observers).

For any two events whatsoever the relationship between the time lapse $\Delta \hat{x}^4$ in the hatted coordinates and the time lapse Δx^4 in the unhatted coordinates is, from [5],

$$\Delta \hat{x}^4 = -\frac{\beta}{\sqrt{1 - \beta^2}} \Delta x^1 + \frac{1}{\sqrt{1 - \beta^2}} \Delta x^4$$

so the two are generally not equal. Consider, in particular, two events on the world line of a point at rest in the unhatted coordinate system, for

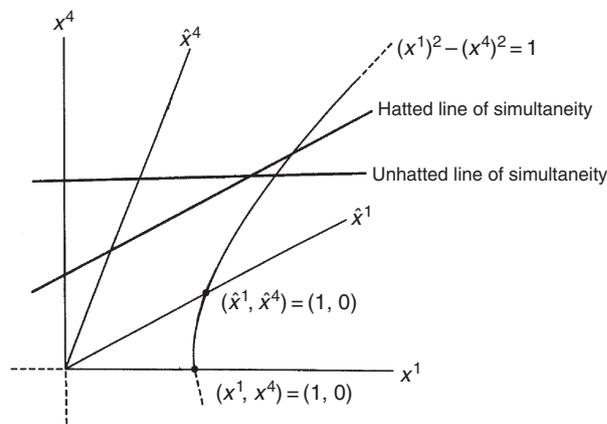


Figure 3 Relativity of simultaneity.

example, two readings on the clock at rest at the origin in this system. Then $\Delta x^1 = 0$ so

$$\Delta \hat{x}^4 = \frac{1}{\sqrt{1 - \beta^2}} \Delta x^4 > \Delta x^4$$

This effect is entirely symmetrical since, if $\Delta \hat{x}^1 = 0$, then [5] implies

$$\Delta x^4 = \frac{1}{\sqrt{1 - \beta^2}} \Delta \hat{x}^4 > \Delta \hat{x}^4$$

Each observer judges the other's clocks to be running slow. This phenomenon is called time dilation and is clearly visible in the spacetime diagram in **Figure 4** (e.g., both observers agree on the time reading “0” for the clock at the origin of the unhatted system, but the line $\hat{x}^4 = 1$ intersects the world line of the clock, i.e., the x^4 -axis, at a point below $(x^1, x^4) = (0, 1)$).

We should emphasize that this phenomenon is quite “real” in the physical sense. For example, certain types of elementary particles (mesons) found in cosmic radiation are so short-lived (at rest) that, even if they could travel at the speed of light, the time required to traverse our atmosphere would be some ten times their normal life span. They should not be able to reach the earth, but they do. Time dilation “keeps them young” in the sense that what seems a normal life time to the meson appears much longer to us.

Finally, since admissible observers generally disagree on which events are simultaneous and since the only way to measure the “length” of a moving object (say, a measuring rod) is to locate its end points “simultaneously,” it should come as no surprise that length, like simultaneity, and time, depends on the admissible observer measuring it. Specifically, let us consider a measuring rod lying at rest along the \hat{x}^1 -axis of the hatted coordinate

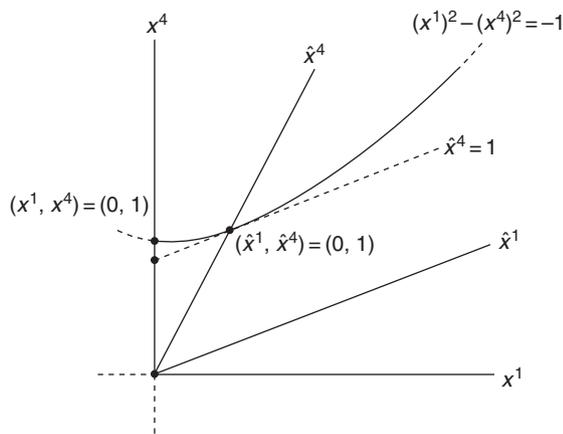


Figure 4 Time dilation.

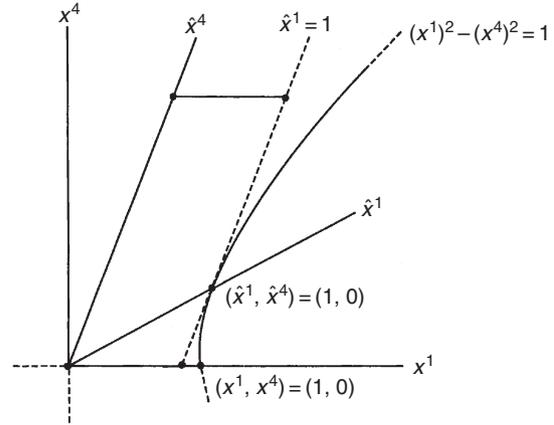


Figure 5 Length contraction.

system. Its “length” in this coordinate system is $\Delta \hat{x}^1$. The world lines of its end points are two straight lines parallel to the \hat{x}^4 -axis. If the unhatted observer locates two events on these world lines “simultaneously” their coordinates will satisfy $\Delta x^4 = 0$ and, by [5] $\Delta \hat{x}^1 = (1/\sqrt{1 - \beta^2}) \Delta x^1$ so

$$\Delta x^1 = \sqrt{1 - \beta^2} \Delta \hat{x}^1 < \Delta \hat{x}^1$$

and the moving measuring rod appears contracted in its direction of motion by a factor of $\sqrt{1 - \beta^2}$. As for time dilation, this phenomenon, known as length contraction, is entirely symmetrical, quite real, and clearly visible in a spacetime diagram (**Figure 5**).

The Relativity Principle

We have found that admissible observers can disagree about some rather startling things (whether or not two events are simultaneous, the time lapse between two events even when no one thinks they are simultaneous, and the length of a measuring rod). This would be a matter of no concern at all, of course, if one could determine, in any given situation, who was really right. Surely, two events are either simultaneous or they are not and we need only sort out which admissible observer has the correct view of the situation? Unfortunately (or fortunately, depending on one's point of view) this distinction between the judgments made by different admissible observers is precisely what physics forbids.

The relativity principle (Einstein *et al.* 1958). *All admissible observers are completely equivalent for the formulation of the laws of physics.*

We must be clear that this is not a mathematical statement. It is rather a statement about the physical world around us and how it should be described, gleaned from observations, some of which are

complex and subtle and some of which are commonplace (a passenger in a smooth, quiet airplane traveling at constant groundspeed cannot “feel” his motion relative to the earth). It is a powerful guide for constructing the laws of relativistic physics, but even more fundamentally it prohibits us from regarding any particular admissible observer as having a privileged view of the universe. In particular, we are forbidden from attaching any objective significance to such questions as, “were the two supernovae simultaneous?”, “How long did the meson survive?”, and “What is the distance between the Crab Nebula and Alpha Centauri?” This is severe, but one must deal with it.

Particles and 4-Momentum

If $I \subseteq \mathbb{R}$ is an interval, then a map $\alpha: I \rightarrow \mathcal{M}$ is a curve in \mathcal{M} . Relative to any admissible basis we can write

$$\alpha(\xi) = x^a(\xi) e_a$$

for each $\xi \in I$. We shall assume that α is smooth in the sense that each $x^a(\xi)$, $a = 1, 2, 3, 4$, is infinitely differentiable (C^∞) on I and the velocity vector

$$\alpha'(\xi) = \frac{dx^a}{d\xi} e_a$$

is nonzero for every $\xi \in I$ (we adopt the usual custom, in a vector space, of identifying the tangent space at each point with the vector space itself). This definition of smoothness clearly does not depend on the choice of admissible basis for \mathcal{M} . The curve α is said to be spacelike, timelike, or null if

$$\alpha'(\xi) \cdot \alpha'(\xi) = \eta_{ab} \frac{dx^a}{d\xi} \frac{dx^b}{d\xi}$$

is positive, negative, or zero, respectively, for each $\xi \in I$. A timelike curve α for which $\alpha'(\xi)$ is future directed for each $\xi \in I$ is called a timelike world line and its image is identified with the set of all events in the history of some (not necessarily free) point material particle. If $I = [\xi_0, \xi_1]$ and $\alpha: [\xi_0, \xi_1] \rightarrow \mathcal{M}$ is a timelike world line, then the proper time length of α is defined by

$$\begin{aligned} L(\alpha) &= \int_{\xi_0}^{\xi_1} \sqrt{|g(\alpha'(\xi), \alpha'(\xi))|} d\xi \\ &= \int_{\xi_0}^{\xi_1} \sqrt{-\eta_{ab} \frac{dx^a}{d\xi} \frac{dx^b}{d\xi}} d\xi \end{aligned}$$

and interpreted as the time lapse between the events $\alpha(\xi_0)$ and $\alpha(\xi_1)$ as recorded by a clock carried along by the particle whose world line is α . This interpretation is easily motivated by writing out a Riemann sum

approximation to the integral and appealing to our interpretation of the proper time separation $\Delta\tau = \sqrt{-\eta_{ab} \Delta x^a \Delta x^b}$. There are subtleties, however, both mathematical and physical (Naber 1992, section 1.4). The mathematical ones are addressed by the following result (which combines theorems 1.4.6 and 1.4.8 of Naber (1992)).

Theorem *Let x_0 and x be two events in \mathcal{M} . Then $x - x_0$ is timelike and future directed if and only if there exists a timelike world line $\alpha: [\xi_0, \xi_1] \rightarrow \mathcal{M}$ in \mathcal{M} with $\alpha(\xi_0) = x_0$ and $\alpha(\xi_1) = x$ and, in this case,*

$$L(\alpha) \leq \tau(x - x_0) \quad [9]$$

with equality holding if and only if α is a parametrization of a timelike straight line.

The inequality [9] asserts that if two material particles experience both x_0 and x , then the one that is free (and so can be regarded as at rest in some admissible coordinate system) has longer to wait for the occurrence of the second event (moving clocks run slow). For many years this basically obvious fact was christened “The Twin Paradox.”

Just as a smooth curve in Euclidean space has an arc length parametrization, so a timelike world line has a proper time parametrization defined as follows. For each ξ in $[\xi_0, \xi_1]$ let

$$\tau = \tau(\xi) = \int_{\xi_0}^{\xi} \sqrt{|g(\alpha'(\zeta), \alpha'(\zeta))|} d\zeta$$

(the proper time length of α from $\alpha(\xi_0)$ to $\alpha(\xi)$). Then $\tau = \tau(\xi)$ has a smooth inverse $\xi = \xi(\tau)$ so α can be reparametrized by τ . We will abuse our notation slightly and write

$$\alpha(\tau) = x^a(\tau) e_a$$

The velocity vector with this parametrization is denoted

$$U = U(\tau) = \frac{dx^a}{d\tau} e_a$$

called the 4-velocity of the world line and is the unit tangent vector field to α , that is,

$$U(\tau) \cdot U(\tau) = -1 \quad [10]$$

for each τ . An admissible observer is, of course, more likely to parametrize a world line by his own time coordinate x^4 . Then

$$\alpha'(x^4) = \frac{dx^1}{dx^4} e_1 + \frac{dx^2}{dx^4} e_2 + \frac{dx^3}{dx^4} e_3 + e_4$$

so

$$|g(\alpha'(x^4), \alpha'(x^4))| = 1 - \|V\|^2$$

where

$$\|\mathbf{V}\| = \sqrt{\left(\frac{dx^1}{dx^4}\right)^2 + \left(\frac{dx^2}{dx^4}\right)^2 + \left(\frac{dx^3}{dx^4}\right)^2}$$

is the usual magnitude of the particle's velocity vector

$$\begin{aligned} \mathbf{V} &= \mathbf{V}(x^4) \\ &= \frac{dx^1}{dx^4} e_1 + \frac{dx^2}{dx^4} e_2 + \frac{dx^3}{dx^4} e_3 \\ &= V^i e_i \end{aligned}$$

in the given admissible coordinate system. One finds then that

$$U = \left(1 - \|\mathbf{V}\|^2\right)^{-1/2} (\mathbf{V} + e_4) \quad [11]$$

We shall identify a material particle in \mathcal{M} with a pair (α, m) , where α is a timelike world line and m is a positive constant called the particle's proper mass (or rest mass). If each $dx^a/d\xi$, $a=1,2,3,4$, is constant, then (α, m) is a free material particle with proper mass m . The 4-momentum of (α, m) is defined by $P = mU$. Thus,

$$P \cdot P = -m^2 \quad [12]$$

In any admissible basis we write

$$\begin{aligned} P &= P^a e_a = mU^a e_a = m \frac{dx^a}{d\tau} e_a \\ &= m \left(1 - \|\mathbf{V}\|^2\right)^{-1/2} (\mathbf{V} + e_4) \end{aligned} \quad [13]$$

The “spatial part” of P in these coordinates is

$$\mathbf{P} = \frac{m}{\sqrt{1 - \|\mathbf{V}\|^2}} \mathbf{V}$$

which, for $\|\mathbf{V}\| \ll 1$, is approximately $m\mathbf{V}$. Identifying m with the inertial mass of Newtonian mechanics (measured by an observer for whom the particle's speed is small), this is simply the classical momentum of the particle. Somewhat more explicitly, if one expands $1/\sqrt{1 - \|\mathbf{V}\|^2}$ by the Binomial Theorem one finds that

$$\begin{aligned} P^i &= \frac{m}{\sqrt{1 - \|\mathbf{V}\|^2}} V^i \\ &= mV^i + \frac{1}{2} mV^i \|\mathbf{V}\|^2 + \dots, \quad i = 1, 2, 3 \end{aligned} \quad [14]$$

which gives the components of the classical momentum plus “relativistic corrections.” In order to preserve a formal similarity with Newtonian mechanics one often sees $m/\sqrt{1 - \|\mathbf{V}\|^2}$ referred

to as the “relativistic mass” of the particle, but we shall avoid this terminology. The fourth component of P is given by

$$\begin{aligned} P^4 &= -P \cdot e_4 \\ &= \frac{m}{\sqrt{1 - \|\mathbf{V}\|^2}} = m + \frac{1}{2} m \|\mathbf{V}\|^2 + \dots \end{aligned} \quad [15]$$

The appearance of the term $(1/2)m\|\mathbf{V}\|^2$ corresponding to the Newtonian kinetic energy suggests that P^4 be denoted E and called the total relativistic energy measured by the given admissible observer for the particle:

$$E = -P \cdot e_4 \quad [16]$$

Now, one must understand that the concept of “energy” in physics is a subtle one and simply giving $-P \cdot e_4$ this name does not ensure that there is any physical content. Whether or not the name is appropriate can only be determined experimentally. In particular, one should ask if the appearance of the term m in [15] is consistent with the view that P^4 represents the “energy” of the particle. Observe that if $\|\mathbf{V}\| = 0$ (i.e., if the particle is at rest relative to the given observer), then [15] gives

$$E = m (= mc^2, \text{ in standard units}) \quad [17]$$

which we interpret as saying that, even when the particle is at rest, it still has energy. If this is really “energy” in the physical sense, then it should be possible to liberate and use it. That this is, indeed, possible has, of course, been rather convincingly demonstrated.

Next we observe that not only material particles, but also photons possess “momentum” and “energy” and therefore should have 4-momentum (witness, e.g., the photoelectric effect in which photons collide with and eject electrons from their orbits in an atom). Unlike a material particle, however, a photon's characteristic feature is not proper mass, but frequency ν , or wavelength $\lambda = 1/\nu$, related to its energy \mathcal{E} by $\mathcal{E} = h\nu$ (h being Planck's constant) and these are highly observer dependent (Doppler effect). There is, moreover, no “proper frequency” analogous to “proper mass” since there is no admissible observer for whom the photon is at rest. In an attempt to model these features we consider a point $x_0 \in \mathcal{M}$, a future directed null vector N and an interval $I \subseteq \mathbb{R}$. The curve $\alpha: I \rightarrow \mathcal{M}$ defined by

$$\alpha(\xi) = x_0 + \xi N \quad [18]$$

is a parametrization of the world line of a photon through x_0 . Being null, N can be written in any admissible basis as

$$N = (-N \cdot e_4)(\mathbf{d} + e_4) \quad [19]$$

where

$$\begin{aligned} \mathbf{d} = & \left[(N \cdot e_1)^2 + (N \cdot e_2)^2 \right. \\ & \left. + (N \cdot e_3)^2 \right]^{-1/2} \left[(N \cdot e_1)e_1 \right. \\ & \left. + (N \cdot e_2)e_2 + (N \cdot e_3)e_3 \right] \end{aligned} \quad [20]$$

is the direction vector of the world line in the corresponding spatial coordinate system. Now, by analogy with [16], we define a photon in \mathcal{M} to be a curve in \mathcal{M} of the form [18], take N to be its 4-momentum and define the energy \mathcal{E} of the photon in the admissible basis $\{e_1, e_2, e_3, e_4\}$ by

$$\mathcal{E} = -N \cdot e_4 \quad [21]$$

Then, by [19],

$$N = \mathcal{E}(\mathbf{d} + e_4) \quad [22]$$

The corresponding frequency ν and wavelength λ are then defined by $\nu = \mathcal{E}/h$ and $\lambda = 1/\nu$. In another admissible basis, one has $N = \hat{\mathcal{E}}(\hat{\mathbf{d}} + \hat{e}_4)$, where $\hat{\mathbf{d}}$ and $\hat{\mathcal{E}}$ are defined by the hatted versions of [20] and [21]. One can then show (Naber 1992, section 1.8) that

$$\begin{aligned} \frac{\hat{\mathcal{E}}}{\mathcal{E}} = \frac{\hat{\nu}}{\nu} &= \frac{1 - \beta \cos \theta}{\sqrt{1 - \beta^2}} \\ &= (1 - \beta \cos \theta) + \frac{1}{2}\beta^2(1 - \beta \cos \theta) + \dots \end{aligned} \quad [23]$$

where β is the relative speed of the two spatial coordinate systems and θ is the angle (in the unhatted spatial coordinate system) between the direction \mathbf{d} of the photon and the direction of motion of the hatted spatial coordinate system. Equation [23] is the formula for the relativistic Doppler effect with the first term in the series being the classical formula.

We conclude this section by examining a few simple interactions between particles of the sort modeled by our definitions, assuming only that 4-momentum is conserved in the interaction. For convenience, we will use the term free particle to refer to either a free material particle or a photon. If \mathcal{A} is a finite set of free particles, then each element of \mathcal{A} has a unique 4-momentum which is a future-directed timelike or null vector. The sum of any such collection of vectors is timelike and future directed, except when all of the vectors are null and

parallel, in which case the sum is null and future directed (Naber 1992, lemma 1.4.3). We call this sum the total 4-momentum of \mathcal{A} . Now we formulate a definition which is intended to model a finite set of free particles colliding at some event with a (perhaps new) set of free particles emerging from the collision (e.g., an electron and proton collide, with a neutron and neutrino emerging from the collision). A contact interaction in \mathcal{M} is a triple $(\mathcal{A}, x, \tilde{\mathcal{A}})$, where \mathcal{A} and $\tilde{\mathcal{A}}$ are two finite sets of free particles, neither of which contains a pair of particles with linearly dependent 4-momenta (which would presumably be physically indistinguishable) and $x \in \mathcal{M}$ is an event such that

1. x is the terminal point of all of the particles in \mathcal{A} (i.e., for each world line $\alpha: [\xi_0, \xi_1] \rightarrow \mathcal{M}$ of a particle in \mathcal{A} , $\alpha(\xi_1) = x$);
2. x is the initial point of all the particles in $\tilde{\mathcal{A}}$, and
3. the total 4-momentum of \mathcal{A} equals the total 4-momentum of $\tilde{\mathcal{A}}$.

Properly (3) is called the conservation of 4-momentum. If \mathcal{A} consists of a single free particle, then $(\mathcal{A}, x, \tilde{\mathcal{A}})$ is called a decay (e.g., a neutron decays into a proton, an electron and an antineutrino).

Consider, for example, an interaction $(\mathcal{A}, x, \tilde{\mathcal{A}})$ for which $\tilde{\mathcal{A}}$ consists of a single photon. The total 4-momentum of $\tilde{\mathcal{A}}$ is null so the same must be true of \mathcal{A} . Since the 4-momenta of the individual particles in \mathcal{A} are timelike or null and future directed their sum can be null only if they are, in fact, all null and parallel. Since \mathcal{A} cannot contain distinct photons with parallel 4-momenta, it must consist of a single photon which, by (3), must have the same 4-momentum as the photon in $\tilde{\mathcal{A}}$. In essence, “nothing happened at x .” We conclude that *no nontrivial interaction of the type modeled by our definition can result in a single photon and nothing else*. Reversing the roles of \mathcal{A} and $\tilde{\mathcal{A}}$ shows that, if 4-momentum is to be conserved, *a photon cannot decay*.

Next let us consider the decay of a single material particle into two material particles, for example, the spontaneous disintegration of an atom through α -emission. Thus, we consider a contact interaction $(\mathcal{A}, x, \tilde{\mathcal{A}})$ in which \mathcal{A} consists of a single free material particle of proper mass m_0 and $\tilde{\mathcal{A}}$ consists of two free material particles with proper masses m_1 and m_2 . Let P_0, P_1 , and P_2 be the 4-momenta of the particles of proper mass m_0, m_1 , and m_2 , respectively. Then $P_0 = P_1 + P_2$. Appealing to the “reversed triangle inequality,” the fact that P_1 and P_2 are linearly independent and future directed, and [12] we conclude that

$$m_0 > m_1 + m_2 \quad [23]$$

The excess mass $m_0 - (m_1 + m_2)$ of the initial particle is regarded, via [17], as a measure of the amount of energy required to split m_0 into two pieces. Stated somewhat differently, when the two particles in $\tilde{\mathcal{A}}$ were held together to form the single particle in \mathcal{A} , the “binding energy” contributed to the mass of this latter particle.

Reversing the roles of \mathcal{A} and $\tilde{\mathcal{A}}$ in the last example gives a contact interaction modelling an inelastic collision (two free material particles with masses m_1 and m_2 collide and coalesce to form a third of mass m_0). The inequality [23] remains true, of course, and a somewhat more detailed analysis (Naber 1992, section 1.8) yields an approximate formula for $m_0 - (m_1 + m_2)$ which can be compared (favorably) with the Newtonian formula for the loss in kinetic energy that results from the collision (energy which, classically, is viewed as taking the form of heat in the combined particle). An analysis of the interaction in which both \mathcal{A} and $\tilde{\mathcal{A}}$ consist of an electron and a photon yields (Naber 1992, section 1.8) a formula for the so-called Compton effect. Many more such examples of this sort are treated in great detail in Synge (1972, chapter VI, § 14).

Charged Particles and Electromagnetic Fields

A charged particle in \mathcal{M} is a triple (α, m, q) , where (α, m) is a material particle and q is a nonzero real number called the charge of the particle. Charged particles do two things of interest to us. By their very presence they create electromagnetic fields and they also respond to the electromagnetic fields created by other charges.

Charged particles “respond” to an electromagnetic field by experiencing changes in 4-momentum. The quantitative nature of this response, that is, the equation of motion, is generally taken to be the so-called Lorentz 4-force law which expresses the proper time rate of change of the particle’s 4-momentum at each point of the world line as a linear function of the 4-velocity. Thus, at each point $\alpha(\tau)$ of the world line

$$\frac{dP(\tau)}{d\tau} = q\tilde{F}_{\alpha(\tau)}(U(\tau)) \quad [24]$$

where $\tilde{F}_{\alpha(\tau)}: \mathcal{M} \rightarrow \mathcal{M}$ is a linear transformation determined, in each admissible coordinate system, by the classical electric E and magnetic B fields (here we are assuming that the contribution of q to the ambient electromagnetic field is negligible, that is,

(α, m, q) is a “test charge”). Let us write [24] more simply as

$$\tilde{F}(U) = \frac{m}{q} \frac{dU}{d\tau} \quad [25]$$

Dotting both sides of [25] with U gives

$$\begin{aligned} \tilde{F}(U) \cdot U &= \frac{m}{q} \frac{dU}{d\tau} \cdot U = \frac{m}{2q} \frac{d}{d\tau} (U \cdot U) \\ &= \frac{m}{2q} \frac{d}{d\tau} (-1) = 0 \end{aligned}$$

Since any future-directed timelike unit vector u is the 4-velocity of some charged particle, we find that $\tilde{F}(u) \cdot u = 0$ for any such vector. Linearity then implies $\tilde{F}(v) \cdot v = 0$ for any timelike vector. Now, if u and v are timelike and future directed, then $u + v$ is timelike so $0 = \tilde{F}(u + v) \cdot (u + v) = \tilde{F}(u) \cdot v + u \cdot \tilde{F}(v)$ and therefore $\tilde{F}(u) \cdot v = -u \cdot \tilde{F}(v)$. But \mathcal{M} has a basis of future-directed timelike vectors so

$$\tilde{F}(x) \cdot y = -x \cdot \tilde{F}(y) \quad [26]$$

for all $x, y \in \mathcal{M}$. Thus, at each point, the linear transformation \tilde{F} must be skew-symmetric with respect to the Lorentz inner product. One could therefore model an electromagnetic field on \mathcal{M} by an assignment to each point of a skew-symmetric linear transformation whose job it is to assign to the 4-velocity of a charged particle whose world line passes through that point the change in 4-momentum that the particle should expect to experience because of the presence of the field. However, a slightly different perspective has proved more convenient. Notice that a skew-symmetric linear transformation $\tilde{F}: \mathcal{M} \rightarrow \mathcal{M}$ and the Lorentz inner product together determine a bilinear form $F: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ given by

$$F(x, y) = \tilde{F}(x) \cdot y$$

which is also skew-symmetric ($F(y, x) = \tilde{F}(y) \cdot x = -F(x, y)$) and that, conversely, a skew-symmetric bilinear form uniquely determines a skew-symmetric linear transformation. Now, an assignment of a skew-symmetric bilinear form to each point of \mathcal{M} is nothing other than a 2-form on \mathcal{M} and it is in the language of forms that we choose to phrase classical electromagnetic theory (a concise introduction to this language is available, for example, in Spivak (1965, chapter 4).

Nature imposes a certain restriction on which 2-forms can reasonably represent an electromagnetic field on \mathcal{M} (“Maxwell’s equations”). To formulate these we introduce a source 1-form J as follows: If

x^1, x^2, x^3, x^4 is any admissible coordinate system on \mathcal{M} , then

$$J = J_1 dx^1 + J_2 dx^2 + J_3 dx^3 - \rho dx^4 \quad [27]$$

where $\rho: \mathcal{M} \rightarrow \mathbb{R}$ is a charge density function and $J = J_1 e_1 + J_2 e_2 + J_3 e_3$ is a current density vector field (these are to be regarded as the usual “smoothed out,” pointwise versions of “charge per unit volume” and “charge flow per unit area per unit time” as measured by the corresponding admissible observer). Now, our formal definition is as follows: The electromagnetic field on \mathcal{M} determined by the source 1-form J on \mathcal{M} is a 2-form F on \mathcal{M} that satisfies Maxwell’s equation

$$dF = 0 \quad [28]$$

and

$$*d^*F = J \quad [29]$$

A few comments are in order here. We have chosen units in which not only the speed of light, but also various other constants that one often finds in Maxwell’s equations (the dielectric constant ϵ_0 and magnetic permeability μ_0) are 1 and a factor of 4π in [29] is “normalized out.” The $*$ in [29] is the Hodge star operator determined by the Lorentz inner product and the chosen orientation of \mathcal{M} . This is a natural isomorphism

$$*: \Omega^p(\mathcal{M}) \rightarrow \Omega^{4-p}(\mathcal{M}), \quad p = 0, 1, 2, 3, 4$$

of the p -forms on \mathcal{M} to the $(4-p)$ -forms on \mathcal{M} and is most simply defined as follows: let x^1, x^2, x^3, x^4 be any admissible coordinate system on \mathcal{M} . If $1 \in \Omega^0(\mathcal{M})$ is the constant function (0-form) on \mathcal{M} whose value is 1 $\in \mathbb{R}$, then

$$*1 = dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4$$

is the volume form on \mathcal{M} . If $1 \leq i_1 < \dots < i_k \leq 4$, then $*(dx^{i_1} \wedge \dots \wedge dx^{i_k})$ is uniquely determined by

$$\begin{aligned} & (dx^{i_1} \wedge \dots \wedge dx^{i_k}) \wedge *(dx^{i_1} \wedge \dots \wedge dx^{i_k}) \\ &= -dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 \end{aligned}$$

Thus, for example, $*dx^2 = dx^1 \wedge dx^3 \wedge dx^4$, $*(dx^1 \wedge dx^2) = -dx^3 \wedge dx^4$, $*(dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4) = -1$, etc. It follows that, if μ is a p -form on \mathcal{M} , then

$$**\mu = (-1)^{p+1} \mu \quad [30]$$

(a more thorough discussion is available in Choquet-Bruhat *et al.* (1977, chapter V A3)). In particular, [29] is equivalent to

$$d^*F = *J \quad [31]$$

On regions in which there are no charges, so that $J = 0$, [28] and [31] become the source free Maxwell equations

$$dF = 0 \quad [32]$$

and

$$d^*F = 0 \quad [33]$$

that is, both F and $*F$ are closed 2-forms.

Any 2-form F on \mathcal{M} can be written in any admissible coordinate system as $F = (1/2)F_{ab} dx^a \wedge dx^b$ (summation convention!), where (F_{ab}) is the skew-symmetric matrix of components of F . In order to make contact with the notation generally employed in physics, we introduce the following names for these components:

$$(F_{ab}) = \begin{pmatrix} 0 & B^3 & -B^2 & E^1 \\ -B^3 & 0 & B^1 & E^2 \\ B^2 & -B^1 & 0 & E^3 \\ -E^1 & -E^2 & -E^3 & 0 \end{pmatrix} \quad [34]$$

Thus,

$$\begin{aligned} F &= E^1 dx^1 \wedge dx^4 + E^2 dx^2 \wedge dx^4 \\ &+ E^3 dx^3 \wedge dx^4 + B^3 dx^1 \wedge dx^2 \\ &+ B^2 dx^3 \wedge dx^1 + B^1 dx^2 \wedge dx^3 \end{aligned} \quad [35]$$

Computing $*F, dF, d^*F$ and $*d^*F$ and writing $E = E^1 e_1 + E^2 e_2 + E^3 e_3$ and $B = B^1 e_1 + B^2 e_2 + B^3 e_3$, one finds that $dF = 0$ is equivalent to

$$\operatorname{div} B = 0 \quad [36]$$

and

$$\operatorname{curl} E + \frac{\partial B}{\partial t} = 0 \quad [37]$$

while $*d^*F = J$ is equivalent to

$$\operatorname{div} E = \rho \quad [38]$$

and

$$\operatorname{curl} B - \frac{\partial E}{\partial t} = J \quad [39]$$

Equations [36]–[39] are the more traditional renderings of Maxwell’s equations.

In another admissible coordinate system $\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{x}^4$ on \mathcal{M} (related to the first by [2]) the 2-form F would be written $F = (1/2)\hat{F}_{ab} d\hat{x}^a \wedge d\hat{x}^b$. Setting $\hat{x}^a = \Lambda^a_{\alpha} x^{\alpha}$ and $\hat{x}^b = \Lambda^b_{\beta} x^{\beta}$ gives $F = (1/2)(\Lambda^a_{\alpha} \Lambda^b_{\beta} \hat{F}_{ab}) dx^{\alpha} \wedge dx^{\beta}$, so

$$F_{\alpha\beta} = \Lambda^a_{\alpha} \Lambda^b_{\beta} \hat{F}_{ab}, \quad \alpha, \beta = 1, 2, 3, 4 \quad [40]$$

Now, suppose that we wish to describe the electromagnetic field of a uniformly moving charge. According to the relativity principle, it does not matter at all whether we view the charge as moving

relative to a “fixed” admissible observer, or the observer as moving relative to a “stationary” charge. Thus, we shall write out the field due to a charge fixed at the origin of the hatted coordinate system (“Coulomb’s law”) and transform, by [40], to an unhatted coordinate system moving relative to it. Relative to $\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{x}^4$, the familiar inverse square law for a fixed point charge q located at the spatial origin gives $\hat{B} = 0$ and $\hat{E} = (q/\hat{r}^3)\hat{r}$, where $\hat{r} = \hat{x}^1\hat{e}_1 + \hat{x}^2\hat{e}_2 + \hat{x}^3\hat{e}_3$ and $\hat{r} = ((\hat{x}^1)^2 + (\hat{x}^2)^2 + (\hat{x}^3)^2)^{1/2}$ (note that \hat{E} is defined only on $\mathcal{M} - \text{Span}\{\hat{e}_4\}$). Thus,

$$(\hat{F}_{ab}) = \frac{q}{\hat{r}^3} \begin{pmatrix} 0 & 0 & 0 & \hat{x}^1 \\ 0 & 0 & 0 & \hat{x}^2 \\ 0 & 0 & 0 & \hat{x}^3 \\ -\hat{x}^1 & -\hat{x}^2 & -\hat{x}^3 & 0 \end{pmatrix} \quad [41]$$

It is a simple matter to verify that, on its domain, (\hat{F}_{ab}) satisfies the source free Maxwell equations. Taking Λ to be the special Lorentz transformation corresponding to [5] and writing out [40] with (\hat{F}_{ab}) given by [41] yields

$$\begin{aligned} E^1 &= q \left(\frac{\hat{x}^1}{\hat{r}^3} \right) \\ E^2 &= \frac{q}{\sqrt{1-\beta^2}} \left(\frac{\hat{x}^2}{\hat{r}^3} \right) \\ E^3 &= \frac{q}{\sqrt{1-\beta^2}} \left(\frac{\hat{x}^3}{\hat{r}^3} \right) \\ B^1 &= 0 \\ B^2 &= \frac{-q\beta}{\sqrt{1-\beta^2}} \left(\frac{\hat{x}^3}{\hat{r}^3} \right) \\ B^3 &= \frac{q\beta}{\sqrt{1-\beta^2}} \left(\frac{\hat{x}^2}{\hat{r}^3} \right) \end{aligned} \quad [42]$$

We wish to express these in terms of measurements made by the unhatted observer at the instant the charge passes through his spatial origin. Setting $x^4 = 0$ in [5] gives

$$\hat{x}^1 = \frac{1}{\sqrt{1-\beta^2}} x^1, \quad \hat{x}^2 = x^2, \quad \hat{x}^3 = x^3$$

and so

$$\hat{r}^2 = \frac{1}{1-\beta^2} (x^1)^2 + (x^2)^2 + (x^3)^2$$

which, for convenience, we write r_β^2 . Making these substitutions in [42] gives

$$\begin{aligned} E &= \frac{q}{\sqrt{1-\beta^2}} \left(\frac{1}{r_\beta^3} \right) (x^1 e_1 + x^2 e_2 + x^3 e_3) \\ &= \frac{q}{\sqrt{1-\beta^2}} \left(\frac{1}{r_\beta^3} \right) \mathbf{r} \end{aligned} \quad [43]$$

and

$$\begin{aligned} \mathbf{B} &= \frac{q}{\sqrt{1-\beta^2}} \left(\frac{1}{r_\beta^3} \right) (0e_1 - \beta x^3 e_2 + \beta x^2 e_3) \\ &= \frac{q}{\sqrt{1-\beta^2}} \left(\frac{1}{r_\beta^3} \right) ((\beta e_1) \times \mathbf{r}) \end{aligned} \quad [44]$$

for the field of a charge moving uniformly with velocity βe_1 at the instant the charge passes through the origin. Observe that when $\beta \ll 1$, $r_\beta \approx r$, so [43] says that the electric field of a slowly moving charge is approximately the Coulomb field. When $\beta \ll 1$, [44] reduces to the Biot–Savart law.

Let us consider one other simple application, that is, the response of a charged particle (α, m, q) to an electromagnetic field which, for some admissible observer, is constant and purely magnetic. For simplicity, we assume that, for this observer $E = 0$ and $\mathbf{B} = b e_3$, where b is a nonzero constant. The corresponding 2-form F has components

$$(F_{ab}) = \begin{pmatrix} 0 & b & 0 & 0 \\ -b & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(from [34]). The corresponding linear transformation \tilde{F} has the same matrix relative to this basis so, with $\alpha(\tau) = x^a(\tau)e_a$ and $U(\tau) = U^a(\tau)e_a$, the Lorentz 4-force law [25] reduces to the system of linear differential equations

$$\begin{aligned} \frac{dU^1}{d\tau} &= \frac{bq}{m} U^2, & \frac{dU^2}{d\tau} &= -\frac{bq}{m} U^1 \\ \frac{dU^3}{d\tau} &= 0, & \frac{dU^4}{d\tau} &= 0 \end{aligned}$$

The system is easily solved and the results easily integrated to give

$$\begin{aligned} \alpha(\tau) &= x_0 + a \sin\left(\frac{bq\tau}{m} + \phi\right) e_1 \\ &\quad + a \cos\left(\frac{bq\tau}{m} + \phi\right) e_2 \\ &\quad + c\tau e_3 + \left(1 + \frac{a^2 b^2 q^2}{m^2} + c^2\right) \tau e_4 \end{aligned} \quad [45]$$

where $x_0 = x_0^a e_a \in \mathcal{M}$ is constant and a, ϕ , and c are real constants with $a > 0$ (we have used $U \cdot U = -1$ to eliminate one other arbitrary real constant). Note that, at each point on α , $(x^1 - x_0^1)^2 + (x^2 - x_0^2)^2 = a^2$. Thus, if $c \neq 0$ the spatial trajectory in this coordinate system is a helix along the e_3 -direction (i.e., along the magnetic field lines). If $c = 0$, the trajectory is a circle in the x^1 – x^2 plane. This case is of some practical significance since one can

introduce constant magnetic fields in a bubble chamber so as to induce a particle of interest to follow a circular path. We show now how to measure the charge-to-mass ratio for such a particle. Taking $c=0$ in [45] and computing $U(\tau)$, then using [11] to solve for the coordinate velocity vector \mathbf{V} of the particle gives

$$\mathbf{V} = \frac{abq/m}{\sqrt{1 - \|\mathbf{V}\|^2}} \left(\cos\left(\frac{bq\tau}{m} + \phi\right) e_1 + \sin\left(\frac{bq\tau}{m} + \phi\right) e_2 \right)$$

From this one computes

$$\|\mathbf{V}\|^2 = \left(1 + \frac{m^2}{a^2 b^2 q^2} \right)^{-1}$$

(note that this is a constant). Solving this last equation for q/m (and assuming $q > 0$ for convenience) one arrives at

$$\frac{q}{m} = \frac{1}{a|b|} \frac{\|\mathbf{V}\|}{\sqrt{1 - \|\mathbf{V}\|^2}}$$

Since a , b , and $\|\mathbf{V}\|$ are measurable, one obtains the desired charge-to-mass ratio.

To conclude we wish to briefly consider the existence and use of “potentials” for electromagnetic fields. Suppose F is an electromagnetic field defined on some connected, open region X in \mathcal{M} . Then F is a 2-form on X which, by [28], is closed. Suppose also that the second de Rham cohomology $H^2(X; \mathbb{R})$ of X is trivial (since \mathcal{M} is topologically \mathbb{R}^4 this will be the case, for example, when X is all of \mathcal{M} , or an open ball in \mathcal{M} , or, more generally, an open “star-shaped” region in \mathcal{M}). Then, by definition, every closed 2-form on X is exact so, in particular, there exists a 1-form A on X satisfying

$$F = dA \quad [46]$$

In particular, such a 1-form A always exists locally on a neighborhood of any point in X for any F . Such an A is not uniquely determined, however, because, if A satisfies [46], then so does $A + df$ for any smooth real-valued function (0-form) f on X ($d^2 = 0$ implies $d(A + df) = dA + d^2 f = dA = F$). Any 1-form A satisfying [46] is called a (gauge) potential for F . The replacement $A \rightarrow A + df$ for some f is called a gauge transformation of the potential and the freedom to make such a replacement without altering [46] is called gauge freedom.

One can show that, given F , it is always possible to locally solve $dA = F$ for A subject to an arbitrary specification of the 0-form $*d^*A$. More precisely, if F

is any 2-form satisfying $dF = 0$ and g is an arbitrary 0-form, then locally, on a neighborhood of any point, there exists a 1-form A satisfying

$$dA = F \quad \text{and} \quad *d^*A = g \quad [47]$$

(a more general result is proved in Parrott (1987, appendix 2) and a still more general one in section 2.9 of this same source). The usefulness of the second condition in [47] can be illustrated as follows. Suppose we are given some (physical) configuration of charges and currents (i.e., some source 1-form J) and we wish to find the corresponding electromagnetic field F . We must solve Maxwell’s equations $dF = 0$ and $*d^*F = J$ (subject to whatever boundary conditions are appropriate). Locally, at least, we may seek instead a corresponding potential A (so that $F = dA$). Then the first of Maxwell’s equations is automatically satisfied ($dF = d(dA) = 0$) and we need only solve $*d^*(dA) = J$. To simplify the notation let us temporarily write $\delta = *d^*$ and consider the operator $\Delta = d \circ \delta + \delta \circ d$ on forms (variously called the Laplace–Beltrami operator, Laplace–de Rham operator, or Hodge Laplacian on Minkowski spacetime). Then

$$\Delta A = d(\delta A) + \delta(dA) = d(*d^*A) + *d^*(dA) \quad [48]$$

According to the result quoted above, we may narrow down our search by imposing the condition $*d^*A = 0$, that is

$$\delta A = 0 \quad [49]$$

(this is generally referred to as imposing the Lorentz gauge). With this, [48] becomes $\Delta A = *d^*(dA)$ and to satisfy the second Maxwell equation we must solve

$$\Delta A = J \quad [50]$$

Thus, we see that the problem of (locally) solving Maxwell’s equations for a given source J reduces to that of solving [49] and [50] for the potential A . To understand how this simplifies the problem, we note that a calculation in admissible coordinates shows that the operator Δ reduces to the componentwise d’Alembertian \square , defined on real-valued functions by

$$\square = \frac{\partial^2}{\partial(x^1)^2} + \frac{\partial^2}{\partial(x^2)^2} + \frac{\partial^2}{\partial(x^3)^2} - \frac{\partial^2}{\partial(x^4)^2}$$

Thus, eqn [50] decouples into four scalar equations

$$\square A_a = J_a, \quad a = 1, 2, 3, 4 \quad [51]$$

each of which is the well-studied inhomogeneous wave equation.

Further Reading

Choquet-Bruhat Y, De Witt-Morette C, and Dillard-Bleick M (1977) *Analysis, Manifolds and Physics*. Amsterdam: North-Holland.

Einstein A *et al.* (1958) *The Principle of Relativity*. New York: Dover.

Naber GL (1992) *The Geometry of Minkowski Spacetime*. Berlin: Springer.

Parrott S (1987) *Relativistic Electrodynamics and Differential Geometry*. Berlin: Springer.

Spivak M (1965) *Calculus on Manifolds*. New York: W A Benjamin.

Synge JL (1972) *Relativity: The Special Theory*. Amsterdam: North-Holland.

Introductory Article: Quantum Mechanics

G F dell'Antonio, Università di Roma "La Sapienza," Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

Historical Background

In this section we shall briefly recall the basic empirical facts and the first theoretical attempts from which the theory and the formalism of present-day quantum mechanics (QM) has grown. In the next sections we shall give the mathematical and computational structure of QM, mention the physical problems that QM has solved with much success, and describe the serious conceptual consistency problems which are posed by QM (and which remain unsolved up to now).

Empirical rules of discretization were observed already, starting from the 1850s, in the absorption and in the emission of light. Fraunhofer noticed that the dark lines in the absorption spectrum of the light of the sun coincide with the bright lines in the emission lines of all elements. G Kirchhoff and R Bunsen reached the conclusion that the relative intensities of the emission and absorption of light implied that the ratio between energy emitted and absorbed is independent of the atom considered. This was the starting point of the analysis by Planck.

On the other hand, by the end of the eighteenth century, the spatial structure of the atom had been investigated; the most successful model was that of Rutherford, in which the atom appeared as a small nucleus of charge Z surrounded by Z electrons attracted by the nucleus according to Coulomb's law. This model represents, for distances of the order of the size of an atom, a complete departure from Newton's laws combined with the laws of classical electrodynamics; indeed, according to these laws, the atom would be unstable against collapse, and would certainly not exhibit a discrete energy spectrum. We must conclude that the classical laws

are inadequate for the description of emission and absorption of light, in which the internal structure of the atom plays a major role.

The birth of the old quantum theory is placed traditionally at the date of M Planck's discussion of the blackbody radiation in 1900.

Planck put forward the postulate that light is emitted and absorbed by matter in discrete energy quanta through "resonators" that have an energy proportional to their frequency. This assumption led, through the use of Gibbs's rules of Statistical Mechanics applied to a gas of resonators, to a law (Planck's law) which reproduces the empirical findings on the radiation from a blackbody. It led Einstein to ascribe to light (which had, since the times of Maxwell, a successful description in terms of waves) a discrete, particle-like nature. Nine years later A Einstein gave further support to Planck's postulate by showing that it can reproduce correctly the energy fluctuations in blackbody radiation and even clarifies the properties of specific heat. Soon afterwards, [Einstein \(1924, 1925\)](#) proved that the putative particle of light satisfied the relativistic laws (relation between energy and momentum) of a particle with zero mass.

This dual nature of light received further support from the experiments on the Compton effect and from description, by Einstein, of the photoelectric effect ([Einstein 1905](#)). It should be emphasized that while Planck considered with light in interaction with matter ν as composed of bits of energy $h\nu$ ($h \simeq 6,6 \times 10^{-27}$ erg s), Einstein's analysis went much further in assigning to the quantum of light properties of a particle-like (localized) object. This marks a complete departure from the laws of classical electromagnetism. Therefore, quoting Einstein,

It is conceivable that the wave theory of light, which retains its effectiveness for the representation of purely optical phenomena and is based on continuous functions over space, will lead to contradiction with the experiments when applied to phenomena in which there is creation or conversion of light; indeed these phenomena can be better

described on the assumption that light is distributed discontinuously in space and described by a finite number of quanta which move without being divided and which must be absorbed or emitted as a whole.

Notice that, for wavelength of $8 \times 10^3 \text{ \AA}$, a 30 W lamp emits roughly 10^{20} photons s^{-1} ; for macroscopic objects the discrete nature of light has no appreciable consequence.

Planck's postulate and energy conservation imply that in emitting and absorbing light the atoms of the various elements can lose or gain energy only by discrete amounts. Therefore, atoms as producers or absorbers of radiation are better described by a theory that assigns to each atom a (possible infinite) discrete set of states which have a definite energy.

The old quantum theory of matter addresses precisely this question. Its main proponent is N Bohr (Bohr 1913, 1918). The new theory is entirely phenomenological (as is Planck's theory) and based on Rutherford's model and on three more postulates (Born 1924):

- (i) The states of the atom are stable periodic orbits, as given by Newton's laws, of energy $E_n, n \in \mathbb{Z}^+$, given by $E_n = h\nu_n f(n)$, where h is Planck's constant, ν_n is the frequency of the electron on that orbit, and $f(n)$ is for each atom a function approximately linear in Z at least for small values of Z .
- (ii) When radiation is emitted or absorbed, the atom makes a transition to a different state. The frequency of the radiation emitted or absorbed when making a transition is $\nu_{n,m} = h^{-1}|E_n - E_m|$.
- (iii) For large values of n and m and small values of $(n - m)/(n + m)$ the prediction of the theory should agree with those of the classical theory of the interaction of matter with radiation.

Later, A Sommerfeld gave a different version of the first postulate, by requiring that the allowed orbits be those for which the classical action is an integer multiple of Planck's constant.

The old quantum theory met success when applied to simple systems (atoms with $Z < 5$) but it soon appeared evident that a new, radically different point of view was needed and a fresh start; the new theory was to contain few free parameters, and the role of postulate (iii) was now to fix the value of these parameters.

There were two (successful) attempts to construct a consistent theory; both required a more sharply defined mathematical formalism. The first one was sparked by W Heisenberg, and further important ideas and mathematical support came from M Born,

P Jordan, W Pauli, P Dirac and, on the mathematical side, also by J von Neumann and A Weyl. This formulation maintains that one should only consider relations between observable quantities, described by elements that depend only on the initial and final states of the system; each state has an internal energy. By energy conservation, the difference between the energies must be proportional (with a universal constant) to the frequency of the radiation absorbed or emitted. This is enough to define the energy of the state of a single atom modulo an additive constant. The theory must also take into account the probability of transitions under the influence of an external electromagnetic field.

We shall give some details later on, which will help to follow the basis of this approach.

The other attempt was originated by L de Broglie following early remarks by HW Bragg and M Brillouin. Instead of emphasizing the discrete nature of light, he stressed the possible wave nature of particles, using as a guide the Hamilton–Jacobi formulation of classical mechanics. This attempt was soon supported by the experiments of Davisson and Germer (1927) of scattering of a beam of ions from a crystal. These experiments showed that, while electrons are recorded as “point particles,” their distribution follows the law of the intensity for the diffraction of a (dispersive) wave. Moreover, the relation between momentum and frequency was, within experimental errors, the same as that obtained by Einstein for photons.

The theory started by de Broglie was soon placed in almost definitive form by E Schrödinger. In this approach one is naturally led to formulate and solve partial differential equations and the full development of the theory requires regularity results from the theory of functions.

Schrödinger soon realized that the relations which were found in the approach of Heisenberg could be easily (modulo technical details which we shall discuss later) obtained within the formalism he was advocating and indeed he gave a proof that the two formalisms were equivalent. This proof was later refined, from the mathematical point of view, by J von Neumann and G Mackey.

In fact, Schrödinger's approach has proved much more useful in the solution of most physical problems in the nonrelativistic domain, because it can rely on the developments and practical use of the theory of functions and of partial differential equations. Heisenberg's “algebraic” approach has therefore a lesser role in solving concrete problems in (nonrelativistic) QM.

If one considers processes in which the number of particles may change in time, one is forced to

introduce a Hilbert space that accommodates states with an arbitrarily large number of particles, as is the case of the theory of relativistic quantized field or in quantum statistical mechanics; it is then more difficult to follow the line of Schrödinger, due to difficulties in handling spaces of functions of infinitely many variables. The approach of Heisenberg, based on the algebra of matrices, has a rather natural extension to suitable algebras of operators; the approach of Schrödinger, based on the description of a state as a (wave) function, encounters more difficulties since one must introduce functionals over spaces of functions and the description of dynamics does not have a simple form.

From this point of view, the generalization of Heisenberg's approach has led to much progress in the understanding of the structure of the resulting theory. Still some relevant results have been obtained in a Schrödinger representation. We shall not elaborate further on this point.

We shall end this introductory section with a short description of the emergence of the structure of QM in Heisenberg's and Schrödinger's approaches; this will provide a motivation for the axiom of QM which we shall introduce in the following section. For an extended analysis, see, for example, Jammer (1979).

The specific form that was postulated by de Broglie (1923) for the wave nature of a particle relies on the relation of geometrical optics with wave propagation and on the formulation of Hamiltonian mechanics as a sort of "wave front propagation" through the solution of the Hamilton–Jacobi equation and the introduction of group velocity.

By the analogy with electromagnetic wave, it is natural to associate with a free nonrelativistic particle of momentum p and mass m the plane wave

$$\phi_p(x, t) = e^{i(px - Et)/\hbar}, \quad \hbar = \frac{h}{2\pi}, \quad E = \frac{p^2}{2m}$$

Schrödinger obtained the equation for a quantum particle in a field of conservative forces with potential $V(x)$ by considering an analogy with the propagation of an electromagnetic wave in a medium with refraction index $n(x, \omega)$ that varies slowly on the scale of the wavelength. Indeed, in this case the "wave" follows the laws of geometrical optics, and has therefore a "particle-like" behavior. If one denotes by $\hat{u}(x, \omega)$ the Fourier transform (with respect to time) of a generic component of the electric field and one assumes that the field be essentially monochromatic (so that the support of $\hat{u}(x, \omega)$ as a function of ω is in a very small

neighborhood of ω_0), one finds that $\hat{u}(x, \omega)$ is an approximate solution of the equation

$$-\Delta \hat{u}(x, \omega) = \frac{\omega_0^2}{c^2} n^2(x, \omega) \hat{u}(x, \omega) \quad [1]$$

Writing $u(x, \omega) = A(x, \omega) e^{i(\omega/c)W(x, \omega)}$ the phase $W(x, \omega)$ satisfies, in the high-frequency limit, the eikonal equation $|\nabla W(x, \omega)|^2 = n^2(x, \omega)$. One can define for the solution a *phase velocity* v_ϕ and it turns out that $v_\phi = c/|\nabla W(x, \omega)|$.

On the other hand, classical mechanics can also be described by propagation of surfaces of constant value for the solution $W(x, t)$ of the Hamilton–Jacobi equation $H(x, \nabla W) = E$, with $H = p^2/2m + V(x)$. Recall that high-frequency (the realm of geometrical optics) corresponds to small distances. This analogy led Schrödinger (1926) to postulate that the dynamics satisfied by the waves associated with the particles was given by the (Schrödinger) equation

$$i\hbar \frac{\partial \psi(x, t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta_x \psi(x, t) + V(x) \psi(x, t) \quad [2]$$

This wave was to describe the particle and its motion, but, being complex valued, it could not represent any measurable property. It is a mathematical property of the solutions of [2] that the quantity $\int |\psi(x, t)|^2 d^3x$ is preserved in time. Furthermore, if one sets

$$\begin{aligned} \rho(x, t) &\equiv |\psi(x, t)|^2 \\ j(x, t) &\equiv -i \frac{\hbar}{2m} [\bar{\psi}(x, t) \nabla \psi(x, t) - \psi(x, t) \nabla \bar{\psi}(x, t)] \end{aligned} \quad [3]$$

one easily verifies the local conservation law

$$\frac{\partial \rho}{\partial t} + \text{div } j(x, t) = 0 \quad [4]$$

These mathematical properties led to the statistical interpretation given by Max Born: in those experiments in which the position of the particles is measured, the integral of $|\psi(x, t)|^2$ over a region Ω of space gives the probability that at time t the particle is localized in the region Ω . Moreover, the current associated with a charged particle is given locally by $j(x, t)$ defined above.

Let us now briefly review Heisenberg's approach. At the heart of this approach are: empirical formulas for the intensities of emission and absorption of radiation (dispersion relations), Sommerfeld's quantum condition for the action and the vague statement "the analogue of the derivative for the discrete action variable is the corresponding finite difference quotient." And, most important, the remark that the correct description of atomic physics was through quantities associated with pairs of states, that is, (infinite) matrices and the

empirical fact that the frequency (or rather the wave number) $\omega_{k,j}$ of the radiation (emitted or absorbed) in the transition between the atomic levels k and j ($k \neq j$) satisfies the Ritz combination principle $\omega_{m,j} + \omega_{j,k} = \omega_{m,k}$. It is easy to see that any doubly indexed family satisfying this relation must have the form $\omega_{m,k} = E_m - E_k$ for suitable constant E_j .

It was empirically verified by Kramers that the dipole moment of an atom in an external monochromatic external field with frequency ν was proportional to the field with a coefficient (of polarization)

$$P = \frac{e^2}{4\pi m} \sum_i \left[\frac{f_i}{\nu_i^2 - \nu^2} - \frac{F_i}{\nu_i^2 - \nu^2} \right] \quad [5]$$

where e , m are the charge and the mass of the electron and f_i, F_i are the probabilities that the frequency ν is emitted or absorbed.

A detailed analysis of the phenomenon of polarization in classical mechanics, with the clearly stated aim “of presenting the results in a way that may give hints for the construction of a New Mechanics” was made by Max Born (1924). He makes use of action-angle variables $\{J_i, \theta_i\}$ assuming that the atom can be considered as a collection of harmonic oscillators with frequency ν_i coupled linearly to the electric field of frequency μ .

In the dipole approximation one obtains the following result for the polarization P (linear response in energy to the electric field):

$$P = - \sum_{(\nu \cdot m) > 0} 2(m \cdot \nabla_j) \frac{|A(J)|^2 (\nu \cdot m)}{((m \cdot \nu)^2 - \mu^2)} \quad [6]$$

where $\nu_k = \partial H / \partial J_k$, H is the interaction Hamiltonian, and $A(J)$ is a suitable matrix. In order to derive the new dynamics, having as a guide the correspondence principle, one has to compare this result with the Kramers dispersion relation, which we write (to make the comparison easier) in the form

$$P = \frac{e^2}{4\pi m} \sum_{n,m} \frac{f_{m,n}}{\nu_{n,m}^2 - \mu^2} - \frac{f_{n,m}}{\nu_{n,m}^2 - \mu^2} \quad E_m > E_n \quad [7]$$

Bohr’s rule implies that $\nu(n + \tau, n) = (E(n + \tau) - E(n)) / \hbar$.

Born and Heisenberg noticed that, for n sufficiently large and k small, one can approximate the differential operator in [6] with the corresponding difference operator, with an error of the order of k/n . Therefore, [6] could be substituted by

$$P = - \hbar^{-1} \sum_{m_k > 0} \left[\frac{|A_{n+m,n}|^2}{\nu(n+m)^2 - \mu^2} - \frac{|A_{n-m,n}|^2}{\nu(n-m)^2 - \mu^2} \right] \quad [8]$$

The conclusion Born and Heisenberg drew is that the matrix A that takes the place of the momentum in the classical theory must be such that $|A_{n+m,n}|^2 = e^2 \hbar m^{-1} f(n+m, n)$. In the same vein, considering the polarization in a static electric field, it is possible to find an expression for the matrix that takes the place of the coordinate x in classical Hamiltonian theory.

In general, the new approach (matrix mechanics) associates matrices with some relevant classical observables (such as functions of position or momentum) with a time dependence that is derived from the empirical dispersion relations of Kramers, the correspondence principle, Bohr’s rule, Sommerfeld action principle and first- (and second-) order perturbation theory for the interaction of an atom with an external electromagnetic field. It was soon clear to Born and Jordan (1925) that this dynamics took the form $i\hbar \dot{A} = AH - HA$ for a matrix H that for the case of the hydrogen atom is obtained for the classical Hamiltonian with the prescription given for the coordinates x and p . It was also seen as plausible the relation $[\hat{x}_b, \hat{p}_k] = i\hbar$ among the matrices \hat{x}_k and \hat{p}_k corresponding to position and momentum. One year later P Dirac (1926) pointed out the structural identity of this relation with the Poisson bracket of Hamiltonian dynamics, developed a “quantum algebra” and a “quantum differentiation” and proved that any $*$ -derivation δ (derivation which preserves the adjoint) of the algebra \mathcal{B}_N of $N \times N$ matrices is inner, that is, is given by $\delta(a) = i[a, b]$ for a Hermitian matrix b . Much later this theorem was extended (with some assumptions) to the algebra of all bounded operators on a separable Hilbert space. Since the derivations are generators of a one-parameter continuous group of automorphisms, that is, of a dynamics, this result led further strength to the ideas of Born and Heisenberg.

The algebraic structure introduced by Born, Jordan, and Heisenberg (1926) was used by Pauli (1927) to give a purely group-theoretical derivation of the spectrum of the hydrogen atom, following the lines of the derivation in symplectic mechanics of the SO(4) symmetry of the Coulomb system. This remarkable success gave much strength to the Heisenberg formulation of QM, which was soon recognized as an efficient instrument in the study of the atomic world.

The algebraic formulation was also instrumental in the description given by Pauli (1928) of the “spin” (a property of electrons empirically postulated by Goudsmid and Uhlenbeck to account for a hyperfine splitting of some emission lines) as “internal” degree of freedom without reference to spatial coordinates and still connected with the

properties of the the system under the group of spatial rotations. This description through matrices has a major role also in the formulation by Pauli of the exclusion principle (and its relation with Fermi–Dirac statistics), which gave further credit to the Heisenberg’s theory by helping in reproducing correctly the classification of the atoms.

These features may explain why the “standard” formulation of the axioms of QM given in the next section shows the influence of Heisenberg’s approach. On the other hand, comparison with experiments is usually set in the framework in Schrödinger’s approach. Posing the problems in terms of properties of the solution of the Schrödinger equation, one is led to a pragmatic use of the formalism, leaving aside difficulties of interpretation. This separation of “the axioms” from the “practical use” may be one of the reasons why a serious analysis of the axioms and of the problems that arise from them is apparently not a concern for most of the research in QM, even from the point of view of mathematical physics.

One should stress that both the approach of Born and Heisenberg and that of de Broglie and Schrödinger are rooted in a mixture of attention to the experimental data, deep understanding of the previous theory, bold analogies and approximations, and deep concern for the consistency of the “new mechanics.”

There is an essential difference between the starting points of the two approaches. In Heisenberg’s approach, the atom has *a priori* no spatial structure; the description is entirely in terms of its properties under emission and absorption of light, and therefore its observable quantities are represented by matrices. Dynamics enters through the study of the interaction with the electromagnetic field, and some analogies with the classical theory of electrodynamics in an asymptotic regime (correspondence principle). In this way, as we have briefly indicated, the special role of some matrices, which have a mutual relation similar to the relation of position and momentum in Hamiltonian theory. Following this analogy, it is possible to extend the theory beyond its original scope and consider phenomena in which the electrons are not bound to an atom.

In the approach of Schrödinger, on the other hand, particles and collections of particles are represented by spatial structures (waves). Spatial coordinates are therefore introduced *a priori*, and the position of a particle is related to the intensity of the corresponding wave (this was stressed by Born). Position and momentum are both basic measurable quantities as in classical mechanics. Physical

interpretation forces the particle wave to be square integrable, and mathematics provides a limitation on the simultaneous localization in momentum and position leading to Heisenberg’s uncertainty principle. Dynamics is obtained from a particle–wave duality and an analogy with the relativistic wave equation in the low-energy regime. The presence of bound states with quantized energies is seen as a consequence of the well-known fact that waves confined to a bounded spatial region have their wave number (and therefore energy) quantized.

Formal Structure

In this section we describe the formal mathematical structure that is commonly associated with QM. It constitutes a coherent mathematical theory, but the interpretation axiom it contains leads to conceptual difficulties.

We state the axioms in the form in which they were codified by J von Neumann (1966); they constitute a mathematically precise rendering of the formalism of Born, Heisenberg, and Jordan. The formalism of Schrödinger *per se* does not require general statements about the category of observables.

Axiom I

- (i) Observables are represented by self-adjoint operators in a complex separable Hilbert space \mathcal{H} .
- (ii) Every such operator represents an observable.

Remark Axiom I (ii) is introduced only for mathematical simplicity. There is no physical justification for part (ii). In principle, an observable must be connected to a procedure of measurement (observation) and for most of the self-adjoint operators on \mathcal{H} (e.g., in the Schrödinger representation for $i\mathbf{x}_k(\partial/\partial\mathbf{x}_k)\mathbf{x}_k$) such procedure has not yet been given).

Axiom II

- (i) Pure states of the systems are represented by normalized vectors in \mathcal{H} .
- (ii) If a measurement of the observable A is made on a system in the state represented by the element $\phi \in \mathcal{H}$, the average of the numerical values one obtains is $\langle \phi, A\phi \rangle$, a real number because A is self-adjoint (we have denoted by $\langle \phi, \psi \rangle$ the scalar product in \mathcal{H}).

Remark Notice that Axiom II makes no statement about the outcome of a single measurement.

Using the natural complex structure of $\mathcal{B}(\mathcal{H})$, pure states can be extended as linear real functionals on $\mathcal{B}(\mathcal{H})$.

One defines a state as any linear real positive functional on $\mathcal{B}(\mathcal{H})$ (all bounded operators on the separable Hilbert space \mathcal{H}) and says that a state is normal if it is continuous in the strong topology. It can be proved that a normal state can be decomposed into a convex combination of at most a denumerable set of pure states. With these definitions a state is pure iff it has no nontrivial decomposition. It is worth stressing that this statement is true only if the operators that correspond to observable quantities generate all of $\mathcal{B}(\mathcal{H})$; one refers to this condition by stating that there are no superselection rules.

By general results in the theory of the algebra $\mathcal{B}(\mathcal{H})$, a normal state ρ is represented by a positive operator of trace class σ through the formula $\rho(A) = \text{Tr}(\sigma A)$. Since a positive trace-class operator (usually referred to as density matrix in analogy with its classical counterpart) has eigenvalues λ_k that are positive and sum up to 1, the decomposition of the normal state ρ takes the form $\sigma = \sum_k \lambda_k \Pi_k$, where Π_k is the projection operator onto the k th eigenstate (counting multiplicity).

It is also convenient to know that if a sequence of normal states σ_k on $\mathcal{B}(\mathcal{H})$ converges weakly (i.e., for each $A \in \mathcal{B}(\mathcal{H})$ the sequence $\sigma_k(A)$ converges) then the limit state is normal. This useful result is false in general for closed subalgebras of $\mathcal{B}(\mathcal{H})$, for example, for algebras that contain no minimal projections.

Note that no pure state is dispersion free with respect to all the observables (contrary to what happens in classical mechanics). Recall that the dispersion of the state ρ_σ with respect to the observable A is defined as $\Delta_\sigma(A) \equiv \sigma(A^2) - (\sigma(A))^2$.

The connection of the state with the outcome of a single measurement of an observable associated with an operator A is given by the following axiom, which we shall formulate only for the case when the self-adjoint operator A has only discrete spectrum. The generalization to the other case is straightforward but requires the use of the spectral projections of A .

Axiom III

- (i) If A has only discrete spectrum, the possible outcomes of a measurement of A are its eigenvalues $\{a_k\}$.
- (ii) If the state of the system immediately before the measurement is represented by the vector $\phi \in \mathcal{H}$, the probability that the outcome be a_k is $\sum_b |\langle \psi, \phi_b^{A;k} \rangle|^2$, where $\phi_b^{A;k}$ are a complete orthonormal set in the Hilbert space spanned by the eigenvectors of A to the eigenvalue a_k .
- (iii) If a system is in the pure state ϕ and one performs a measurement of the observable A with outcome $a_j \in (b - \delta, b + \delta)$ for some

$b, \delta \in \mathbb{R}$ then immediately after the measurement the system can be in any (not necessarily pure) state which lies in the convex hull of the pure states which are in the spectral subspace of the operator A in the interval $\Delta_{b,\delta} \equiv (b - \delta, b + \delta)$.

Note Statements (ii) and (iii) can be extended without modification to the case in which the initial state is not a pure state, and is represented by a density matrix σ .

Remark 1 Axiom III makes sure that if one performs, immediately after the first, a further measurement of the same observable A the outcome will still lie in the interval $\Delta_{b,\delta}$. This is needed to give some objectivity to the statement made about the outcome; notice that one must place the condition “immediately after” because the evolution may not leave invariant the spectral subspaces of A . If the operator A has, in the interval $\Delta_{b,\delta}$, only discrete (pure point) spectrum, one can express Axiom III in the following way: the outcome can be any state that can be represented by a convex affine superposition of the eigenstates of A with eigenvalues contained in $\Delta_{b,\delta}$.

In the very special case when A has only one eigenvalue in $\Delta_{b,\delta}$ and this eigenvalue is not degenerate, one can state Axiom III in the following form (commonly referred to as “reduction of the wave packet”): the system after the measurement is pure and is represented by an eigenstate of the operator A .

Remark 2 Notice that the third axiom makes a statement about the state of the system after the measurement is completed.

It follows from Axiom III that one can measure “simultaneously” only observables which are represented by self-adjoint operators that commute with each other (i.e., their spectral projections mutually commute). It follows from the spectral representation of the self-adjoint operators that a family $\{A_k\}$ of commuting operators can be considered (i.e., there is a representation in which they are) functions over a common measure space.

Axioms I–III give a mathematically consistent formulation of QM and allow a statistical description (and statistical prediction) of the outcome of the measurement of any observable. It is worth remarking that while the predictions will have only a statistical nature, the dynamical evolution of the observables (and by duality of the states) will be described by deterministic laws. The intrinsically statistical aspect of the predictions comes only from

the third postulate, which connects the mathematical content of the theory with the measurement process.

The third axiom, while crucial for the connection of the mathematical formalism with the experimental data, contains the seed of the conceptual difficulties which plague QM and have not been cured so far.

Indeed, the third axiom indicates that the process of measurement is described by laws that are intrinsically different from the laws that rule the evolution without measurement. This privileged role of the changing by effect of a measurement leads to serious conceptual difficulties since the changing is independent of whether or not the result is recorded by some observer; one should therefore have a way to distinguish between measurements and generic interactions with the environment.

A related problem that is originated by Axiom III is that the formulation of this axiom refers implicitly to the presence of a classical observer that certifies the outcomes of measurements and is allowed to make use of classical probability theory. This observer is not subjected therefore to the laws of QM.

These two aspects of the conceptual difficulties have their common origin in the separation of the measuring device and of the measured systems into disjoint entities satisfying different laws. The difficulties in the theory of measurement have not yet received a satisfactory answer, but various attempts have been made, with various degree of success, and some of them are described briefly in the section “**Interpretation problems.**” It appears therefore that QM in its present formulation is a refined and successful instrument for the description of the nonrelativistic phenomena at the Planck scale, but its internal consistency is still standing on shaky ground.

Returning to the axioms, it is worth remarking explicitly that according to Axiom II a state is a linear functional over the observables, but it is represented by a sesquilinear function on the complex Hilbert space \mathcal{H} . Since Axiom II states that any normalized element of \mathcal{H} represents a state (and elements that differ only by a phase represent the same state) together with ϕ, γ also $\xi \equiv a\phi + b\psi$, $|a|^2 + |b|^2 = 1$ represent a state superposition of ϕ and ψ (superposition principle).

But for an observable A , one has in general $\rho_\xi(A) \neq |a|^2 \rho_\phi(A) + |b|^2 \rho_\psi(A)$, due to the cross-terms in the scalar product. The superposition principle is one of the characteristic features of QM. The superposition of the two pure states ϕ and ψ has properties completely different from those of a

statistical mixture of the same two states, defined by the density matrix $\sigma = |a|^2 \Pi_\phi + |b|^2 \Pi_\psi$, where we have denoted by Π_ϕ the orthogonal projection onto the normalized vector ϕ . Therefore, the search for these interference terms is one of the means to verify the predictions of QM, and their smallness under given conditions is a sign of quasiclassical behavior of the system under study.

Strictly connected to superposition are entanglement and the partial trace operation. Suppose that one has two systems which when considered separately are described by vectors in two Hilbert spaces $\mathcal{H}_i, i = 1, 2$, and which have observables $A_i \in \mathcal{B}(\mathcal{H}_i)$. When we want to study their mutual interaction, it is natural to describe both of them in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and to consider the observables $A_1 \otimes I$ and $I \otimes A_2$.

When the systems interact, the interaction will not in general commute with the projection operator Π_1 onto \mathcal{H}_1 . Therefore, even if the initial state is of the form $\phi_1 \otimes \phi_2, \phi_i \in \mathcal{H}_i$, the final state (after the interaction) is a vector $\xi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ which cannot be written as $\xi = \zeta_1 \otimes \zeta_2$ with $\zeta_i \in \mathcal{H}_i$. It can be shown, however, that there always exist two orthonormal family vectors $\phi_n \in \mathcal{H}_1$ and $\psi_n \in \mathcal{H}_2$ such that $\xi = \sum c_n \phi_n \otimes \psi_n$ for suitable $c_n \in \mathbb{C}$, $\sum |c_n|^2 = 1$ (this decomposition is not unique in general).

Recalling that $\rho_{\phi \otimes \psi}(A_1 \otimes I) = \rho_\phi(A_1)$, one can write

$$\begin{aligned} \rho_\xi(A_1 \otimes I) &= \sum |c_n|^2 \rho_{\zeta_n}(A_1) = \rho_\sigma(A_1) \\ \sigma &\equiv \sum_n |c_n|^2 \Pi_{\phi_n} \end{aligned}$$

The map $\Gamma_2: \rho_\xi \rightarrow \rho_{\sigma_1}$ is called reduction or also conditioning) with respect to \mathcal{H}_2 ; it is also called “partial trace” with respect to \mathcal{H}_2 . The first notation reflects the analogy with conditioning in classical probability theory.

The map Γ_2 can be extended by linearity to a map from normal states (density matrices) on $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ to normal states on $\mathcal{B}(\mathcal{H}_1)$ and gives rise to a positivity-preserving and trace-preserving map.

One can in fact prove (Takesaki 1971) that any conditioning for normal states of a von Neumann algebra \mathcal{M} is completely positive in the sense that it remains positive after tensorization of \mathcal{M} with $\mathcal{B}(\mathcal{K})$, where \mathcal{K} is an arbitrary Hilbert space.

It can also be proved that a partial converse is true, that is, that every completely positive trace-preserving map Φ on normal states of a von Neumann algebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ can be written, for a suitable choice of a larger Hilbert space \mathcal{K} and partial isometries V_k , in the form (Kraus form) $\Phi(a) = \sum_k V_k^* a V_k$.

But it must be remarked that, if $U(t)$ is a one-parameter group of unitary operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$ and σ is a density matrix, the one-parameter family of maps $\Gamma(t) \equiv \sigma \rightarrow \Gamma_2(U(t)\sigma U^*(t))$ does not, in general, have the semigroup property $\Gamma(t+s) = \Gamma(t) \cdot \Gamma(s)$, $s, t > 0$ and therefore there is in general no generator (of a reduced dynamics) associated with it. Only in special cases and under very strong hypothesis and approximations is there a reduced dynamics given by a semigroup (Markov property).

Since entanglement and (nontrivial) conditioning are marks of QM, and on the other side the Markov property described above is typical of conditioning in classical mechanics, it is natural to search for conditions and approximations under which the Markov property is recovered, and more generally under which the coherence properties characteristic of QM are suppressed (decoherence). We shall discuss briefly this problem in the section “**Interpretation problems,**” devoted to the attempts to overcome the serious conceptual difficulties that descend from Axiom III.

It is seen from the remarks and definitions above that normal states (density matrices) play the role that in classical mechanics is attributed to measures over phase space, with the exception that pure states in QM do not correspond to Dirac measures (later on we shall discuss the possibility of describing a quantum-mechanical states with a function (Wigner function) on phase space).

In this correspondence, evaluation of an observable (a measurable function over phase space) over a state (a normalized, positive measure) is related to finding the (Hilbert space) trace of the product of an operator in $\mathcal{B}(\mathcal{H})$ with a density matrix. Notice that the trace operation shares some of the properties of the integral, in particular $\text{tr} AB = \text{tr} BA$ if A is in trace class and $B \in \mathcal{B}(\mathcal{H})$ (cf. $g \in L^1$ and $f \in L^\infty$) and $\text{tr} AB > 0$ if A is a density matrix and B is a positive operator. This suggests to define functions over the density matrices that correspond to quantities which are important in the theory of dynamical systems, in particular the entropy.

This is readily done if the Hilbert space is finite dimensional, and in the infinite-dimensional case if one takes as observables all Hermitian bounded operators. In quantum statistical mechanics one is led to consider an infinite collection of subsystems, each one described with a Hilbert space (finite or infinite dimensional) $\mathcal{H}_i, i = 1, 2, \dots$, the space of representation is a subspace \mathcal{K} of $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \dots$, and the observables are a (weakly closed) subalgebra \mathcal{A} of $\mathcal{B}(\mathcal{K})$ (typically constructed as an inductive limit of elements of the form $I \otimes I \cdots \otimes A_k \otimes I \cdots$). In this context one also considers normal states on \mathcal{A} and defines a trace operation, with the properties

described above for a trace. Most of the definitions (e.g., of entropy) can be given in this enlarged context, but differences may occur, since in general \mathcal{A} does not contain finite-dimensional projections, and therefore the trace function is not the trace commonly defined in a Hilbert space. We shall not describe further this very interesting and much developed theory, of major relevance in quantum statistical mechanics. For a thorough presentation see [Ohya and Petz \(1993\)](#).

The simplest and most-studied example is the case when each Hilbert space \mathcal{H}_i is a complex two-dimensional space. The resulting system is constructed in analogy with the Ising model of classical statistical mechanics, but in contrast to that system it possesses, for each value of the index i , infinitely many pure states. The corresponding algebra of observables is a closed subalgebra of $(C^2 \times C^2)^{\otimes Z}$ and generically does not contain any finite-dimensional projection.

This model, restricted to the case $(C^2 \times C^2)^K$, K a finite integer, has become popular in the study of quantum information and quantum computation, in which case a normalized element of \mathcal{H}_i is called a q-bit (in analogy with the bits of information in classical information theory). It is clear that the unit sphere in $(C^2 \times C^2)$ contains many more than four points, and this gives much more freedom for operations on the system. This is the basis of quantum computation and quantum information, a very interesting field which has received much attention in recent years.

Quantization and Dynamics

The evolution in nonrelativistic QM is described by the Schrödinger equation in the representation in which for an N -particle system the Hilbert space is $L^2(\mathbb{R}^{3N} \otimes C^k)$, where C^k is a finite-dimensional space which accounts for the fact that some of the particles may have a spin content.

Apart from (often) inessential parameters, the Schrödinger equation for spin-0 particles can be written typically as

$$\begin{aligned} i\hbar \frac{\partial \phi}{\partial t} &= H\phi \\ H &\equiv \sum_{k=1}^N m_k (i\hbar \nabla_k + A_k)^2 \\ &\quad + \sum_{k=1}^N V_k(x_k) + \sum_{i \neq k, 1}^N V_{i,k}(x_i - x_k) \quad [9] \end{aligned}$$

where \hbar is Planck's constant, A_k are vector-valued functions (vector potentials), and V_k and $V_{i,k}$ are scalar-valued function (scalar potentials) on \mathbb{R}^3 .

If some particles have of spin 1/2, the corresponding kinetic energy term should read $-(i\hbar\sigma \cdot \nabla)^2$, where σ_k , $k = 1, 2, 3$, are the Pauli matrices and one must add a term $W(x)$ which is a matrix field with values in $C^k \otimes C^k$ and takes into account the coupling between the spin degrees of freedom. Notice that the local operator $i\sigma \cdot \nabla$ is a “square root” of the Laplacian.

A relativistic extension of the Schrödinger equation for a free particle of mass $m \geq 0$ in dimension 3 was obtained by Dirac in a space of spinor-valued functions $\psi_k(x, t)$, $k = 0, 1, 2, 3$, which carries an irreducible representation of the Lorentz group. In analogy with the electromagnetic field, for which a linear partial differential equation (PDE) can be written using a four-dimensional representation of the Lorentz group, the relativistic Dirac equation is the linear PDE

$$i \sum_{k=0}^3 \gamma_k \frac{\partial}{\partial x_k} \psi = m\psi, \quad x_0 \equiv ct$$

where the γ_k generate the algebra of a representation of the Lorentz group. The operator $\sum (\partial/\partial x_k)\gamma_k$ is a local square root of the relativistically invariant d'Alembert operator $-\partial^2/\partial x_0^2 + \Delta - m \cdot I$.

When one tries to introduce (relativistically invariant) local interactions, one faces the same problem as in the classical mechanics, namely one must introduce relativistically covariant fields (e.g., the electromagnetic field), that is, systems with an infinite number of degrees of freedom. If this field is considered as external, one faces technical problems, which can be overcome in favorable cases. But if one tries to obtain a fully quantized theory (by also quantizing the field) the obstacles become unsurmountable, due also to the nonuniqueness of the representation of the canonical commutation relations if these are taken as the basis of quantization, as in the finite-dimensional case.

In a favorable case (e.g., the interaction of a quantum particle with the quantized electromagnetic field) one can set up a perturbation scheme in a parameter α (the physical value of α in natural units is roughly 1/137). We shall come back later to perturbation schemes in the context of the Schrödinger operator; in the present case one has been able to find procedures (renormalization) by which the series in α that describe relevant physical quantities are well defined term by term. But even in this favorable case, where the sum of the first few terms of the series is in excellent agreement with the experimental data, one has reasons to believe that the series is not convergent, and one does not even know whether the series is asymptotic.

One is led to wonder whether the structure of fields (operator-valued elements in the dual of compactly supported smooth functions on classical spacetime), taken over in a simple way from the field structure of classical electromagnetism, is a valid instrument in the description of phenomena that take place at a scale incomparably smaller than the scale (atomic scale) at which we have reasons to believe that the formalisms of Schrödinger and Heisenberg provide a suitable model for the description of natural phenomena.

The phenomena which are related to the interaction of a quantum nonrelativistic particle interacting with the quantized electromagnetic field take place at the atomic scale. These phenomena have been the subject of very intense research in theoretical physics, mostly within perturbation theory, and the analysis to the first few orders has led to very spectacular results (although there is at present no proof that the perturbation series are at least asymptotic).

In this field rigorous results are scarce, but recently some progress has been made, establishing, among other things, the existence of the ground state (a nontrivial result, because there is no gap separating the ground-state energy from the continuous part of the spectrum) and paving the way for the description of scattering phenomena; the latter result is again nontrivial because the photon field may lead to an anomalous infrared (long-range) behavior, much in the same way that the long-range Coulomb interaction requires a special treatment in nonrelativistic scattering theory.

This contribution to the Encyclopedia is meant to be an introduction to QM and therefore we shall limit ourselves to the basic structure of nonrelativistic theory, which deals with systems of a finite number of particles interacting among themselves and with external (classical) potential fields, leaving for more specialized contributions a discussion of more advanced items in QM and of the successes and failures of a relativistically invariant theory of interaction between quantum particles and quantized fields.

We shall return therefore to basics.

One may begin a section on dynamics in QM by discussing some properties of the solutions of the Schrödinger equation, in particular dispersive effects and the related scattering theory, the problem of bound states and resonances, the case of time-dependent perturbation and the ionization effect, the binding of atoms and molecules, the Rayleigh scattering, the Hall effect and other effects in nanophysics, the various multiscale and adiabatic limits, and in general all the physical problems that

have been successfully solved by Schrödinger's QM (as well as the very many interesting and unsolved problems).

We will consider briefly these issues and the approximation schemes that have been developed in order to derive explicit estimates for quantities of physical interest. Since there are very many excellent reviews of present-day research in QM (e.g., Araki and Ezawa (2004), Blanchard and Dell'Antonio (2004), Cycon *et al.* (1986), Islop and Sigal (1996), Lieb (1990), Le Bris (2005), Simon (2002), and Schlag (2004)) we refer the reader to the more specialized contributions to this Encyclopedia for a detailed analysis and precise statements about the results.

We prefer to come back first to the foundations of the theory; we shall take the point of view of Heisenberg and start discussing the mapping properties of the algebra of observables and of the states. Since transition probabilities play an important role, we consider only transformations α which are such that, for any pair of pure states ϕ_1 and ϕ_2 , one has $\langle \alpha(\phi_1), \alpha(\phi_2) \rangle = \langle \phi_1, \phi_2 \rangle$. We call these maps Wigner automorphisms.

A result of Wigner (see Weyl (1931)) states that if α is a Wigner automorphism then there exists a unique operator U_α , either unitary or antiunitary, such that $\alpha(P) = U_\alpha^* P U_\alpha$ for all projection operators. If there is a one-parameter group of such automorphisms, the corresponding operators are all unitary (but they need not form a group).

A generalization of this result is due to Kadison. Denoting by $I_{1,+}$ the set of density matrices, a Kadison automorphism β is, by definition, such that for all $\sigma_1, \sigma_2 \in I_{1,+}$ and all $0 < s < 1$ one has $\beta(s\sigma_1 + (1-s)\sigma_2) = s\beta(\sigma_1) + (1-s)\beta(\sigma_2)$. For Kadison automorphisms the same result holds as for Wigner's.

A similar result holds for automorphisms of the observables. Notice that the product of two Hermitian operators is not Hermitian in general, but Hermiticity is preserved under Jordan's product defined as $A \times B \equiv (1/2)[AB + BA]$.

A Segal automorphism is, by definition, an automorphism of the Hermitian operators that preserves the Jordan product structure. A theorem of Segal states that γ is a Segal automorphism if and only if there exist an orthogonal projector E , a unitary operator U in $E\mathcal{H}$, and an antiunitary operator V in $(I - E)\mathcal{H}$ such that $\gamma(A) = W A W^*$, where $W \equiv U \oplus V$.

We can study now in more detail the description of the dynamics in terms of automorphism of Wigner or Kadison type when it refers to states and of Segal type when it refers to observables. We require that the evolution be continuous in suitable

topologies. The strongest result refers to Wigner's case. One can prove that if a one-parameter group of Wigner automorphism α_t is measurable in the weak topology (i.e., $\alpha_t \sigma(A)$ is measurable in t for every choice of A and σ) then it is possible to choose the $U(t)$ provided by Wigner's theorem in such a way that they form a group which is continuous in the strong topology. Similar results are obtained for the cases of Kadison and Segal automorphism, but in both cases one has to assume continuity of α_t in a stronger topology (the strong operator topology in the Segal case, the norm topology in Kadison's). Weak continuity is sufficient if the operator product is preserved (in this case one speaks of automorphisms of the algebra of bounded operators). The existence of the continuous group $U(t)$ defines a Hamiltonian evolution. One has indeed:

Theorem 1 (Stone). *The map $t \rightarrow U(t), t \in \mathbb{R}$ is a weakly continuous representation of \mathbb{R} in the set of unitary operators in a Hilbert space \mathcal{H} if and only if there exists a self-adjoint operator H on (a dense set of) \mathcal{H} such that $U(t) = e^{itH}$ and therefore*

$$\phi \in D(H) \rightarrow i \frac{dU(t)}{dt} \phi = HU(t)\phi \quad [10]$$

The operator H is called generator of the dynamics described by $U(t)$.

Note In Schrödinger's approach the operator described in Stone's theorem is called Hamiltonian, in analogy with the classical case. In the case of one particle of mass m in \mathbb{R}^3 subject to a conservative force with potential energy $V(x)$ it has the following form, in units in which $\hbar = 1$:

$$H = -\frac{1}{2m} \Delta + V(x), \quad \Delta = \sum_k \frac{\partial^2}{\partial x_k^2} \quad [11]$$

If the potential V depends on time, Stone's theorem is not directly applicable but still the spectral properties of the self-adjoint operators H_t and of the Kernel of the group $\tau \rightarrow e^{iH_t \tau}$ are essential to solve the (time-dependent) Schrödinger equation.

The semigroup $t \rightarrow e^{-tH_0}$ is usually a positivity-preserving semigroup of contractions and defines a Markov process; in favorable cases, the same is true of $t \rightarrow e^{-tH}$ (Feynmann-Kac formula).

There is an analogous situation in the general theory of dynamical systems on a von Neumann algebra; in analogy with the case of elliptic operators, one defines as "dissipation" a map Δ on a von Neumann algebra \mathcal{M} which satisfies $\Delta(a^*a) \geq a^* \Delta(a) + \Delta(a^*)a$ for all $a \in \mathcal{M}$. The positive dissipation Δ is called completely positive if it remains positive after tensorization with $\mathcal{B}(\mathcal{K})$ for any

Hilbert space \mathcal{K} . Notice that according to this definition every $*$ -derivation is a completely positive dissipation. For dissipations there is an analog of the theorem of Stinespring, and often bounded dissipation can be written as

$$\Delta(a) = i[b, a] + \sum V_k^* a V_k - \left(\frac{1}{2}\right) \sum \{V_k^* V_k, a\}$$

for $a \in \mathcal{M}$

(the symbols $\{.,.\}$ denote the anticommutator).

In general terms, by quantization is meant the construction of a theory by deforming a commutative algebra of functions on a classical phase X in such a way that the dynamics of the quantum system can be derived from the prescription of deformation, usually by deforming the Poisson brackets if X is a cotangent bundle $T^*\mathcal{M}$ (Halbut 2002, Landsman 2002). We shall discuss only the Weyl quantization (Weyl 1931) that has its roots in Heisenberg's formulation of QM and refers to the case in which the configuration space is R^N , or, with some variant (Floquet–Zak) the N -dimensional torus. We shall add a few remarks on the Wick (anti-Weyl) quantization. More general formulations are needed when one tries to quantize a classical system defined on the cotangent bundle of a generic variety and even more so if it defined on a generic symplectic manifold.

The Weyl quantization is a mathematically accurate rendering of the essential content of the procedure adopted by Born and Heisenberg to construct dynamics by finding operators which play the role of symplectic coordinates.

Consider a system with one degree of freedom. The first naive attempt would be to find operators \hat{q}, \hat{p} that satisfy the relation

$$[\hat{q}, \hat{p}] \subset iI \quad [12]$$

and to construct the Hamiltonian in analogy with the classical case. To play a similar role, the operators \hat{q} and \hat{p} must be self-adjoint and satisfy [12] at least in a weak sense. If both are bounded, [12] implies $e^{-ib\hat{p}}\hat{q}e^{-ib\hat{p}} = \hat{q} + bI$ (the exponential is defined through a convergent series) and therefore the spectrum of \hat{q} is the entire real line, a contradiction. Therefore, that inclusion sign in [12] is strict and we face domain problems, and as a consequence [12] has many inequivalent solutions (“equivalence” here means “unitary equivalence”).

Apart from “pathological” ones, defined on L^2 -spaces over multiple coverings of R , there are inequivalent solutions of [12] which are effectively used in QM.

The most common solution is on the Hilbert space $L^2(R)$ (with Lebesgue measure), with \hat{x} defined as

the essentially self-adjoint operator that acts on the smooth functions with compact support as multiplication by the coordinate x and \hat{p} is defined similarly in Fourier space. This representation can be trivially generalized to construct operators \hat{q}_k and \hat{p}_k in $L^2(R^N)$.

Another frequently used representation of [12] is on $L^2(S^1)$ (and when generalized to N degrees of freedom, on T^N). In this representation, the operator \hat{p} is defined by $c_k \rightarrow kc_k$ on functions $f(\theta) = \sum_{k=-M}^N c_k e^{ik\theta/2\pi}$, $0 \leq M, N < \infty$. In this case the operator \hat{q} is defined as multiplication by the angle coordinate θ . It is easy to check that this representation is inequivalent to the previous one and that [12] is satisfied (as an identity) on the (dense) set of vectors which are in the domain both of $\hat{p}\hat{q}$ and of $\hat{q}\hat{p}$. But notice that the domain of essential self-adjointness of \hat{p} is not left invariant by the action of \hat{q} ($\theta f(\theta)$ is a function on S^1 only if $f(2\pi) = 0$).

We shall denote \hat{p} in this representation by the symbol $\partial/\partial\theta_{\text{per}}$ and refer to it as the Bloch representation. It can be modified by setting the action of \hat{p} as $c_n \rightarrow nc_n + \alpha$, $0 < \alpha < 2\pi$, and this gives rise to the various Bloch–Zak and magnetic representations.

The Bloch representation can be extended to periodic functions on R^1 noticing that $L^2(R) = L^2(S^1) \otimes l^2(N)$; similarly, the Bloch–Zak and the magnetic representation can be extended to $L^2(R^N)$.

The difference between the representations can be seen more clearly if one considers the one-parameter groups of unitary operators generated by the “canonical operators” \hat{q} and \hat{p} . In the Schrödinger representation on $L^2(R)$, these groups satisfy

$$U(a)V(b) = e^{iab}V(b)U(a) \\ U(a) = e^{ia\hat{q}}, \quad V(b) = e^{ib\hat{p}}$$

and therefore, setting $z = a + ib$ and $W(z) \equiv e^{-iab/2}V(b)U(a)$ one has

$$W(z)W(z') = e^{-i\omega(z, z')/2}W(z + z') \\ z \in C, \quad \omega(z, z') = \text{Im}(\bar{z}, z') \quad [13]$$

The unitary operators $W(z)$ are therefore projective representations of the additive group C . This generalizes immediately to the case of N degrees of freedom; the representation is now of the additive group C^N and ω is the standard symplectic form on C^N .

In the Bloch representation, the unitaries $U(a)V(b)U^*(a)V^*(b)$ are not multiples of the identity, and have no particularly simple form. The map $C^N \ni z \rightarrow W(z)$ with the structure [13] is called Weyl system; it plays a major role in QM. The following

theorem has therefore a major importance in the mathematical theory of QM.

Theorem 2 (von Neumann 1965). *There exists only one, modulo unitary equivalence, irreducible representation of the Weil system.*

The proof of this theorem follows a general pattern in the theory of group representations. One introduces an algebra $\mathcal{W}^{(N)}$ of operators

$$W_f \equiv \int f(z) W(z) dz, \quad f \in L^1(C^N)$$

called Weyl algebra.

It is easy to see that $|W_f| = |f|_1$ and that $f \rightarrow W_f$ is a linear isomorphism of algebras if one considers $\mathcal{W}^{(N)}$ with its natural product structure and L^1 as a noncommutative algebra with product structure

$$f * g \equiv \int dz' f(z - z') g(z') \exp \frac{i}{2} \omega(z, z') \quad [14]$$

So far the algebra $\mathcal{W}^{(N)}$ is a concrete algebra of bounded operators on $L^2(R^2)$. But it can also be considered an abstract C^* -algebra which we still denote by $\mathcal{W}^{(N)}$.

It is easy to see that, according to [14], if f_0 is chosen to be a suitable Gaussian, then W_{f_0} is a projection operator which commutes with all the W_f 's. Moreover, $W_f W_g = \phi_{f,g} W_{f*g}$ for a suitable phase factor ϕ . Considering the Gelfand–Neumark–Segal construction for the C^* -algebra $\mathcal{W}^{(N)}$, one finds that these properties lead to a decomposition of any representation in cyclic irreducible equivalent ones, completing the proof of the theorem.

The Weyl system has a representation (equivalent to the Schrödinger one) in the space $L^2(R^N, g)$, where g is Gauss's measure. This allows an extension in which C^N is replaced by an infinite-dimensional Banach space equipped with a Gauss measure (weak distribution (Segal 1965, Gross 1972, Wiener 1938)). Uniqueness fails in this more general setting (uniqueness is strictly connected with the compactness of the unit ball in C^N). Notice that in the Schrödinger representation (and, therefore, in any other representation) the Hamiltonian for the harmonic oscillator defines a positive self-adjoint operator

$$N = \sum_1^N N_k, \quad N_k = -\frac{\partial^2}{\partial x_{k^2}} + x_k^2 - 1$$

The spectrum of each of the commuting operators N_k consists of the positive integers (including 0) and is therefore called number operator for the k th degree of freedom. The operator N_k can be written as $N_k = a_k^* a_k$, where $a_k = (1/\sqrt{2})(x_k + \partial/\partial x_k)$ and a_k^*

is the formal adjoint of a_k in $L^2(R)$. One has $|a_k(N_k + 1)^{-1/2}| < 1$. In the domain of N these operators satisfy the following relations (canonical commutation relations)

$$\begin{aligned} [a_k, a_b^*] &= \delta_{k,b}, & [a_b, a_k] &= 0 \\ [N_k, a_b] &= -a_b \delta_{b,k}, & [N_b, a_k^*] &= a_k^* \delta_{b,k} \end{aligned} \quad [15]$$

In view of the last two relations, the operator a_k is called the annihilation operator (relative to the k th degree of freedom) and its formal adjoint is called the creation operator. The operators a_k have as spectrum the entire complex plane, the operators a_k^* have empty spectrum; the eigenvectors of N_k are the Hermite polynomials in the variable x_k . The eigenvectors of a_k (i.e., the solutions in $L^2(R)$ of the equation $a_k \phi_\lambda = \lambda \phi_\lambda, \lambda \in C$) are called coherent states; they have a major role in the Bargmann–Fock–Segal quantization and in general in the semiclassical limit.

The operators $\{N_k\}$ generate a maximal abelian system and therefore the space $L^2(R^N)$ has a natural representation as the symmetrized subspace of $\oplus_k (C^N)^k$ (Fock representation). In this representation, a natural basis is given by the common eigenvectors $\phi_{\{n_k\}}, k = 1, \dots, N$, of the operators N_k . A generic vector can be written as

$$\psi = \sum_{\{n_k\}} c_{\{n_k\}} \phi_{\{n_k\}}, \quad \sum_{\{n_k\}} |c_{\{n_k\}}|^2 < \infty$$

and therefore can be represented by the sequence $c_{\{n_k\}}$.

Notice that the creation operators do not create particles in R^N but rather act as a shift in the basis of the Hermite polynomials.

It is traditional to denote by $\gamma(L^2(R^N))$ the Fock representation (also called second quantization because for each degree of freedom the wave function is written in the quantized basis of the harmonic oscillator) and to denote by $\Gamma(A)$ the lift of a matrix $A \in B(C^N)$. These notations are especially used if C^N is substituted with a Banach space X . This terminology was introduced by Segal in his work on quantization of the wave equation; it is used ever since, mostly in a perturbative context.

In the theory of quantized fields, the space C^N is substituted with a Banach space, X , of functions. In this setting, “second quantization” (Segal 1965, Nelson 1974) considers the state $\phi_{\{n_k\}}$ as representing a configuration of the system in which there are precisely n_k particles in the k th physical state (this presupposes having chosen a basis in the space of distribution on R^3). There is no problem in doing this (Gross 1972) and one can choose for X a suitable Sobolev space (which one depends on the Gaussian measure given in X) if one wants that the

generalization of the commutation relations [15] be of the form $[a^*(f), a(g)] = \langle f, g \rangle$ with a suitable scalar product $\langle \cdot, \cdot \rangle$ in X . The problem with quantization of relativistic fields is that, in order to ensure locality, one is forced to use a Sobolev space of negative index (depending on the dimension of physical space), and this gives rise to difficulties in the definition of the dynamics for nonlinear vector fields.

One should notice that in the work of Segal (1965), and then in Constructive field theory (Nelson 1974), the Fock representation is placed in a Schrödinger context exhibiting the relevant operators as acting on a space $L^2(X, g)$, where X is a subspace of the space of Schwartz distributions on the physical space of the particles one wants to describe and g is a suitably defined Gauss measure on X .

The Fock representation is related to the Bargmann–Fock–Segal representation (Bargmann 1967), a representation in a space of holomorphic functions on C^N square integrable with respect to a Gaussian measure. For its development, this representation relies on the properties of Toeplitz operators and on Tauberian estimates. It is much used in the study of the semiclassical limit and in the formulation of QM in systems for which the classical version has, for phase space, a manifold which is not a cotangent bundle (e.g., the 2-sphere).

Remark The Fock representation associated with the Weyl system in the infinite-dimensional context can describe only particles obeying Bose–Einstein statistics; indeed, the states are qualified by their particle content for each element of the basis chosen and there is no possibility of identifying each particle in an N -particle state. This is obvious in the finite-dimensional case: the Hermite polynomial of order 2 cannot be seen as “composed” of two polynomials of order 1.

In the infinite-dimensional context, if one wants to treat particles which obey Fermi–Dirac statistics, one must rely on the Pauli exclusion principle (Pauli 1928), which states that two such particles cannot be in the same configuration; to ensure this, the wave function must be antisymmetric under permutation of the particle symbols. It is a matter of fact (and a theorem in relativistic quantum field theory which follows in that theory from covariance, locality and positivity of the energy (Streater and Wightman 1964) that particles with half-integer spin obey the Fermi–Dirac statistics. Therefore, to quantize such systems, one must introduce (commutation) relations different from those of Weyl. Since it must now be that $(a^*)^2 = 0$, due to antisymmetry, it

is reasonable to introduce the following relations (canonical anticommutation relations):

$$\begin{aligned} \{a_k, a_b^*\} &= \delta_{k,b}, & \{a_b, a_k\} &= 0 \\ [N_k, a_b] &= -a_b \delta_{b,k}, & \{A, B\} &\equiv AB - BA \end{aligned} \quad [16]$$

The Hilbert space is now $\otimes^N \mathcal{H}_2$, where \mathcal{H}_2 is a two-dimensional complex Hilbert space. Notice that \mathcal{H}_2 carries an irreducible two-dimensional representation of $sU(2) \equiv o(3)$ (spin representation) so that this quantization associates spin 1/2 and antisymmetry.

The operators in [16] are all bounded (in fact bounded by 1 in norm). The Fock representation is constructed as in the case of Weyl (see Araki (1988)), with n_k equal 0 or 1 for each index k . The infinite-dimensional case is defined in the same way, and leads to inequivalent irreducible representations (Araki 1988); only in one of them is the number operator defined and bounded below. Some of these representations can be given a Schrödinger-like form, with the introduction of a gauge and an integration formalism based on a trace (Gross 1972). This system is much used in quantum statistical mechanics because it deals with bounded operators and can take advantage of strong results in the theory of C^* -algebras. In the finite-dimensional case (and occasionally also in the general case) it is used in quantum information (the space \mathcal{H}_2 is the space of a quantum bit).

Returning to the Weyl system, we now introduce the strictly related Wigner function which plays an important role in the analysis of the semiclassical limit and in the discussion of some scaling limits, in particular the hydrodynamical limit and the Bose–Einstein condensation when $N \rightarrow \infty$.

The Wigner function W_ϕ for a pure state ϕ is a real-valued function on the phase space of the classical system which represents the state faithfully. It is defined as

$$W_\psi(x, \xi) = (2\pi)^{-n} \int_{R^n} e^{-i(\xi, x)} \psi\left(x + \frac{y}{2}\right) \bar{\psi}\left(x - \frac{y}{2}\right) dy$$

The Wigner function is not positive in general (the only exceptions are those Gaussian states that satisfy $\Delta(x) \cdot \Delta(p) \geq \hbar$). But it has the interesting property that its marginals reproduce correctly the Born rule. In fact, one has $\int W_\phi(x, \xi) dx = |\hat{\phi}(\xi)|^2$. If the function $\phi(t, x)$ $x \in R^n$ is a solution of the free Schrödinger equation $i\hbar \partial \phi / \partial t = -\hbar^2 \Delta$ then its Wigner function satisfies the Liouville (transport) equation $\partial W_\phi / \partial t + \xi \cdot \nabla W = 0$.

The Wigner function is strictly linked with the Weyl quantization. This quantization associates with every function $\sigma(p, x)$ in a given regularity

class an operator $\sigma(D, x)$ (the Weyl symbol of the function σ) defined by

$$(\sigma(D, x)f, g) \equiv \int \sigma(\xi, x)W(f, g)(\xi, x) d\xi dx$$

$$W(f, g)(\xi, x) \equiv \int e^{-i(\xi, p)} f\left(x + \frac{p}{2}, x - \frac{p}{2}\right) dp$$

It can be verified that the action of F preserves the Schwartz classes S and S' and is unitary in $L^2(R^{2N})$. Moreover, one has $\sigma(D, x)^* = \bar{\sigma}(D, x)$.

The relation between Weyl's quantization and Wigner functions can be readily seen from the natural duality between bounded operators and pure states:

$$\text{tr}(\hat{A} \hat{\rho}) \equiv \int a(p, q)\rho(p, q) dp dq$$

$$\rho(p, q) = \int e^{i(p, q')} \rho(q', q) dq'$$

We give now a brief discussion of the general structure of a quantization, and apply it to the Weyl quantization. By quantization of a Hamiltonian system we mean a correspondence, parametrized by a small parameter \hbar , between classical observables (real functions on a phase space \mathcal{F}) and quantum observables (self-adjoint operators on a Hilbert space \mathcal{H}) with the property that the corresponding structures coincide in the limit $\hbar \rightarrow 0$ and the difference for $\hbar \neq 0$ can be estimated in a suitable topology.

This last requirement is important for the applications and, from this point of view, Weyl's quantization gives stronger results than the other formalisms of quantization.

We limit our analysis to the case $\mathcal{F} \equiv T^*X$, with $X \equiv R^N$, and we make use of the realization of \mathcal{H} as $L^2(R^N)$.

Let $\{x_i\}$ be Cartesian coordinates in R^N and consider a correspondence $A \rightarrow \hat{A}$ that satisfies the following requirements:

1. $A \leftrightarrow \hat{A}$ is linear;
2. $x_k \leftrightarrow \hat{x}_k$ where \hat{x}_k is multiplication by x_k ;
3. $p_k \leftrightarrow -i\hbar\partial/\partial x_k$;
4. if f is a continuous function in R^N , one has $f(x) \leftrightarrow f(\hat{x})$ and $\hat{f}(p) = (Ff)(\hat{x})$, where F denotes a Fourier transform;
5. $L_\zeta \leftrightarrow \hat{L}_\zeta, \zeta \equiv (\alpha, \beta), \alpha, \beta \in R^N$, where L_ζ is the generator of the translations in phase space in the direction ζ and \hat{L}_ζ is the generator of the one-parameter group $t \rightarrow W(t\zeta)$ associated with ζ by the Weyl system.

Note that (1) and (4) imply (2) and (3) through a limit procedure.

Under the correspondence $A \leftrightarrow \hat{A}$, linear symplectic maps correspond to unitary transformations. This is not in general the case for nonlinear maps.

One can prove that conditions (1)–(5) give a complete characterization of the map $A \leftrightarrow \hat{A}$. Moreover, the correspondence cannot be extended to other functions in phase space. Indeed, one has:

Theorem 3 (van Hove). *Let G be the class of functions C^∞ on R^{2N} which are generators of global symplectic flows. For $g \in G$ let $\Phi_g(t)$ be the corresponding group. There cannot exist for every g a correspondence $g \leftrightarrow \hat{g}$, with \hat{g} self-adjoint, such that $\hat{g}(x, p) = g(\hat{x}, \hat{p})$.*

We described the Weyl quantization as a correspondence between functions in the Schwartz class S and a class of bounded operators. Weyl's quantization can be extended to a much wider class of functions. Operators that can be so constructed are called Fourier integral operators. One uses the notation $\hat{\sigma} \equiv \sigma(D, x)$.

We have the following useful theorems (Robert 1987):

Theorem 4 *Let l_1, \dots, l_K be linear functions on R^N such that $\{l_i, l_k\} = 0$. Let P be a polynomial and let $\sigma(\xi, x) \equiv P[l_1(\xi, x), l_K(\xi, x)]$. Then*

- (i) $\sigma(D, x)$ maps S in $L^2(R^N)$ and self-adjoint;
- (ii) if g is continuous, then $(g(\sigma)(D, x) = g(\sigma(D, x)))$.

One proves that $\sigma(D, x)$ extends to a continuous map $S'(X) \rightarrow S'(X)$ and, moreover,

Theorem 5 (Calderon–Vaillancourt). *If $\sigma_0 \equiv \sum_{|\alpha|+|\beta| \leq 2N+1} |D_\xi^\alpha D_x^\beta \sigma| < \infty$ the norm of the operator $\sigma(D, x)$ is bounded by σ_0 .*

Any operator obtained from a suitable class of functions through Weyl's quantization is called a pseudodifferential operator. If $\sigma(q, p) = P(p)$, where P is a polynomial, $\hat{\sigma}(p, q)$ is a differential operator.

Moreover, if $\sigma(p, x) \in L^2$ then $\sigma(D, x)$ is a Hilbert–Schmidt operator and

$$|\sigma(D, x)|_{\text{HS}} = (2\pi\hbar)^{-n/2} \left[\int |A(z)|^2 dz \right]^{1/2}$$

Pseudodifferential operators turn out to be very important in particular in the quantum theory of molecules (Le Bris 2003), where adiabatic analysis and Peierls substitution rules force the use of pseudodifferential operators.

The next important problem in the theory of quantization is related to dynamics.

Let β be a quantization procedure and let $H(p, q)$ be a classical Hamiltonian on phase space. Let A_t be

the evolution of a classical observable A under the flow defined by H and assume that $\beta(A_t)$ is well defined or all t .

Is there a self-adjoint operator \hat{H} such that $\beta(A_t) = e^{it\hat{H}}\beta(A)e^{-it\hat{H}}$? If so, can one estimate $|\hat{H} - \beta(H)|$? Conversely, if the generator of the quantized flow is, by definition, \hat{H} (as is usually assumed), is it possible to give an estimate of the difference $|\beta(A_t) - (\beta(A))_t|$ for a dense set of $\phi \in \mathcal{H}$, where $A_t \equiv e^{it\hat{H}}Ae^{-it\hat{H}}$, or to estimate $|\hat{A}_t - A_t|_\infty$, where \hat{A}_t is defined by $\beta(\hat{A}_t) = (\beta(A))_t$. Is it possible to write an asymptotic series in \hbar for the differences?

For the Weyl quantization some quantitative results have been obtained if one makes use of the semiclassical observables (Robert 1987). We shall not elaborate further on this point.

For completeness, we briefly mention another quantization procedure which is often used in mathematical physics.

Wick Quantization

This quantization assigns positive operators to positive functions, but does not preserve polynomial relations. It is strictly related to the Bargmann–Fock–Segal representation.

Call coherent state centered in the point (y, η) of phase space the normalized solution of $(i\hat{p} + \hat{x} - i\eta + x)\phi_{y,\eta}(x) = 0$.

Wick's quantization of the classical observable A is by definition the map $A \rightarrow \text{Op}^{\text{W}}(A)$, where

$$\text{Op}^{\text{W}}(A)\psi \equiv (2\pi\hbar)^{-n} \int A(y, \eta)(\psi, \bar{\phi}_{y,\eta})\phi_{y,\eta} \, dy \, d\eta$$

One can prove, either directly or going through Weyl's representation, that

1. if $A \geq 0$ then $\text{Op}_\hbar^{\text{W}}(A) \geq 0$;
2. the Weyl symbol of the operator $\text{Op}_\hbar^{\text{W}}(A)$ is

$$(\pi\hbar)^{-n} \int \int A(y, \eta) e^{-\frac{1}{\hbar}[(x-y)^2 + (\xi-\eta)^2]} \, dy \, d\eta$$

3. for every $A \in O(0)$ one has $\|\text{Op}_\hbar^{\text{W}}(A) - \hat{A}\| = O(\hbar)$.

Wick's quantization associates with every vector $\phi \in \mathcal{H}$ a positive Radon measure μ_ϕ in phase space, called Husimi measure. It is defined by $\int A \, d\mu_\psi = (\text{Op}_\hbar^{\text{W}}(A)\psi, \psi)$, $A \in S(z)$. Wick's quantization is less adapted to the treatment of nonrelativistic particles, in particular Eherenfest's rule does not apply, and the semiclassical propagation theorem has a more complicated formulation. It is very much used for the analysis in Fock space in the theory of quantized relativistic fields, where a special role is assigned to Wick ordering, according to which the polynomials in \hat{x}_\hbar and \hat{p}_\hbar are reordered in terms of creation and

annihilation operators by placing all creation operators to the left.

We now come back to Schrödinger's equation and notice that it can be derived within Heisenberg's formalism and Weyl's quantization scheme from the Hamiltonian of an N -particle system in Hamiltonian mechanics (at least if one neglects spin, which has no classical analog).

Apart from (often) inessential parameters, the Schrödinger equation for N scalar particles in R^3 can be written as

$$i\hbar \frac{\partial \phi}{\partial t} = \sum_{k=1}^N (i\hbar \nabla_k + A_k)^2 \phi + V \phi \equiv H \phi \quad [17]$$

$$\phi \in L^2(R^{3N})$$

where A_k are vector-valued functions (vector potentials) and $V = V_k(x_k) + V_{i,k}(x_i - x_k)$ are scalar-valued function (scalar potentials) on R^3 .

Typical problems in Schrödinger's quantum mechanics are:

1. Self-adjointness of H , existence of bound states (discrete spectrum of the operator), their number and distribution, and, in general, the properties of the spectrum.
2. Existence, completeness, and continuity properties of the wave operators

$$W_\pm \equiv s - \lim_{\mp\infty} e^{itH_0} e^{-itH} \quad [18]$$

and the ensuing existence and properties of the S -matrix and of the scattering cross sections. In [18] H_0 is a suitable reference operator, usually $-\Delta$ (with periodic boundary conditions if the potentials are periodic in space), for which Schrödinger's equation can be somewhat analytically controlled.

3. Existence and property of a semiclassical limit.

In [17] and [18] we have implicitly assumed that H is time independent; very interesting problems arise when H depends on time, in particular if it is periodic or quasiperiodic in time, giving rise to ionization phenomena. In the periodic case, one is helped by Floquet's theory, but even in this case many interesting problems are still unsolved.

If the potentials are sufficiently regular, the spectrum of H consists of an absolutely continuous part (made up of several bands in the space-periodic case) and a discrete part, with few accumulation points.

On the contrary, if $V(x, \omega)$ is a measurable function on some probability space Ω , with a suitable distribution (e.g., Gaussian), the spectrum may have totally different properties almost surely.

For example, in the case $N = 1$ (so that the terms $V_{i,j}$ are absent) in one and two spatial dimensions the spectrum is pure point and dense, with eigenfunctions which decrease at infinity exponentially fast (although not uniformly); as a consequence, the evolution group does not give rise to a dispersive motion. The same is true in three dimensions if the potential is sufficiently strong and the kinetic energy content of the initial state is sufficiently limited. This very interesting behavior is due roughly to the randomness of the “barriers” generated by the potential and is also present, to a large extent, for potentials quasiperiodic in space (Pastur and Figotin 1992).

In these as well as in most problems related to Schrödinger’s equation, a crucial role is taken by the resolvent operator $(H - \lambda I)^{-1}$, where λ is any complex number outside the spectrum of H ; many of the results are obtained when the difference $(H - \lambda I)^{-1} - (H_0 - \lambda I)^{-1}$ is a compact operator.

Problems of type (1) and (2) are of great physical interest, and are of course common with theoretical physics and quantum chemistry (Le Bris 2003), although the instruments of investigation are somewhat different in mathematical physics. The semiclassical limit is often more of theoretical interest, but its analysis has relevance in quantum chemistry and its methods are very useful whenever it is convenient to use multiscale methods, as in the study of molecular spectra.

We start with a brief description of point (3); it provides a valid instrument in the description of quantum-mechanical systems at a scale where it is convenient to use units in which the physical constant \hbar has a very small value ($\hbar \simeq 10^{-27}$ in CGS units). From Heisenberg’s commutation relations, $[\hat{x}, \hat{p}] \subset \hbar I$, it follows that the product of the dispersion (uncertainty) of the position and momentum variables is proportional to \hbar and therefore at least one of these two quantities must have very large values (compared to \hbar). One considers usually the case in which these dispersions have comparable values, which is therefore very small, of the order of magnitude $\hbar^{1/2}$ (but very large as compared with \hbar). In order to make connection with the Hamilton–Jacobi formalism of classical mechanics one can also consider the case in which the dispersion in momentum is of the order \hbar (the WKB method).

The semiclassical limit takes advantage mathematically from the fact that the parameter \hbar is very small in natural units, and performs an asymptotic analysis, in which the terms of “lowest order” are exactly described and the difference is estimated. The problem one faces is that the Schrödinger equation becomes, in the “mathematical limit”

$\hbar \rightarrow 0$, a very singular PDE (the coefficients of the differential terms go to zero in this limit).

Dividing each term of the equation by \hbar (because we do not want to change the scale of time) leads, in the case of one quantum particle in R^3 in potential field $V(x)$ (we treat, for simplicity, only this case), to the equation

$$i \frac{\partial \phi(x, t)}{\partial t} = -\hbar \Delta \phi(x, t) + \hbar^{-1} V(x) \phi(x, t) \quad [19]$$

It is convenient therefore to “rescale” the spatial variables by a factor $\hbar^{1/2}$ (i.e., choose different units) setting $x = \sqrt{\hbar} X$ and look for solutions of [19] which remain regular in the limit $\hbar \rightarrow 0$ as functions of the rescaled variable X . One searches therefore for solutions that on the “physical scale” have support that becomes “vanishingly small” in the limit. It is therefore not surprising that, in the limit, these solutions may describe point particles; the main result of semiclassical analysis is that the coordinates of these particles obey Hamilton’s laws of classical mechanics.

This can be roughly seen as follows (accurate estimates are needed to make this empirical analysis precise). Using multiscale analysis, one may write the solution in the form $\phi(X, x, t)$ and seek solutions which are smooth in X and x . Both terms on the right-hand side of [19] contain contributions of order -2 and -1 in $\sqrt{\hbar}$ and in order to have regular solutions one must have cancellations between equally singular contributions. For this, one must perform an expansion to the second order of the potential (assumed at least twice differentiable) around a suitable trajectory $q(t)$, $q \in R^3$, and choose this trajectory in such a way that the cancellations take place.

A formal analysis shows that this is achieved only if the trajectory chosen is precisely a solution of the classical Lagrange equations. Of course, a more refined analysis and good estimates are needed to make this argument precise, and to estimate the error that is made when one neglects in the resulting equation terms of order $\sqrt{\hbar}$; in favorable cases, for each chosen T the error in the solution for most initial conditions of the type described is of order $\sqrt{\hbar}$ for $|t| < T$.

This semiclassical result is most easily visualized using the formalism of Wigner functions (the technical details, needed to make into a proof the formal arguments, take advantage of regularity estimates in the theory of functions).

In natural units, one defines

$$W_{\hbar, \rho}(x, \xi, t) = \left(\frac{i}{2\pi} \right)^N W_{\rho} \left(x, \frac{\xi}{\hbar}, t \right)$$

In terms of the Wigner function $W_{\hbar,\rho}$ the Schrödinger equation [19] takes the form

$$\frac{\partial f^{\hbar}}{\partial t} + \xi \cdot \nabla_x f^{\hbar} + K_{\hbar} * f^{\hbar} = 0 \tag{20}$$

$$\rho^{\hbar}(t=0) = \rho_0(\hbar)$$

where

$$K_{\hbar} = \frac{i}{(2\pi)^N} e^{-i\xi \cdot y} \hbar^{-1} \left[V\left(x + \frac{\hbar y}{2}\right) - V\left(x - \frac{\hbar y}{2}\right) \right]$$

It can be proved (Robert 1987) that if the potential is sufficiently regular and if the initial datum converges in a suitable topology to a positive measure f_0 , then, for all times, $W_{\hbar,\rho}(x, t)$ converges to a (weak) solution of the Liouville equation

$$\frac{\partial f}{\partial t} + \xi \cdot \nabla_x f - \nabla V(x) \cdot \nabla_{\xi} f = 0$$

This leads to the semiclassical limit if, for example, one considers a sequence of initial data ρ_{ϕ_n} where ϕ_n is a sequence of functions centered at x_0 with Fourier transform centered at p_0 and dispersion of order $\hbar^{1/2}$ both in position and in momentum. In this case, the limit measure is a Dirac measure centered on the classical paths.

In the course of the proof of the semiclassical limit theorem, one becomes aware of the special status of the Hamiltonians that are at most quadratic in \hat{x} and \hat{p} . Indeed, it is easy to verify that for these Hamiltonians the expectation values of \hat{x} and \hat{p} obey the classical equation of motion (P Ehrenfest rule).

From the point of view of Heisenberg, this can be understood as a consequence of the fact that operators at most bilinear in a and a^* form an algebra \mathcal{D} under commutation and, moreover, the homogeneous part of order 2 is a closed subalgebra such that its action on \mathcal{D} (by commutation) has the same structure as the algebra of generators of the Hamiltonian flow and its tangent flow. Apart from (important) technicalities, the proof of the semiclassical limit theorem reduces to the proof that one can estimate the contribution of the terms of order higher than 2 in the expansion of the quantum Hamiltonian at the classical trajectory as being of order $\hbar^{1/2}$ in a suitable topology (Hepp 1974).

We end this overview by giving a brief analysis of problems (1) and (2), which refer to the description of phenomena that are directly accessible to comparison with experimental data, and therefore have been extensively studied in theoretical physics and quantum chemistry (Mc Weeny 1992); some of them have been analyzed with the instruments of mathematical physics, often with considerable

success. We give here a very naive introduction to these problems and refer the reader to the more specialized contributions to this Encyclopedia for a rigorous analysis and exact statements.

Of course, most of the problems of physical interest are not “exactly solvable,” in the sense that rarely the final result is given explicitly in terms of simple functions. As a consequence, exact numerical results, to be compared with experimental data, are rarely obtained in physically relevant problems, and most often one has to rely on approximation schemes with (in favorable cases) precise estimates on the error.

Formal perturbation theory is the easiest of such schemes, but it seldom gives reliable results to physically interesting problems. One writes

$$H_{\epsilon} \equiv H + \epsilon V \tag{21}$$

where ϵ is a small real parameter, and sets a formal scheme in case (1) by writing

$$H_{\epsilon} \phi_{\epsilon} \equiv E_{\epsilon} \phi_{\epsilon}, \quad E_{\epsilon} \equiv \sum_0^{\infty} \epsilon^k E_k, \quad \phi_{\epsilon} \equiv \sum_0^{\infty} \epsilon^k \phi_k$$

and, in case (2), iterating Duhamel’s formula

$$e^{-itH_{\epsilon}} = e^{-itH_0} + i\epsilon \int_0^t e^{-i(t-s)H_{\epsilon}} V e^{-isH_0} ds \tag{22}$$

Very seldom the perturbation series converges, and one has to resort to more refined procedures.

In some cases, it turns out to be convenient to consider the formal primitive \tilde{E}_{ϵ} of E_{ϵ} (as a differentiable function of ϵ) and prove that it is differentiable in ϵ for $0 < \epsilon < \epsilon_0$ (but not for $\epsilon = 0$). In favorable cases, this procedure may lead to

$$E_{\epsilon} = \sum_0^N \epsilon^k E_k + R_N(\epsilon), \quad \lim_{N \rightarrow \infty} |R_N|(\epsilon) = +\infty$$

with explicit estimates of $|R_N(\epsilon)|$ for $0 \leq \epsilon < \epsilon_0$.

Re-summation techniques of the formal power series may be of help in some cases.

The estimate of the lowest eigenvalues of an operator bounded below is often done by variational analysis, making use of min–max techniques applied to the quadratic form $Q(\phi) \equiv (\phi, H\phi)$.

Semiclassical analysis can be useful to search for the distribution of eigenvalues and in the study of the dynamics of states whose dispersions both in position and in momentum are very large in units in which $\hbar = 1$.

A case of particular interest in molecular and atomic physics occurs when the physical parameters which appear in H_{ϵ} (typically the masses of the particles involved in the process) are such that one

can *a priori* guess the presence of coordinates which have a rapid dependence on time (fast variables) and a complementary set of coordinates whose dependence on time is slow. This suggests that one can try an asymptotic analysis, often in connection with adiabatic techniques. Seldom one deals with cases in which the hypotheses of elementary adiabatic theorems are satisfied, and one has to refine the analysis, mostly through subtle estimates which ensure the existence of quasi invariant subspaces.

Asymptotic techniques and refined estimates are also needed to study the effective description of a system of N interacting identical particles when N becomes very large; for example, in statistical mechanics, one searches for results which are valid when $N \rightarrow \infty$.

The most spectacular results in this direction are the proof of stability of matter by E Lieb and collaborators, and the study of the phenomenon of Bose–Einstein condensation and the related Gross–Pitaevskii (nonlinear Schrödinger) equation. The experimental discovery of the state of matter corresponding to a Bose–Einstein condensate is a clear evidence of the nonclassical behavior of matter even at a comparatively macroscopic size. From the point of view of mathematical physics, the ongoing research in this direction is very challenging.

One should also recognize the increasing role that research in QM is taking in applications, also in connection with the increasing success of nanotechnology. In this respect, from the point of view of mathematical physics, the study of nanostructure (quantum-mechanical systems constrained to very small regions of space or to lower-dimensional manifolds, such as sheets or graphs) is still in its infancy and will require refined mathematical techniques and most likely entirely new ideas.

Finally, one should stress the important role played by numerical analysis (Le Bris 2003) and especially computer simulations. In problems involving very many particles, present-day analytical techniques provide at most qualitative estimates and in favorable cases bounds on the value of the quantities of interest. Approximation schemes are not always applicable and often are not reliable.

Hints for a progress in the mathematical treatment of some relevant physical phenomena of interest in QM (mostly in condensed matter physics) may come from the *ab initio* analysis made by simulations on large computers; this may provide a qualitative and, to a certain extent, quantitative behavior of the solutions of Schrödinger’s equation corresponding to “typical” initial conditions. In recent times the availability of more efficient computing tools has made computer simulation more reliable and more

apt to concur with mathematical investigation to a fuller comprehension of QM.

Interpretation Problems

In this section we describe some of the conceptual problems that plague present-day QM and some of the attempts that have been made to cure these problems, either within its formalism or with an altogether different approach.

Approaches within the QM Formalism

We begin with the approaches “from within.” We have pointed out that the main obstacle in the measurement problem is the description of what occurs during an act of measurement. Axiom III claims that it must be seen as a “destruction” act, and the outcome is to some extent random. The final state of the system is one of the eigenstates of the observable, and the dependence on the initial state is only through an *a priori* probability assignment; the act of measurement is therefore not a causal one, contrary to the (continuous) causal reversible description of the interaction with the environment. One should be able to distinguish *a priori* the acts of measurement from a generic interaction.

There is a further difficulty. Due to the superposition principle, if a system \mathcal{S} on which we want to make a measurement of the property associated with the operator A “interacts” with an instrument \mathcal{I} described by the operator S , the final state ξ of the combined system will be a coherent superposition of tensor product of (normalized) eigenstates of the two systems

$$\xi = \sum_{n,m} c_{n,m} \phi_n^A \otimes \psi_m^S, \quad \sum_{n,m} |c_{n,m}|^2 = 1 \quad [23]$$

Measurement as described by Axiom III of QM claims that once the measurement is over, the measured system is, with probability $\sum_m |c_{n,m}|^2$, in the state ϕ_n^A and the instrument is in a state which carries the information about the final state of the system (after all, what one reads at the end is an indicator of the final state of the instrument).

It is therefore convenient to write ξ in the form

$$\xi = \sum_n d_n \phi_n^A \otimes \zeta_n, \quad \sum_n |d_n|^2 = 1 \quad [24]$$

(this defines ζ_n if the spectrum of A is pure point and nondegenerate). It is seen from [24] that, due to the reduction postulate, we know that the the measured system is in the state $\phi_{n_0}^A$ if a measurement of an observable T with nondegenerate spectrum,

eigenvectors $\{\zeta_n\}$, and eigenvalues $\{z_n\}$ gives the results z_{n_0} .

Along these lines, one does not solve the measurement problem (the outcome is still probabilistic) but at least one can find the reason why the measuring apparatus may be considered “classical.”

It is more convenient to go back to [23] and to assume that one is able to construct the measuring apparatus in such a way that one divides (roughly) its pure (microscopic) states in sets Φ_n (each corresponding to a “macroscopic” state) which are (roughly) in one-to-one correspondence to the eigenstates of A . The sets Φ_n contain a very large number, N_{Φ_n} , of elements, so that the sets Φ_n need not be given with extreme precision. And the sets Φ_n must be in a sense “stable” under small external perturbations.

It is clear from this rough description that the apparatus should contain a large number of small components and still its interaction with the “small” system A should lead to a more or less sudden change of the sets Φ_n .

A concrete model of this mechanism has been proposed by K Hepp (1972) for the case when A is a 2×2 matrix, and the measuring apparatus is made of a chain of N spins, $N \rightarrow \infty$; the analysis was recently completed by Sewell (2005) with an estimate on the error which is made if N is finite but large. This is a dynamical model, in which the observable A (a spin) interacts with a chain of spins (“moves over the spins”) leaving the trace of its passage. It is this trace (final macroscopic state of the apparatus) which is measured and associated with the final state of A . The interaction is not “instantaneous” but may require a very short time, depending on the parameters used to describe the apparatus and the interaction.

We call “decoherence” the weakening of the superposition principle due to the interaction with the environment.

Two different models of decoherence have been analyzed in some detail; we shall denote them thermal-bath model and scattering model; both are dynamical models and both point to a solution, to various extents, of the problem of the reduction to a final density matrix which commutes with the operator A (and therefore to the suppression of the interference terms).

The thermal-bath model makes use of the Heisenberg representation and relies on results of the theory of C^* -algebras. This approach is closely linked with (quantum) statistical mechanics; its aim is to prove, after conditioning with respect to the degrees of freedom of the bath, that a special role emerges for a commuting set of operators of the

measured system, and these are the observables that specify the outcome of the measurement in probabilistic terms.

The scattering approach relies on the Schrödinger approach to QM, and on results from the theory of scattering. This approach describes the interaction of the system \mathcal{S} (typically a heavy particle) with an environment made of a large number of light particles and seeks to describe the state of \mathcal{S} after the interaction when one does not have any information on the final state of the light particle. One seeks to prove that the reduced density matrix is (almost) diagonal in a given representation (typically the one given by the spatial coordinates). This defines the observable (typically, position) that can be measured and the probability of each outcome.

Both approaches rely on the loss of information in the process to cancel the effect of the superposition principle and to bring the measurement problem within the realm of classical probability theory. None of them provides a causal dependence of the result of the measurement on the initial state of the system.

We describe only very briefly these attempts.

In its more basic form, the “scattering approach” has as starting point the Schrödinger equation for a system of two particles, one of which has mass very much smaller than the other one. The heavy particle may be seen as representing the system on which a measurement is being made. The outline of the method of analysis (which in favorable cases can be made rigorous) (Joos and Zeh 1985, Tegmark 1993) is the following. One chooses units in which the mass of the heavy particle is 1, and one denotes by ϵ the mass of the light particle. If x is the coordinate of the heavy particle and y that of the light one, and if the initial state of the system is denoted by $\Phi_0(x, y)$, the solution of the equation for the system is (apart from inessential factors)

$$\Phi_t = \exp\{i(-\Delta_x - \epsilon^{-1}\Delta_y + W(x) + V(x - y))t\}\Phi_0$$

Making use of center-of-mass and relative coordinates, one sees that when ϵ is very small one should be able to describe the system on two timescales, one fast (for the light particle) and one slow (for the heavy one) and, therefore, place oneself in a setting which may allow the use of adiabatic techniques. In this setting, for the measure of the heavy particle (e.g., its position) one may be allowed to consider the light particle in a scattering regime, and use the wave operator corresponding to a potential $V_x(y) \equiv V(y - x)$.

Taking the partial trace with respect to the degrees of freedom of the light particle (this

corresponds to no information of its final state) one finds, at least heuristically, that the state of the heavy particle is now described (due to the trace operation) by a density matrix σ for which in the coordinate representation the off-diagonal terms $\sigma_{x,x'}$ are slightly suppressed by a factor $\xi_{x,x'} = 1 - (W_x^+ \psi, W_{x'}^+ \psi)$ where ψ represents the initial state of the light particle and W_x^+ is the wave operator for the motion of the light particle in the potential ϵV_x . One must assume that function ϕ which represents the initial state of the heavy particle is sufficiently localized so that $\xi_{x,x'} < 1$ for every $x' \neq x$ in its support.

If the environment is made of very many particles (their number $N(\epsilon)$ must be such that $\lim_{\epsilon \rightarrow 0} \epsilon N(\epsilon) = \infty$) and the heavy particle can be supposed to have separate interactions with all of them, the off-diagonal elements of the density matrix tend to 0 as $\epsilon \rightarrow 0$ and the resulting density matrix tends to have the form $\Phi(x, x') = \delta(x - x')$ $\rho(x), \rho(x) \geq 0, \int \rho(x) dx = 1$. If it can be supposed that all interactions take place within a time $T(\epsilon) \leq \epsilon^\alpha$, $\alpha > 0$ one has $\rho(x) = |\psi(x)|^2$.

If the interactions are not independent, the analysis becomes much more involved since it has to be treated by many-body scattering theory; this suggests that the scattering approach can be hardly used in the context of the “thermal-bath model.” In any case, the selection of a “preferred basis” (the coordinate representation) depends on the fact that one is dealing with a scattering phenomenon. A few steps have been made for a rigorous analysis (Teta 2004) but we are very far from a mathematically satisfactory answer.

The thermal-bath approach has been studied within the algebraic formulation of QM and stands on good mathematical ground (Alicki 2002, Blanchard *et al.* 2003, Sewell 2005). Its drawback is that it is difficult to associate the formal scheme with actual physical situations and it is difficult to give a realistic estimate on the decoherence time.

The thermal-bath approach attributes the decoherence effect to the practical impossibility of distinguishing between a vast majority of the pure states of the systems and the corresponding statistical mixtures. In this approach, the observables are represented by self-adjoint elements of a weakly closed subalgebra \mathcal{M} of all bounded operators $\mathcal{B}(\mathcal{H})$ on a Hilbert space \mathcal{H} . This subalgebra may depend on the measuring apparatus (i.e., not all the apparatuses are fit to measure a set of observables). A “classical” observable by definition commutes with all other observables and therefore must belong to the center of \mathcal{A} which is isomorphic to a collection of functions on a probability space \mathcal{M} .

So the appearance of classical properties of a quantum system corresponds to the “emergence” of an algebra with nontrivial center. Since automorphic evolutions of an algebra preserve its center, this program can be achieved only if we admit the loss of quantum coherence, and this requires that the quantum systems we describe are open and interact with the environment, and moreover that the commutative algebra which emerges be stable for time evolution.

It may be shown that one must consider quantum environment in the thermodynamic limit, that is, consider the interaction of the system to be measured with a thermal bath. A discussion of the possible emergence of classical observables and of the corresponding dynamics is given by Gell-Mann (1993). In all these approaches, the commutative subalgebra is selected by the specific form of the interaction; therefore, the measuring apparatus determines the algebra of classical observables.

On the experimental side, a number of very interesting results have been obtained, using very refined techniques; these experiments usually also determine the “decoherence time.” The experimental results, both for the collision model (Hornberger *et al.* 2003) and for the thermal-bath model (Hackermueller *et al.* 2004), are done mostly with fullerene (a molecule which is heavy enough and is not deflected too much after a collision with a particle of the gas). They show a reasonable accordance with the (rough) theoretical conclusions.

The most refined experiments about decoherence are those connected with quantum optics (circularly polarized atoms in superconducting cavities). These are not related to the wave nature of the particles but in a sense to the “wave nature” of a photon as a single unit. The electromagnetic field is now regarded as an incoherent superposition of states with an arbitrarily large number of photons. Polarized photons can be produced one by one, and they retain their individuality and their polarization until each of them interacts with “the environment” (e.g., the boundary of the cavity or a particle of the gas). In a sense, these experimental results refer to a “decoherence by collision” theory.

The experiments by Haroche (2003) prove that coherence may persist for a measurable interval of time and are the most controlled experiments on coherence so far.

Other Approaches

We end this section with a brief discussion of the problem of “hidden variables” and a presentation of an entirely different approach to QM, originated by

D Bohm (1952) and put recently on firm mathematical grounds by Duerr *et al.* (1999). The approach is radically different from the traditional one and it is not clear at present whether it can give a solution to the measurement problem and a description of all the phenomena which traditional QM accounts for. But it is very interesting from the point of view of the mathematics involved.

We have remarked that the formulation of QM that is summarized in the three axioms given earlier has many unsatisfactory aspects, mainly connected with the superposition principle (described in its extremal form by the Schrödinger's cat "paradox") and with the problem of measurement which reveals, for example, through the Einstein–Rosen–Podolski "paradox," an intrinsic nonlocality if one maintains that their "objective" properties can be attributed to systems which are far apart. From the very beginning of QM, attempts have been made to attribute these features to the presence of "hidden variables"; the statistical nature of the predictions of QM is, from this point of view, due to the incompleteness of the parameters used to describe the systems. The impossibility of matching the statistical prediction of QM (confirmed by experimental findings) with a local theory based on hidden variables and classical probability theory has been known for sometime (Kochen and Specker 1967), also through the use of "Bell inequalities" (Bell 1964) among correlations of outcomes of separate measurements performed on entangled system (mainly two photons or two spin-1/2 particles created in a suitable entangled state).

A proof of the intrinsic nonlocality of QM (in the above sense) was given by L Hardy (see Haroche (2003)).

While experimental results prove that one cannot substitute QM with a "naive" theory of hidden variables, more refined attempts may have success. We shall only discuss the approach of Bohm (following a previous attempt by de Broglie) as presented in Duerr *et al.* (1999). It is a dynamical theory in which representative points follow "classical paths" and their motion is governed by a time-dependent vector "velocity" field (in this sense, it is not Newtonian). In a sense, Bohmian mechanics is a minimal completion of QM if one wants to keep the position as primitive observable. To these primitive objects, Bohm's theory adds a complex-valued function ϕ (the "guiding wave" in Bohm's terminology) defined on the configuration space \mathcal{Q} of the particles. In the case of particles with spin, the function ϕ is spinor-valued. Dynamics is given by two equations: one for the coordinates of the particles and one for the guiding wave. If $x \equiv x_1, \dots, x_N$ describes the

configuration of the points, the dynamics in a potential field $V(x)$ is described in the following way: for the wave ϕ by a nonrelativistic Schrödinger equation with potential V and for the coordinates by the ordinary differential equation (ODE)

$$\dot{x}_k = (\hbar/m_k) \text{Im} \left[\frac{\phi^* \nabla_k \phi}{\phi^* \phi} \right] (x), \quad x_k \in R^3$$

where m_k is the mass of the m th particle.

Notice that the vector field is singular at the zeros of the wave function, therefore global existence and uniqueness must be proved. To see why Bohmian mechanics is empirically equivalent to QM, at least for measurement of position, notice that the equation for the points coincides with the continuity equation in QM. It follows that if one has at time zero a collection of points distributed with density $|\phi_0|^2$, the density at time t will be $|\phi(t)|^2$ where $\phi(t)$ is the solution of the Schrödinger equation with initial datum ϕ_0 .

Bohm (1952) formulated the theory as a modification of Newton's laws (and in this form it has been widely used) through the introduction of a "quantum potential" V_Q . This was achieved by writing the wave function in its polar form $\phi = R e^{iS/\hbar}$ and writing the continuity equation as a modified Hamilton–Jacobi equation. The version of Bohm's theory discussed in Duerr *et al.* (1999) introduces only the guiding wave function and the coordinates of the points, and puts the theory on firm mathematical grounds. Through an impressive series of mathematical results, these authors and their collaborators deal with the completeness of the velocity vector field, the asymptotic behavior of the points trajectories (both for the scattering regime and for the trapped trajectories, which are shown to correspond to bound states in QM), with a rigorous analysis of the theorem on the flux across a surface (a cornerstone in scattering theory) and the detailed analysis of the "two-slit" experiment through a study of the interaction with the measuring apparatus. The theory is completely causal, both for the trajectories of the points and for the time development of the pilot wave, and can also accommodate points with spin. It leads to a mathematically precise formulation of the semiclassical limit, and it may also resolve the measurement problem by relating the pilot wave of the entire system to its approximate decomposition in incoherent superposition of pilot wave associated with the particle and to the measuring apparatus (this would be the way to see the "collapse of the wave function" in QM). A weak point of this approach is the relation of the representative points with observable quantities.

Further Reading

- Alicki R (2004) Pure decoherence in quantum systems. *Open Syst. Inf. Dyn.* 11: 53–61.
- Araki H (1988) In: Jorgensen P and Muhly P (eds.) *Operator Algebras and Mathematical Physics*, Contemporary Mathematics 62. Providence, RI: American Mathematical Society.
- Araki H and Ezawa H (eds.) (2004) *Topics in the theory of Schroedinger Operators*. River Edge, NJ: World Scientific.
- Bach V, Froelich J, and Sigal IM (1998) Quantum electrodynamics of constrained non relativistic particles. *Advanced Mathematics* 137: 299–395.
- Bargmann V (1967) On a Hilbert space of analytic functions and an associated integral transform. *Communications of Pure and Applied Mathematics* 20: 1–101.
- Bell J (1966) On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* 38: 4247–4280.
- Blanchard P and Dell’Antonio GF (eds.) (2004) Multiscale methods in quantum mechanics, theory and experiments. *Trends in Mathematics*. Boston: Birkhauser.
- Blanchard P and Olkiewicz R (2003) Decoherence in the Heisenberg representation. *International Journal of Physics B* 18: 501–507.
- Bohm D (1952) A suggested interpretation of quantum theory in terms of “hidden” variables I, II. *Physical Reviews* 85: 161–179, 180–193.
- Bohr N (1913) On the constitution of atoms and molecules. *Philosophical Magazine* 26: 1–25, 476–502, 857–875.
- Bohr N (1918) On the quantum theory of line spectra. *Kongelige Danske Videnskabernes Selskabs Skrifter Series 8, IV, 1, 1–118*.
- Born M (1924) Über quantenmechanik. *Zeitschrift für Physik* 32: 379–395.
- Born M and Jordan P (1925) Zur quantenmechanik. *Zeitschrift für Physik* 34: 858–888.
- Born M, Jordan P, and Heisenberg W (1926) Zur quantenmechanik II. *Zeitschrift für Physik* 35: 587–615.
- Cycon HL, Frese RG, Kirsh W, and Simon B (1987) Schroedinger operators with application to quantum mechanics and geometry. *Texts and Monographs in Physics*. Berlin: Springer Verlag.
- de Broglie L (1923) Ondes et quanta. *Comptes Rendue* 177: 507–510.
- Dell’Antonio GF (2004) On decoherence. *Journal of Mathematical Physics* 44: 4939–4955.
- Dirac PAM (1925) The fundamental equations of quantum mechanics. *Proceedings of the Royal Society of London A* 109: 642–653.
- Dirac PAM (1926) The quantum algebra. *Proceedings of the Cambridge Philosophical Society* 23: 412–428.
- Dirac PAM (1928) The quantum theory of the electron. *Proc. Royal Soc. London A* 117: 610–624, 118: 351–361.
- Duerr D, Golstein S, and Zanghi N (1996) Bohmian mechanics as the foundation of quantum mechanics. *Boston Studies Philosophical Society* 184: 21–44. Dordrecht: Kluwer Academic.
- Einstein A (1905) *The Collected Papers of Albert Einstein*, vol. 2, pp. 347–377, 564–585. Princeton, NJ: Princeton University Press.
- Einstein A (1924–1925) Quantentheorie des einatomigen idealen gases. *Berliner Berichte* (1924) 261–267, (1925) 3–14.
- Gell-Mann M and Hartle JB (1997) Strong decoherence. *Quantum-Classical Correspondence*, pp. 3–35. Cambridge, MA: International Press.
- Gross L (1972) Existence and uniqueness of physical ground state. *Journal of Functional Analysis* 19: 52–109.
- Haroche S (2003) Quantum Information in cavity quantum electrodynamics. *Royal Society of London Philosophical Transactions Serial A Mathematical and Physics Engineering Science* 361: 1339–1347.
- Heisenberg W (1925) Über Quantentheoretische Umdeutung Kinematisches und Mechanischer Beziehungen. *Zeitschrift für Physik* 33: 879–893.
- Heisenberg W (1926) Über quantentheoretische Kinematik und Mechanik. *Mathematisches Annalen* 95: 694–705.
- Hepp K (1974) The classical limit of quantum correlation functions. *Communications in Mathematical Physics* 35: 265–277.
- Hepp K (1975) Results and problem in the irreversible statistical mechanics of open systems. *Lecture Notes in Physics* 39. Berlin: Spriger Verlag.
- Hornberger K and Sype EJ (2003) Collisional decoherence reexamined. *Physical Reviews A* 68: 012105, 1–16.
- Islop P and Sigal S (1996) Introduction to spectral theory with application to Schroedinger operators. *Applied Mathematical Sciences* 113. New York: Springer Verlag.
- Jammer M (1989) *The Conceptual Development of Quantum Mechanics*, 2nd edn. Tomash Publishers, American Institute of Physics.
- Joos E et al. (eds.) (2003) *Quantum Theory and the Appearance of a Classical World*, second edition. Berlin: Springer Verlag.
- Kochen S and Speker EP (1967) The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17: 59–87.
- Le Bris C (2002) Problematiques numeriques pour la simulation moleculaire. *ESAIM Proceedings of the 11th Society on Mathematics and Applied Industries*, pp. 127–190. Paris.
- Le Bris C and Lions PL (2005) From atoms to crystals: a mathematical journey. *Bulletin of the American Mathematical Society (NS)* 42: 291–363.
- Lieb E (1990) From atoms to stars. *Bulletin of the American Mathematical Society* 22: 1–49.
- Mackey GW (1963) *Mathematical Foundations of Quantum Mechanics*. New York–Amsterdam: Benjamin.
- Mc Weeny E (1992) An overview of molecular quantum mechanics. *Methods of Computational Molecular Physics*. New York: Plenum Press.
- Nelson E (1973) Construction of quantum fields from Markoff fields. *Journal of Functional Analysis* 12: 97–112.
- von Neumann J (1996) Mathematical foundation of quantum mechanics. *Princeton Landmarks in Mathematics*. Princeton NJ: Princeton University Press.
- Nielsen M and Chuang I (2000) *Quantum Computation and Quantum Information*. Cambridge, MA: Cambridge University Press.
- Ohya M and Petz D (1993) *Quantum Entropy and Its Use*. Text and Monographs in Physics. Berlin: Springel Verlag.
- Pauli W (1927) Zur Quantenmechanik des magnetische Elektron. *Zeitschrift für Physik* 43: 661–623.
- Pauli W (1928) *Collected Scientific Papers*, vol. 2. 151–160, 198–213, 1073–1096.
- Robert D (1987) *Aoutur de l’approximation semi-classique*. Progress in Mathematics 68. Boston: Birkhauser.
- Schroedinger E (1926) Quantizierung als Eigenwert probleme. *Annalen der Physik* 79: 361–376, 489–527, 80: 437–490, 81: 109–139.
- Segal I (1996) *Quantization, Nonlinear PDE and Operator Algebra*, pp. 175–202. Proceedings of the Symposium on Pure Mathematics 59. Providence, RI: American Mathematical Society.
- Sewell J (2004) Interplay between classical and quantum structure in algebraic quantum theory. *Rend. Circ. Mat. Palermo Suppl.* 73: 127–136.
- Simon B (2000) Schrodinger operators in the twentieth century. *Journal of Mathematical Physics* 41: 3523–3555.

- Streater RF and Wightman AS (1964) *PCT, Spin and Statistics and All That*. New York–Amsterdam: Benjamin.
- Takesaki M (1971) One parameter automorphism groups and states of operator algebras. *Actes du Congrès International des Mathématiciens Nice*, 1970, Tome 2, pp. 427–432. Paris: Gauthier Villars.
- Teta A (2004) On a rigorous proof of the Joos–Zeh formula for decoherence in a two-body problem. *Multiscale Methods in Quantum Mechanics*, pp. 197–205. Trends in Mathematics. Boston: Birkhauser.
- Weyl A (1931) *The Theory of Groups and Quantum Mechanics*. New York: Dover.
- Wiener N (1938) The homogeneous chaos. *American Journal of Mathematics* 60: 897–936.
- Wigner EP (1952) Die Messung quantenmechanischer operatoren. *Zeitschrift fur Physik* 133: 101–108.
- Yafaev DR (1992) *Mathematical scattering theory. Transactions of Mathematical Monographs*. Providence, RI: American Mathematical Society.
- Zee HI (1970) On the interpretation of measurement in quantum theory. *Foundations of Physics* 1: 69–76.
- Zurek WH (1982) Environment induced superselection rules. *Physical Reviews D* 26(3): 1862–1880.

Introductory Article: Topology

Tsou Sheung Tsun, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

This will be an elementary introduction to general topology. We shall not even touch upon algebraic topology, which will be dealt with in Cohomology Theories, although in some mathematics departments it is introduced in an advanced undergraduate course.

We believe such an elementary article is useful for the encyclopaedia, purely for quick reference. Most of the concepts will be familiar to physicists, but usually in a general rather vague sense. This article will provide the rigorous definitions and results whenever they are needed when consulting other articles in the work. To make sure that this is the case, we have in fact experimentally tested the article on physicists for usefulness.

Topology is very often described as “rubber-sheet geometry,” that is, one is allowed to deform objects without actually breaking them. This is the all-important concept of continuity, which underlies most of what we shall study here.

We shall give full definitions, state theorems rigorously, but shall not give any detailed proofs. On the other hand, we shall cite many examples, with a view to applications to mathematical physics, taking for granted that familiar more advanced concepts there need not be defined. By the same token, the choice of topics will also be so dictated.

"1,5,1,0,0pc,0pc,0pc,0pc>Essential Concepts

Definition 1 Let X be a set. A collection \mathcal{T} of subsets of X is called a *topology* if the following are satisfied:

- (i) $\emptyset, X \in \mathcal{T}$.
- (ii) Let \mathcal{I} be an index set. then

$$A_\alpha \in \mathcal{T}, \alpha \in \mathcal{I} \implies \bigcup_{\alpha \in \mathcal{I}} A_\alpha \in \mathcal{T}$$

- (iii) $A_i \in \mathcal{T}, i = 1, \dots, n \implies \bigcap_{i=1}^n A_i \in \mathcal{T}$.

Definition 2 A member of the topology \mathcal{T} is called an *open set* (of X with topology \mathcal{T}).

Remark The last two properties are more easily put as *arbitrary* unions of open sets are open, and *finite* intersections of open sets are open. One can easily see the significance of this: if we take the “usual topology” (which will be defined in due course) of the real line, then the intersection of all open intervals $(-1/n, 1/n)$, n a positive integer, is just the single point $\{0\}$, which is manifestly not open in the usual sense.

Example If we postulate that \emptyset , and the entire set X , are the only open subsets, we get what is called the *indiscrete* or *coarsest* topology. At the other extreme, if we postulate that all subsets are open, then we get the *discrete* or *finest* topology. Both seem quite unnatural if we think in terms of the real line or plane, but in fact it would be more unnatural to explicitly exclude them from the definition. They prove to be quite useful in certain respects.

Definition 3 A subset of X is *closed* if its complement in X is open.

Remarks

- (i) One could easily build a topology using closed sets instead of open sets, because of the simple relation that the complement of a union is the intersection of the complements.
- (ii) From the definitions, there is nothing to prevent a set being both open and closed, or neither

Definition 4 A set equipped with a topology is called a *topological space* (with respect to the given topology). Elements of a topological space are sometimes called *points*.

Definition 5 Let $x \in X$. A *neighborhood* of x is a subset of X containing an open set which contains x .

Remark This seems a clumsy definition, but turns out to be more useful in the general case than restricting to open neighborhoods, which is often done.

Definition 6 A subcollection of open sets $\mathcal{B} \subseteq \mathcal{T}$ is called a *basis* for the topology \mathcal{T} if every open set is a union of sets of \mathcal{B} .

Definition 7 A subcollection of open sets $\mathcal{S} \subseteq \mathcal{T}$ is called a *sub-basis* for the topology \mathcal{T} if every open set is a union of finite intersections of sets of \mathcal{S} .

Definition 8 The *closure* \bar{A} of a subset A of X is the smallest closed set containing A .

Definition 9 The *interior* $\overset{\circ}{A}$ of a subset A of X is the largest open set contained in A .

Remark It is sometimes useful to define the *boundary* of A as the set $\bar{A} \setminus \overset{\circ}{A} = \{x \in \bar{A}, x \notin \overset{\circ}{A}\}$.

Definition 10 Let A be a subset of a topological space X . A point $x \in X$ is called a *limit point* of A if every open set containing x contains some point of A other than x .

Definition 11 A subset A of X is said to be *dense* in X if $\bar{A} = X$.

Definition 12 A topological space X is called a *Hausdorff space* if for any two distinct points $x, y \in X$, there exist an open neighborhood A of x and an open neighborhood B of y such that A and B are disjoint (that is, $A \cap B = \emptyset$).

Remark and Examples

- (i) This is looking more like what we expect. However, certain mildly non-Hausdorff spaces turn out to be quite useful, for example, in twistor theory. A “pocket” furnishes such an example. Explicitly, consider X to be the subset of the real plane consisting of the interval $[-1, 1]$ on the x -axis, together with the interval $[0, 1]$ on the line $y = 1$, where the following pairs of points are identified: $(x, 0) \cong (x, 1)$, $0 < x \leq 1$. Then the two points $(0, 0)$ and $(0, 1)$ do not have any disjoint neighborhoods. Strictly speaking, one needs the notion of a quotient topology, introduced below.
- (ii) For a more “truly” non-Hausdorff topology, consider the space of positive integers $\mathbb{N} = \{1, 2, 3, \dots\}$, and take as open sets the following: \emptyset , \mathbb{N} , and the sets $\{1, 2, \dots, n\}$ for each $n \in \mathbb{N}$.

This space is neither Hausdorff nor compact (see later for definition of compactness).

Definition 13 Let X and Y be two topological spaces and let $f: X \rightarrow Y$ be a map from X to Y . We say that f is *continuous* if $f^{-1}(A)$ is open (in X) whenever A is open (in Y).

Remark Continuity is the single most important concept here. In this general setting, it looks a little different from the “ ϵ - δ ” definition, but this latter works only for metric spaces, which we shall come to shortly.

Definition 14 A map $f: X \rightarrow Y$ is a *homeomorphism* if it is a continuous bijective map such that its inverse f^{-1} is also continuous.

Remark Homeomorphisms are the natural maps for topological spaces, in the sense that two homeomorphic spaces are “indistinguishable” from the point of view of topology. Topological invariants are properties of topological spaces which are preserved under homeomorphisms.

Definition 15 Let $B \subseteq A$. Then one can define the *relative topology* of B by saying that a subset $C \subseteq B$ is open if and only if there exists an open set D of A such that $C = D \cap B$.

Definition 16 A subset $B \subseteq A$ equipped with the relative topology is called a *subspace* of the topological space A .

Remark Thus, if for subsets of the real line, we consider $A = [0, 3]$, $B = [0, 2]$, then $C = (1, 2]$ is open in B , in the relative topology induced by the usual topology of \mathbb{R} .

Definition 17 Given two topological spaces X and Y , we can define a *product topological space* $Z = X \times Y$, where the set is the Cartesian product of the two sets X and Y , and sets of the form $A \times B$, where A is open in X and B is open in Y , form a basis for the topology.

Remark Note that the open sets of $X \times Y$ are not always of this product form ($A \times B$).

Definition 18 Suppose there is a partition of X into disjoint subsets $A_\alpha, \alpha \in \mathcal{I}$, for some index set \mathcal{I} , or equivalently, there is defined on X an equivalence relation \sim . Then one can define the *quotient topology* on the set of equivalence classes $\{A_\alpha, \alpha \in \mathcal{I}\}$, usually denoted as the quotient space $X/\sim = Y$, as follows. Consider the map $\pi: X \rightarrow Y$, called the canonical projection, which maps the element $x \in X$ to its equivalence class $[x]$. Then a subset $U \subseteq Y$ is open if and only if $\pi^{-1}(U)$ is open.

Proposition 1 Let \mathcal{T} be the quotient topology on the quotient space Y . Suppose \mathcal{T}' is another

topology on Y such that the canonical projection is continuous, then $T' \subseteq T$.

Definition 19 An (open) cover $\{U_\alpha : \alpha \in \mathcal{I}\}$ for X is a collection of open sets $U_\alpha \subseteq X$ such that their union equals X . A subcover of this cover is then a subset of the collection which is itself a cover for X .

Definition 20 A topological space X is said to be compact if every cover contains a finite subcover.

Remark So for a compact space, however one chooses to cover it, it is always sufficient to use a finite number of open subsets. This is one of the essential differences between an open interval (not compact) and a closed interval (compact). The former is in fact homeomorphic to the entire real line.

Definition 21 A topological space X is said to be connected if it cannot be written as the union of two nonempty disjoint open sets.

Remark A useful equivalent definition is that any continuous map from X to the two-point set $\{0, 1\}$, equipped with the discrete topology, cannot be surjective.

Definition 22 Given two points x, y in a topological space X , a path from x to y is a continuous map $f : [0, 1] \rightarrow X$ such that $f(0) = x, f(1) = y$. We also say that such a path joins x and y .

Definition 23 A topological space X is path-connected if every two points in X can be joined by a path lying entirely in X .

Proposition 2 A path-connected space is connected.

Proposition 3 A connected open subspace of \mathbb{R}^n is path-connected.

Definition 24 Given a topological space X , define an equivalence relation by saying that $x \sim y$ if and only if x and y belong to the same connected subspace of X . Then the equivalence classes are called (connected) components of X .

Examples

- (i) The Lie group $O(3)$ of 3×3 orthogonal matrices has two connected components. The identity connected component is $SO(3)$ and is a subgroup.
- (ii) The proper orthochronous Lorentz transformations of Minkowski space form the identity component of the group of Lorentz transformations.

Metric Spaces

A special class of topological spaces plays an important role: metric spaces.

Definition 25 A metric space is a set X together with a function $d : X \times X \rightarrow \mathbb{R}$ satisfying

- (i) $d(x, y) \geq 0$,
- (ii) $d(x, y) = 0 \Leftrightarrow x = y$,
- (iii) $d(x, z) \leq d(x, y) + d(y, z)$ (“triangle inequality”).

Remarks

- (i) The function d is called the metric, or distance function, between the two points.
- (ii) This concept of metric is what is generally known as “Euclidean” metric in mathematical physics. The distinguishing feature is the positive definiteness (and the triangle inequality). One can, and does, introduce indefinite metrics (for example, the Minkowski metric) with various signatures. But these metrics are not usually used to induce topologies in the spaces concerned.

Definition 26 Given a metric space X and a point $x \in X$, we define the open ball centred at x with radius r (a positive real number) as

$$B_r(x) = \{y \in X : d(x, y) < r\}$$

Given a metric space X , we can immediately define a topology on it by taking all the open balls in X as a basis. We say that this is the topology induced by the given metric. Then we can recover our usual “ ϵ - δ ” definition of continuity.

Proposition 4 Let $f : X \rightarrow Y$ be a map from the metric space X to the metric space Y . Then f is continuous (with respect to the corresponding induced topologies) at $x \in X$ if and only if given any $\epsilon > 0, \exists \delta > 0$ such that $d(x, x') < \delta$ implies $d(f(x), f(x')) < \epsilon$.

Note that we do not bother to give two different symbols to the two metrics, as it is clear which spaces are involved. The proof is easily seen by taking the relevant balls as neighborhoods. Equally easy is the following:

Proposition 5 A metric space is Hausdorff.

Definition 27 A map $f : X \rightarrow Y$ of metric spaces is uniformly continuous if given any $\epsilon > 0$ there exists $\delta > 0$ such that for any $x_1, x_2 \in X, d(x_1, x_2) < \delta$ implies $d(f(x_1), f(x_2)) < \epsilon$.

Remark Note the difference between continuity and uniform continuity: the latter is stronger and requires the same δ for the whole space.

Definition 28 Two metrics d_1 and d_2 defined on X are equivalent if there exist positive constants a and b such that for any two points $x, y \in X$ we have

$$ad_1(x, y) \leq d_2(x, y) \leq bd_1(x, y)$$

Remark This is clearly an equivalence relation. Two equivalent metrics induce the same topology.

Examples

- (i) Given a set X , we can define the discrete metric as follows: $d_0(x, y) = 1$ whenever $x \neq y$. This induces the discrete topology on X . This is quite a convenient way of describing the discrete topology.
- (ii) In \mathbb{R} , the usual metric is $d(x, y) = |x - y|$, and the usual topology is the one induced by this.
- (iii) More generally, in \mathbb{R}^n , we can define a metric for every $p \geq 1$ by

$$d_p(x, y) = \left\{ \sum_{k=1}^n |x_k - y_k|^p \right\}^{1/p}$$

where $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n)$. In particular, for $p = 2$ we have the usual Euclidean metric, but the other cases are also useful. To continue the series, one can define

$$d_\infty = \max_{1 < k < n} \{|x_k - y_k|\}$$

All these metrics induce the same topology on \mathbb{R}^n .

- (iv) In a vector space V , say over the real or the complex field, a function $\|\cdot\| : V \rightarrow \mathbb{R}^+$ is called a *norm* if it satisfies the following axioms:
 - (a) $\|x\| = 0$ if and only if $x = 0$,
 - (b) $\|\alpha x\| = |\alpha| \|x\|$, and
 - (c) $\|x + y\| \leq \|x\| + \|y\|$.

Then it is easy to see that a metric can be defined using the norm

$$d(x, y) = \|x - y\|$$

In many cases, for example, the metrics defined in example (iii) above, one can define the norm of a vector as just the distance of it from the origin. One obvious exception is the discrete metric.

A slightly more general concept is found to be useful for spaces of functions and operators: that of seminorms. A *seminorm* is one which satisfies the last two of the conditions, but not necessarily the first, for a norm, as listed above.

Definition 29 Given a metric space X , a sequence of points $\{x_1, x_2, \dots\}$ is called a *Cauchy sequence* if, given any $\epsilon > 0$, there exists a positive integer N such that for any $k, \ell > N$ we have $d(x_k, x_\ell) < \epsilon$.

Definition 30 Given a sequence of points $\{x_1, x_2, \dots\}$ in a metric space X , a point $x \in X$ is called a *limit* of the sequence if given any $\epsilon > 0$, there exists a positive integer N such that for any $n > N$ we have $d(x, x_n) < \epsilon$. We say that the sequence *converges* to x .

Definition 31 A metric space X is *complete* if every Cauchy sequence in X converges to a limit in it.

Examples

- (i) The closed interval $[0, 1]$ on the real line is complete, whereas the open interval $(0, 1)$ is not. For example, the Cauchy sequence $\{1/n, n = 2, 3, \dots\}$ has no limit in this open interval. (Considered as a sequence on the real line, it has of course the limit point 0.)
- (ii) The spaces \mathbb{R}^n are complete.
- (iii) The Hilbert space ℓ^2 consisting of all sequences of real numbers $\{x_1, x_2, \dots\}$ such that $\sum_1^\infty x_k^2$ converges is complete with respect to the obvious metric which is a generalization to infinite dimension of d_2 above. For arbitrary $p \geq 1$, one can similarly define ℓ^p , which are also complete and are hence Banach spaces.

Remarks Completeness is not a topological invariant. For example, the open interval $(-1, 1)$ and the whole real line are homeomorphic (with respect to the usual topologies) but the former is not complete while the latter is. The homeomorphism can conveniently be given in terms of the trigonometric function tangent.

Definition 32 A subset B of the metric space X is *bounded* if there exists a ball of radius R ($R > 0$) which contains it entirely.

Theorem 1 (Heine–Borel) *Any closed bounded subset of \mathbb{R}^n is compact.*

Remark The converse is also true. We have thus a nice characterization of compact subsets of \mathbb{R}^n as being closed and bounded.

Proposition 6 *Any bounded sequence in \mathbb{R}^n has a convergent subsequence.*

Definition 33 Consider a sequence $\{f_n\}$ of real-valued functions on a subset A (usually an interval) of \mathbb{R} . We say that $\{f_n\}$ *converges pointwise* in A if the sequence of real numbers $\{f_n(x)\}$ converges for every $x \in A$. We can then define a function $f : A \rightarrow \mathbb{R}$ by $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, and write $f_n \rightarrow f$.

Definition 34 A sequence of functions $f_n : A \rightarrow \mathbb{R}, A \subseteq \mathbb{R}$ is said to *converge uniformly* to a function $f : A \rightarrow \mathbb{R}$ if given any $\epsilon > 0$, there exists a positive integer N such that, for all $x, |f_n(x) - f(x)| < \epsilon$ whenever $n > N$.

Theorem 2 *Let $f_n : (a, b) \rightarrow \mathbb{R}$ be a sequence of functions continuous at the point $c \in (a, b)$, and suppose f_n converges uniformly to f on (a, b) . Then f is continuous at c .*

Remark and Example The pointwise limit of continuous functions need not be continuous, as can be shown by the following example: $f_n(x) = x^n, x \in [0, 1]$. We see that the limit function f is not continuous:

$$f(x) = \begin{cases} 0 & x \neq 1 \\ 1 & x = 1 \end{cases}$$

Definition 35 Let X be a metric space. A map $f : X \rightarrow X$ is a *contraction* if there exists $c < 1$ such that $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$.

Theorem 3 (Banach) *If X is a complete metric space and f is a contraction in X , then f has a unique fixed point $x \in X$, that is, $f(x) = x$.*

Some Function and Operator Spaces

The spaces of functions and operators can be equipped with different topologies, given by various concepts of convergence and of norms (or sometimes seminorms), very often with different such concepts for the same space. As we saw earlier, a norm in a vector space gives rise to a metric, and hence to a topology. Similarly with the concept of convergence for sequences of functions and operators, as one then knows what the limit points, and hence closed sets, are.

But before we do that, let us introduce, in a slightly different context, a topology which is in some sense the natural one for the space of continuous maps from one space to another.

Definition 36 Consider a family F of maps from a topological space X to a topological space Y , and define $W(K, U) = \{f : f \in F, f(K) \subseteq U\}$. Then the family of all sets of the form $W(K, U)$ with K compact (in X) and U open (in Y) form a sub-basis for the *compact open topology* for F .

Consider a topological space X and sequences of functions (f_n) on it. Let $D \subseteq X$. We can then define pointwise convergence and uniform convergence exactly as for functions on subsets of the real line.

Definition 37 Let X, D and (f_n) as above.

- (i) The functions f_n *converge pointwise* on D to a function f if the sequence of numbers $f_n(x) \rightarrow f(x), \forall x \in D$.
- (ii) The functions f_n *converge uniformly* on D to a function f if given $\epsilon > 0$, there exists N such that for all $n > N$ we have $|f_n(x) - f(x)| < \epsilon, \forall x \in D$.

Next we consider the Lebesgue spaces L^p , that is, functions f defined on subsets of \mathbb{R}^n , such that $|f(x)|^p$ is Lebesgue integrable, for real numbers $p \geq 1$. To define these spaces, we tacitly

take equivalence classes of functions which are equal almost everywhere (that is, up to a null set), but very often we can take representatives of these classes and just deal with genuine functions instead. Note that of all L^p , only L^2 is a Hilbert space.

Definition 38 In the space L^p , we define its norm by

$$\|f\| = \left(\int |f(x)|^p dx \right)^{1/p}$$

Now we turn to general normed spaces, and operators on them.

Definition 39 Convergence in the norm is also called strong convergence. In other words, a sequence (x_n) in a normed space X is said to *converge strongly* to x if

$$\lim_{n \rightarrow \infty} \|x_n - x\| = 0$$

Definition 40 A sequence (x_n) in a normed space X is said to *converge weakly* to x if

$$\lim_{n \rightarrow \infty} f(x_n) = f(x)$$

for all bounded linear functionals f .

Consider the space $B(X, Y)$ of bounded linear operators T from X to Y . We can make this into a normed space by defining the following norm:

$$\|T\| = \sup_{x \in X, \|x\|=1} \|Tx\|$$

Then we can define three different concepts of convergence on $B(X, Y)$. There are in fact more in current use in functional analysis.

Definition 41 Let X and Y be normed spaces and let (T_n) be a sequence of operators $T_n \in B(X, Y)$.

- (i) (T_n) is *uniformly convergent* if it converges in the norm.
- (ii) (T_n) is *strongly convergent* if $(T_n x)$ converges strongly for every $x \in X$.
- (iii) (T_n) is *weakly convergent* if $(T_n x)$ converges weakly for every $x \in X$.

Remark Clearly we have: uniform convergence \implies strong convergence \implies weak convergence, and the limits are the same in all three cases. However, the converses are in general not true.

Homotopy Groups

The most elementary and obvious property of a topological space X is the number of connected components it has. The next such property, in a certain sense, is the number of holes X has. There

are higher analogues of these, called the *homotopy groups*, which are topological invariants, that is, they are invariant under homeomorphisms. They play important roles in many topological considerations in field theory and other topics of mathematical physics. The articles Topological Defects and Their Homotopy Classification and Electric-Magnetic Duality contain some examples.

Definition 42 Given a topological space X , the zeroth homotopy set, denoted $\pi_0(X)$, is the set of connected components of X . One sometimes writes $\pi_0(X) = 0$ if X is connected.

To define the fundamental group of X , or $\pi_1(X)$, we shall need the concept of closed loops, which we shall find useful in other ways too. For simplicity, we shall consider based loops (that is, loops passing through a fixed point in X). It seems that in most applications, these are the relevant ones. One could consider loops of various smoothness (when X is a manifold), but in view of applications to quantum field theory, we shall consider continuous loops, which are also the ones relevant for topology.

Definition 43 Given a topological space X and a point $x_0 \in X$, a (*closed*) (*based*) *loop* is a continuous function of the parametrized circle to X :

$$\xi : [0, 2\pi] \rightarrow X$$

satisfying $\xi(0) = \xi(2\pi) = x_0$.

Definition 44 Given a connected topological space X and a point $x_0 \in X$, the space of all closed based loops is called the (parametrized based) *loop space* of X , denoted ΩX .

Remarks

- (i) The loop space ΩX inherits the relative compact-open topology from the space of continuous maps from the closed interval $[0, 2\pi]$ to X . It also has a natural base point: the constant function mapping all of $[0, 2\pi]$ to x_0 . Hence it is easy to iterate the construction and define $\Omega^k X$, $k \geq 1$.
- (ii) Here we have chosen to parametrize the circle by $[0, 2\pi]$, as is more natural if we think in terms of the phase angle. We could easily have chosen the unit interval $[0, 1]$ instead. This would perhaps harmonize better with our previous definition of paths and the definitions of homotopies below.

Proposition 7 *The fundamental group of a topological space X , denoted $\pi_1(X)$, consists of classes of closed loops in X which cannot be continuously deformed into one another while preserving the base point.*

Definition 45 A space X is called *simply connected* if $\pi_1(X)$ is trivial.

To define the higher homotopy groups, let us go into a little detail about homotopy.

Definition 46 Given two topological spaces X and Y , and maps

$$p, q : X \rightarrow Y$$

we say that h is a *homotopy* between the maps p, q if

$$h : X \times I \rightarrow Y$$

is a continuous map such that $h(x, 0) = p(x)$, $h(x, 1) = q(x)$, where I is the unit interval $[0, 1]$. In this case, we write $p \simeq q$.

Definition 47 A map $f : X \rightarrow Y$ is a *homotopy equivalence* if there exists a map $g : Y \rightarrow X$ such that $g \circ f \simeq \text{id}_X$ and $f \circ g \simeq \text{id}_Y$.

Remark This is an equivalence relation.

Definition 48 For a topological space X with base point x_0 , we define $\pi_n(X)$, $n \geq 0$ as the set of homotopy equivalence classes of based maps from the n -sphere S^n to X .

Remark This coincides with the previous definitions for π_0 and π_1 .

There is a very nice relation between homotopy classes and loop spaces.

Proposition 8 $\pi_n(X) = \pi_{n-1}(\Omega X) = \cdots = \pi_0(\Omega^n X)$.

Remarks

- (i) When we consider the gauge group G in a Yang–Mills theory, its fundamental group classifies the monopoles that can occur in the theory.
- (ii) For $n \geq 1$, $\pi_n(X)$ is a group, the group action coming from the joining of two loops together to form a new loop. On the other hand, $\pi_0(X)$ in general is not a group. However, when X is a Lie group, then $\pi_0(X)$ inherits a group structure from X , because it can be identified with the quotient group of X by its identity-connected component. For example, the two components of $O(3)$ can be identified with the two elements of the group \mathbb{Z}_2 , the component where the determinant equals 1 corresponding to 0 in \mathbb{Z}_2 and the component where the determinant equals -1 corresponding to 1 in \mathbb{Z}_2 .
- (iii) For $n \geq 2$, the group $\pi_n(X)$ is always abelian.
- (iv) Examples of nonabelian π_1 are the fundamental groups of some Riemann surfaces.
- (v) Since π_1 is not necessarily abelian, much of the direct-sum notation we use for the homotopy

groups should more correctly be written multiplicatively. However, in most literature in mathematical physics, the additive notation seems to be preferred.

Examples

- (i) $\pi_n(X \times Y) = \pi_n(X) + \pi_n(Y)$, $n \geq 1$.
- (ii) For the spheres, we have the following results:

$$\begin{aligned} \pi_i(S^n) &= \begin{cases} 0 & \text{if } i > n \\ \mathbb{Z} & \text{if } i = n \end{cases} \\ \pi_i(S^1) &= 0 & \text{if } i > 1 \\ \pi_{n+1}(S^n) &= \mathbb{Z}_2 & \text{if } n \geq 3 \\ \pi_{n+2}(S^n) &= \mathbb{Z}_2 & \text{if } n \geq 2 \\ \pi_6(S^3) &= \mathbb{Z}_{12} \end{aligned}$$

- (iii) From the theory of sphere bundles, we can deduce:

$$\begin{aligned} \pi_i(S^2) &= \pi_{i-1}(S^1) + \pi_i(S^3) & \text{if } i \geq 2 \\ \pi_i(S^4) &= \pi_{i-1}(S^3) + \pi_i(S^7) & \text{if } i \geq 2 \\ \pi_i(S^8) &= \pi_{i-1}(S^7) + \pi_i(S^{15}) & \text{if } i \geq 2 \end{aligned}$$

and the first of these relations give the following more succinct result:

$$\pi_i(S^3) = \pi_i(S^2) \quad \text{if } i \geq 3$$

- (iv) A result of Serre says that all the homotopy groups of spheres are in fact finite except $\pi_n(S^n)$ and $\pi_{4n-1}(S^{2n})$, $n \geq 1$.

Definition 49 Given a connected space X , a map $\pi : B \rightarrow X$ is called a *covering* if (i) $\pi(B) = X$, and (ii) for each $x \in X$, there exists an open connected neighborhood V of x such that each component of $\pi^{-1}(V)$ is open in B , and π restricted to each component is a homeomorphism. The space B is called a *covering space*.

Examples

- (i) The real line \mathbb{R} is a covering of the group $U(1)$.
- (ii) The group $SU(2)$ is a double cover of the group $SO(3)$.
- (iii) The group $SL(2, \mathbb{C})$ is a double cover of the Lorentz group $SO(1, 3)$.
- (iv) The group $SU(2, 2)$ is a 4-fold cover of the conformal group in four dimensions. This local isomorphism is of great importance in twistor theory.

Remarks

- (i) By considering closed loops in X and their coverings in B it is easily seen that the fundamental group $\pi_1(X)$ acts on the coverings of X . If we further assume that the action is

transitive, then we have the following nice result: coverings of X are in 1–1 correspondence with normal subgroups of $\pi_1(X)$.

- (ii) Given a connected space X , there always exists a unique connected simply connected covering space \tilde{X} , called the universal covering space. Furthermore, \tilde{X} covers all the other covering spaces of X . For the higher homotopy groups, one has

$$\pi_n(X) = \pi_n(\tilde{X}), \quad n \geq 2$$

One very important class of homotopy groups are those of Lie groups. To simplify matters, we shall consider only connected groups, that is, $\pi_0(G) = 0$. Also we shall deal mainly with the classical groups, and in particular, the orthogonal and unitary groups.

Proposition 9 *Suppose that G is a connected Lie group.*

- (i) *If G is compact and semi-simple, then $\pi_1(G)$ is finite. This implies that \tilde{G} is still compact.*
- (ii) $\pi_2(G) = 0$.
- (iii) *For G compact, simple, and nonabelian, $\pi_3(G) = \mathbb{Z}$.*
- (iv) *For G compact, simply connected, and simple, $\pi_4(G) = 0$ or \mathbb{Z}_2 .*

Examples

- (i) $\pi_1(SU(n)) = 0$.
- (ii) $\pi_1(SO(n)) = \mathbb{Z}_2$.
- (iii) Since the unitary groups $U(n)$ are topologically the product of $SU(n)$ with a circle S^1 , their homotopy groups are easily computed using the product formula. We remind ourselves that $U(1)$ is topologically a circle and $SU(2)$ topologically S^3 .
- (iv) For $i \geq 2$, we have:

$$\begin{aligned} \pi_i(SO(3)) &= \pi_i(SU(2)) \\ \pi_i(SO(5)) &= \pi_i(Sp(2)) \\ \pi_i(SO(6)) &= \pi_i(SU(4)) \end{aligned}$$

Just for interest, and to show the richness of the subject, some isomorphisms for homotopy groups are shown in [Table 1](#) and some homotopy groups for low $SU(n)$ and $SO(n)$ are listed in [Table 2](#).

Table 1 Some isomorphisms for homotopy groups

Isomorphism	Range
$\pi_i(SO(n)) \cong \pi_i(SO(m))$	$n, m \geq i + 2$
$\pi_i(SU(n)) \cong \pi_i(SU(m))$	$n, m \geq \frac{1}{2}(i + 1)$
$\pi_i(Sp(n)) \cong \pi_i(Sp(m))$	$n, m \geq \frac{1}{4}(i - 1)$
$\pi_i(G_2) \cong \pi_i(SO(7))$	$2 \leq i \leq 5$
$\pi_i(F_4) \cong \pi_i(SO(9))$	$2 \leq i \leq 6$
$\pi_i(SO(9)) \cong \pi_i(SO(7))$	$i \leq 13$

Table 2 Some homotopy groups for low $SU(n)$ and $SO(n)$

	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
$SU(2)$	\mathbb{Z}_2	\mathbb{Z}_2	\mathbb{Z}_{12}	\mathbb{Z}_2	\mathbb{Z}_2	\mathbb{Z}_3	\mathbb{Z}_{15}
$SU(3)$	0	\mathbb{Z}	\mathbb{Z}_6	0	\mathbb{Z}_{12}	\mathbb{Z}_3	\mathbb{Z}_{30}
$SU(4)$	0	\mathbb{Z}	0	\mathbb{Z}	\mathbb{Z}_{24}	\mathbb{Z}_2	$\mathbb{Z}_{120} + \mathbb{Z}_2$
$SU(5)$	0	\mathbb{Z}	0	\mathbb{Z}	0	\mathbb{Z}	\mathbb{Z}_{120}
$SU(6)$	0	\mathbb{Z}	0	\mathbb{Z}	0	\mathbb{Z}	\mathbb{Z}_3
$SO(5)$	\mathbb{Z}_2	\mathbb{Z}_2	0	\mathbb{Z}	0	0	\mathbb{Z}_{120}
$SO(6)$	0	\mathbb{Z}	0	\mathbb{Z}	\mathbb{Z}_{24}	\mathbb{Z}_2	$\mathbb{Z}_{120} + \mathbb{Z}_2$
$SO(7)$	0	0	0	\mathbb{Z}	$\mathbb{Z}_2 + \mathbb{Z}_2$	$\mathbb{Z}_2 + \mathbb{Z}_2$	\mathbb{Z}_{24}
$SO(8)$	0	0	0	$\mathbb{Z} + \mathbb{Z}$	$\mathbb{Z}_2 + \mathbb{Z}_2 + \mathbb{Z}_2$	$\mathbb{Z}_2 + \mathbb{Z}_2 + \mathbb{Z}_2$	$\mathbb{Z}_{24} + \mathbb{Z}_{24}$
$SO(9)$	0	0	0	\mathbb{Z}	$\mathbb{Z}_2 + \mathbb{Z}_2$	$\mathbb{Z}_2 + \mathbb{Z}_2$	\mathbb{Z}_{24}
$SO(10)$	0	0	0	\mathbb{Z}	\mathbb{Z}_2	$\mathbb{Z} + \mathbb{Z}_2$	\mathbb{Z}_{12}

Appendix: A Mathematician’s Basic Toolkit

The following is a drastically condensed list, most of which is what a mathematics undergraduate learns in the first few weeks. The rest is included for easy reference. These notations and concepts are used universally in mathematical writing. We have not endeavored to arrange the material in a logical order. Furthermore, given structures such as sets, groups, etc., one can usually define “substructures” such as subsets, subgroups, etc., in a straightforward manner. We shall therefore not spell this out.

Sets

- $A \cup B = \{x : x \in A \text{ or } x \in B\}$ union
- $A \cap B = \{x : x \in A \text{ and } x \in B\}$ intersection
- $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$ complement
- $A \times B = \{(x, y) : x \in A, y \in B\}$ Cartesian product

Maps

1. A map or mapping $f : A \rightarrow B$ is an assignment of an element $f(x)$ of B for every $x \in A$.
2. A map $f : A \rightarrow B$ is injective if $f(x) = f(y) \implies x = y$. This is sometimes called a 1–1 map, a term to be avoided.
3. A map $f : A \rightarrow B$ is surjective if for every $y \in B$ there exists an $x \in A$ such that $y = f(x)$. This is sometimes called an “onto” map.
4. A map $f : A \rightarrow B$ is bijective if it is both surjective and injective. This is also sometimes called a 1–1 map, a term to be equally avoided.
5. For any map $f : A \rightarrow B$ and any subset $C \subseteq B$, the inverse image $f^{-1}(C) = \{x : f(x) \in C\} \subseteq A$ is always defined, although, of course, it can be empty. On

the other hand, the map f^{-1} is defined if and only if f is bijective.

6. A map from a set to either the real or complex numbers is usually called a function.
7. A map between vector spaces, and more particularly normed spaces (including Hilbert spaces), is called an operator. Most often, one considers linear operators.
8. An operator from a vector space to its field of scalars is called a functional. Again, one considers almost exclusively linear functionals.

Relations

1. A relation \sim on a set A is a subset $R \subseteq A \times A$. We say that $x \sim y$ if $(x, y) \in R$.
2. We shall only be interested in equivalence relations. An equivalence relation \sim is one satisfying, for all $x, y, z \in A$:
 - (a) $x \sim x$ (“reflexive”),
 - (b) $x \sim y \implies y \sim x$ (“symmetric”),
 - (c) $x \sim y, y \sim z \implies x \sim z$ (“transitive”).
3. If \sim is an equivalence relation in A , then for each $x \in A$, we can define its equivalence class:

$$[x] = \{y \in A : y \sim x\}$$

It can be shown that equivalence classes are nonempty, any two equivalence classes are either equal or disjoint, and they together partition the set A . Subgroup equivalence classes are called cosets.

4. An element of an equivalence class is called a representative.

Groups

A group is a set G with a map, called multiplication or group law

$$G \times G \longrightarrow G$$

$$(x, y) \longmapsto xy$$

satisfying

1. $(xy)z = x(yz), \forall x, y, z \in G$ (“associative”);
2. there exists a neutral element (or identity) 1 such that $1x = x1 = x, \forall x \in G$; and
3. every element $x \in G$ has an inverse x^{-1} , that is, $xx^{-1} = x^{-1}x = 1$.

A map such as the multiplication in the definition is an example of a binary operation. Note that we have denoted the group law as multiplication here. It is usual to denote it additively if the group is abelian, that is, if $xy = yx, \forall x, y \in G$. In this case, we may write the condition as $x + y = y + x$, and call the identity element 0.

Rings

A ring is a set R equipped with two binary operations, $x + y$ called addition, and xy called multiplication, such that

1. R is an abelian group under addition;
2. the multiplication is associative; and
3. $(x + y)z = xz + yz, x(y + z) = xy + xz, \forall x, y, z \in R$ (“distributive”).

If the multiplication is commutative ($xy = yx$) then the ring is said to be commutative. A ring may contain a multiplicative identity, in which case it is called a ring with unit element.

An ideal I of R is a subring of R , satisfying in addition

$$r \in R, a \in I \implies ra \in I, ar \in I$$

One can define in an obvious fashion a left-ideal and a right-ideal. The above definition will then be for a two-sided ideal.

Modules

Given a ring R , an R -module is an abelian group M , together with an operation, $M \times R \rightarrow M$, denoted multiplicatively, satisfying, for $x, y \in M, r, s \in R$,

1. $(x + y)r = xr + yr$,
2. $x(r + s) = xr + xs$,
3. $x(rs) = (xr)s$, and
4. $x1 = x$

The term right R -module is sometimes used, to distinguish it from obviously defined left R -modules.

Fields

A field F is a commutative ring in which every nonzero element is invertible.

The additive identity 0 is never invertible, unless $0 = 1$, so it is usual to assume that a field has at least two elements, 0 and 1.

The most common fields we come across are, of course, the number fields: the rationals, the reals, and the complex numbers.

Vector Spaces

A vector space, or sometimes linear space, V , over a field F , is an abelian group, written additively, with a map $F \times V \rightarrow V$ such that, for $x, y \in V, \alpha, \beta \in F$,

1. $\alpha(x + y) = \alpha x + \alpha y$ (“linearity”),
2. $(\alpha + \beta)x = \alpha x + \beta x$,
3. $(\alpha\beta)x = \alpha(\beta x)$, and
4. $1x = x$.

A vector space is then a right (or left) F -module. The elements of V are called vectors, and those of F scalars.

Algebras

An algebra A over a field F is a ring which is a vector space over F , such that

$$\alpha(ab) = (\alpha a)b = a(\alpha b), \quad \alpha \in F, a, b \in A$$

Note that in some older literature, particularly the Russian school, an algebra of operators is called a ring of operators.

Further Reading

- Borel A (1955) Topology of Lie groups and characteristic classes. *Bulletin American Mathematical Society* 61: 397–432.
- Kelly JL (1955) *General Topology*. New York: Van Nostrand Reinhold.
- Kreyszig E (1978) *Introductory Functional Analysis with Applications*. New York: Wiley.
- Mc Carty G (1967) *Topology: An Introduction with Application to Topological Groups*. New York: McGraw-Hill.
- Simmons GF (1963) *Introduction to Topology and Modern Analysis*. New York: McGraw-Hill.

A

Abelian and Nonabelian Gauge Theories Using Differential Forms

A C Hirshfeld, Universität Dortmund, Dortmund, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Quantum electrodynamics is the theory of the electromagnetic interactions of photons and electrons. When attempting to generalize this theory to other interactions it turns out to be necessary to identify its essential components. The essential properties of electrodynamics are contained in its formulation as an “abelian gauge theory.” The generalization to include other interactions is then reduced to incorporating the structure of nonabelian groups. This becomes particularly clear when we formulate the theory in the language of differential forms.

Here we first present the formulation of electrodynamics using differential forms. The electromagnetic fields are introduced via the Lorentz force equation. They are recognized as the components of a differential 2-form. This form fulfills two differential conditions, which are equivalent to Maxwell’s equations. These are expressed with the help of a differential operator and its Hermitian conjugate, the codifferential operator. We consider the effects of charge conservation and introduce electromagnetic potentials, which are defined up to gauge transformations. We finally consider Weyl’s argument for the existence of the electromagnetic interaction as a consequence of the local phase invariance of the electron wave function.

We then go on to present the nonabelian generalization. The gauge bosons appear in a theory with fermions by requiring invariance of the theory with respect to local gauge transformations. When the fermions group into symmetry multiplets this gives rise to a gauge group $SU(N)$ involving N^2-1 gauge bosons mediating the interaction, where N is the dimension of the Lie algebra. The interaction arises through the necessity of replacing the usual derivatives by covariant derivatives, which transform in a natural way in order to preserve the gauge

invariance. The covariant derivatives involve the gauge potentials, whose transformation properties are dictated by those of the covariant derivative. Whereas for an abelian gauge theory such as electromagnetism scalar-valued p -forms are sufficient (actually only $p=1,2$), a nonabelian gauge theory involves the use of Lie-algebra-valued p -forms. These are introduced and used to construct the Yang–Mills action, which involves the field strength tensor which is determined from the gauge potentials. This action leads to the Yang–Mills equations for the gauge potentials, which are the nonabelian generalizations of the Maxwell equations.

Relativistic Kinematics

The trajectory of a mass point is described as $x^\mu(\tau)$, where τ is the invariant proper time interval:

$$d\tau^2 = dt^2 - dx \cdot dx = dt^2(1 - v^2) \quad [1]$$

with $v = dx/dt$. With the abbreviation $\gamma = (1 - v^2)^{-1/2}$ this yields $d\tau = (1/\gamma)dt$.

The 4-velocity of a point is defined as $u^\mu = dx^\mu/d\tau = \gamma(dx^\mu/dt)$. The quantity

$$u^2 = g_{\mu\nu}u^\mu u^\nu = \frac{dx^\mu dx_\mu}{d\tau^2} = 1 \quad [2]$$

is a relativistic invariant. Here

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad [3]$$

is the metric of Minkowski space.

The 4-momentum of a particle is $p^\mu = m_0 u^\mu = (m_0\gamma, m_0\gamma\mathbf{v})$, and $p^\mu p_\mu = m_0^2$. The 4-force is

$$f^\mu = \frac{dp^\mu}{d\tau} = \gamma \frac{dp^\mu}{dt} = \gamma \left(\frac{dp^0}{dt}, \mathbf{f} \right) \quad [4]$$

with the 3-force

$$\mathbf{f} = \frac{d(m_0\gamma\mathbf{v})}{dt} \quad [5]$$

Differentiate $p^2 = m_0^2$ with respect to τ , this yields

$$2p^\mu f_\mu = 2m_0\gamma^2 \left(\frac{dp^0}{dt} - \mathbf{f} \cdot \mathbf{v} \right) = 0 \quad [6]$$

or

$$\frac{dp^0}{dt} = \mathbf{f} \cdot \mathbf{v} = \mathbf{f} \cdot \frac{d\mathbf{x}}{dt} \quad [7]$$

This says that

$$dp^0 = \mathbf{f} \cdot d\mathbf{x} = dW \quad [8]$$

where W is the work done and p^0 is the energy.

For a charged particle, the Lorentz force is

$$\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad [9]$$

where q is the charge of the particle, \mathbf{E} is the electric, and \mathbf{B} the magnetic field strength. Since $\mathbf{f} \cdot \mathbf{v} = q\mathbf{E} \cdot \mathbf{v}$, we have the four-dimensional form of the Lorentz force:

$$f^\mu = q\gamma(\mathbf{E} \cdot \mathbf{v}, \mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad [10]$$

The Lorentz Force Equation with Differential Forms

We write the Lorentz force equation as an equation for a differential form $f = f_\mu dx^\mu$, with $f_\mu = g_{\mu\nu} f^\nu$. The velocity-dependent Lorentz force is

$$f = -qi_u F \quad [11]$$

with

$$u = \gamma \left(\frac{\partial}{\partial t} + v^x \frac{\partial}{\partial x} + v^y \frac{\partial}{\partial y} + v^z \frac{\partial}{\partial z} \right) \quad [12]$$

the 4-velocity and F the electromagnetic field strength:

$$F = \mathcal{E} \wedge dt + \mathcal{B} \quad [13]$$

where \mathcal{E} is a 1-form in three dimensions,

$$\mathcal{E} = E_x dx + E_y dy + E_z dz \quad [14]$$

and \mathcal{B} is a 2-form in three dimensions,

$$\mathcal{B} = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy \quad [15]$$

The symbol i_u indicates a contraction of a 2-form with a vector, which is defined as

$$i_u F(v) = F(u, v) \quad [16]$$

for an arbitrary vector v . The contraction of a 2-form with a vector yields a 1-form.

It is easily seen that a 2-form can be expressed in terms of a polar vector and an axial vector: if it is to be invariant with respect to parity transformations with

$$t \rightarrow t, \quad x \rightarrow -x, \quad y \rightarrow -y, \quad z \rightarrow -z \quad [17]$$

the fields in eqn [13] must transform as

$$\mathbf{E} \rightarrow -\mathbf{E}, \quad \mathbf{B} \rightarrow \mathbf{B} \quad [18]$$

Now we check the validity of eqn [11]. We have

$$\begin{aligned} f &= -qi_u F \\ &= q\gamma(\mathbf{v} \cdot \mathbf{E})dt - q\gamma[(E^x + (\mathbf{v} \times \mathbf{B})^x)dx \\ &\quad + (E^y + (\mathbf{v} \times \mathbf{B})^y)dy + (E^z + (\mathbf{v} \times \mathbf{B})^z)dz] \end{aligned} \quad [19]$$

in agreement with eqn [10]. We remember to change the signs in $E_x = -E^x$, $B_x = -B^x$, etc.

The Codifferential Operator

The space of p -forms on an n -dimensional manifold is an

$$\binom{n}{p} = \binom{n}{n-p} = \frac{n!}{(n-p)!p!} \quad [20]$$

dimensional vector space. The space of p -forms is thus isomorphic to the space of $(n-p)$ -forms. The Hodge dual operator maps the p -forms into the $(n-p)$ -forms, and is defined by

$$\alpha \wedge * \beta = \langle \alpha, \beta \rangle dx^1 \wedge \cdots \wedge dx^n \quad [21]$$

Here $\langle \alpha, \beta \rangle$ is the scalar product of two p -forms:

$$\langle \alpha, \beta \rangle = \alpha_{i_1 \dots i_p} \beta^{i_1 \dots i_p} \quad [22]$$

where $\alpha_{i_1 \dots i_p}$ are the coefficients of the form α ,

$$\alpha = \alpha_{i_1 \dots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p} \quad [23]$$

$\beta_{j_1 \dots j_p}$ are the coefficients of the form β ,

$$\beta = \beta_{j_1 \dots j_p} dx^{j_1} \wedge \cdots \wedge dx^{j_p} \quad [24]$$

and

$$\beta^{i_1 \dots i_p} = g^{i_1 j_1} \cdots g^{i_p j_p} \beta_{j_1 \dots j_p} \quad [25]$$

The indices satisfy $i_1 < \cdots < i_p$ and $j_1 < \cdots < j_p$.

The basis elements are orthogonal with respect to this scalar product, and

$$\begin{aligned} \langle dx^{i_1} \wedge \cdots \wedge dx^{i_p}, dx^{j_1} \wedge \cdots \wedge dx^{j_p} \rangle \\ = g_{i_1 i_1} \cdots g_{i_p i_p} \end{aligned} \quad [26]$$

The Hodge dual has the property that

$$\begin{aligned} & * \left(dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \right) \\ &= g_{\sigma(1)\sigma(1)} \cdots g_{\sigma(p)\sigma(p)} (\text{sign } \sigma) \\ & \quad \times \left(dx^{\sigma(p+1)} \wedge \cdots \wedge dx^{\sigma(n)} \right) \end{aligned} \quad [27]$$

where σ is a permutation of the indices $(1, \dots, n)$, $\sigma(1) < \cdots < \sigma(p)$, and $\sigma(p+1) < \cdots < \sigma(n)$. We also have

$$\begin{aligned} & * \left(dx^{\sigma(p+1)} \wedge \cdots \wedge dx^{\sigma(n)} \right) \\ &= g_{\sigma(p+1)\sigma(p+1)} \cdots g_{\sigma(n)\sigma(n)} (-1)^{p(n-p)} (\text{sign } \sigma) \\ & \quad \times \left(dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \right) \end{aligned} \quad [28]$$

We therefore find that the application of the Hodge dual to a p -form twice yields

$$\begin{aligned} & ** \left(dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \right) \\ &= g_{\sigma(1)\sigma(1)} \cdots g_{\sigma(p)\sigma(p)} (\text{sign } \sigma) * \left(dx^{\sigma(p+1)} \wedge \cdots \wedge dx^{\sigma(n)} \right) \\ &= g_{\sigma(1)\sigma(1)} \cdots g_{\sigma(n)\sigma(n)} (-1)^{p(n-p)} dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \end{aligned} \quad [29]$$

or

$$** = (-1)^{p(n-p)} (-1)^{\text{Ind } g} \text{Id} \quad [30]$$

where $\text{Ind } g$ is the number of times (-1) occurs along the diagonal of g .

Now let α be a $(p-1)$ -form, and β a p -form. Then $d*\beta$ is an $(n-p+1)$ -form, and

$$\begin{aligned} d(\alpha \wedge *\beta) &= d\alpha \wedge *\beta + (-1)^{p-1} \alpha \wedge d*\beta \\ &= d\alpha \wedge *\beta + (-1)^{(p-1)} (-1)^{(n-p+1)(p-1)} \\ & \quad \times (-1)^{\text{Ind } g} \alpha \wedge (**d*\beta) \\ &= d\alpha \wedge *\beta + (-1)^{n(p-1)} (-1)^{\text{Ind } g} \\ & \quad \times \alpha \wedge (*d*\beta) \end{aligned} \quad [31]$$

We then have

$$(d\alpha, \beta) - (\alpha, d*\beta) = \int_M d(\alpha \wedge *\beta) \quad [32]$$

with

$$d* = -(-1)^{n(p-1)} (-1)^{\text{Ind } g} *d* \quad [33]$$

We are here using the scalar product of two p -forms

$$(\alpha, \beta) := \int_M (\alpha \wedge *\beta) \quad [34]$$

With the help of Stokes' theorem the last integral in eqn [32] may be turned into a surface term at infinity, which vanishes for α and β with compact support. d^* is the adjoint operator to d with respect

to the scalar product $(,)$. Whereas the differential operator d maps p -forms into $(p+1)$ -forms, the codifferential operator d^* maps p -forms into $(p-1)$ -forms.

The relation $d^2 = 0$ leads to

$$(d^*)^2 \propto (*d*)(*d*) \propto *d^2* = 0 \quad [35]$$

This fact plays an essential role in connection with the conservation laws.

Finally, we want to obtain a coordinate expression for $d^*\beta$. Indeed $d^*\beta = -\text{Div } \beta$ for

$$(\text{Div } \beta)_K = \frac{\partial \beta_K^j}{\partial x^j} \quad [36]$$

where K is the multi-index of the coefficients in $\beta = \beta_K dx^K$, and \underline{K} indicates that $K = (k_1, \dots, k_p)$ is in the order $k_1 < \cdots < k_p$. We will show that $(\alpha, d^*\beta) = (\alpha, -\text{Div } \beta)$ for an arbitrary $(p-1)$ -form α . It is a fact that

$$(\alpha, d^*\beta) = (d\alpha, \beta) = \int (d\alpha)_{\underline{I}} \beta^{\underline{I}} * 1 \quad [37]$$

Now we have the coordinate expressions

$$d\alpha = (d\alpha_{\underline{I}}) \wedge dx^{\underline{I}} \quad [38]$$

and $(dx^{\underline{I}})_K = \delta_K^{\underline{I}}$. It follows that

$$(d\alpha)_{\underline{I}} = (d\alpha_{\underline{I}} \wedge dx^{\underline{I}})_{\underline{I}} = \delta_{\underline{I}}^{j\underline{K}} \frac{\partial \alpha_{\underline{I}}}{\partial x^j} \delta_K^{\underline{I}} \quad [39]$$

or

$$(d\alpha)_{\underline{I}} = \delta_{\underline{I}}^{j\underline{K}} \frac{\partial \alpha_K}{\partial x^j} \quad [40]$$

Here we use

$$(\alpha \wedge \beta)_{\underline{I}} = \delta_{\underline{I}}^{KL} \alpha_K \beta_L \quad [41]$$

where

$$\delta_{\underline{I}}^{KL} = \begin{cases} 1 & \text{if (KL) is an even} \\ & \text{permutation of I} \\ -1 & \text{if (KL) is an odd} \\ & \text{permutation of I} \\ 0 & \text{otherwise} \end{cases} \quad [42]$$

Use of the Leibnitz rule yields

$$\begin{aligned} \int (d\alpha)_{\underline{I}} \beta^{\underline{I}} * 1 &= \int \delta_{\underline{I}}^{j\underline{K}} \frac{\partial \alpha_K}{\partial x^j} \beta^{\underline{I}} * 1 \\ &= \int \frac{\partial (\delta_{\underline{I}}^{j\underline{K}} \alpha_K \beta^{\underline{I}})}{\partial x^j} * 1 \\ & \quad - \int \alpha_K \delta_{\underline{I}}^{j\underline{K}} \frac{\partial \beta^{\underline{I}}}{\partial x^j} * 1 \end{aligned} \quad [43]$$

The first term corresponds to a surface integration and we can neglect it. We then have $\delta_L^{jK}\beta^l = \beta^{jK}$ from the antisymmetry of β , so that

$$(\alpha, d^*\beta) = -\int \alpha_K \frac{\partial \beta^{jK}}{\partial x^j} * 1 = (\alpha, -\text{Div}\beta) \quad [44]$$

The Maxwell Equations

The Maxwell equations become remarkably concise when expressed in terms of differential forms, namely

$$dF = 0, \quad d^*F = -j \quad [45]$$

where F is the field strength and j is the current density. We wish to demonstrate this. We use a $(3+1)$ -separation of the exterior derivative into a timelike and a spacelike part:

$$d = d + dt \wedge \frac{\partial}{\partial t} \quad [46]$$

We then get

$$dF = \left(d\mathcal{E} + \frac{\partial \mathcal{B}}{\partial t} \right) \wedge dt + d\mathcal{B} = 0 \quad [47]$$

By comparing coefficients, we arrive at

$$d\mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t}, \quad d\mathcal{B} = 0 \quad [48]$$

In vector notation

$$\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \text{div } \mathbf{B} = 0 \quad [49]$$

the usual form of the homogeneous Maxwell equations.

By direct application of the formula [27], one finds

$$*F = -*\mathcal{B} \wedge dt + *\mathcal{E} \quad [50]$$

where $*$ means the Hodge dual in three space dimensions. One finds

$$d^*F = d*\mathcal{E} - \left(d*\mathcal{B} - \frac{\partial *\mathcal{E}}{\partial t} \right) \wedge dt \quad [51]$$

Therefore,

$$\begin{aligned} d^*F = & -(\text{div } \mathbf{E}) dx \wedge dy \wedge dz \\ & + \left((\text{curl } \mathbf{B})^x - \frac{\partial E^x}{\partial t} \right) dy \wedge dz \wedge dt \\ & + \left((\text{curl } \mathbf{B})^y - \frac{\partial E^y}{\partial t} \right) dz \wedge dx \wedge dt \\ & + \left((\text{curl } \mathbf{B})^z - \frac{\partial E^z}{\partial t} \right) dx \wedge dy \wedge dt \end{aligned} \quad [52]$$

We apply again the Hodge dual:

$$\begin{aligned} *d^*F = & -(\text{div } \mathbf{E}) dt + \left((\text{curl } \mathbf{B})^x - \frac{\partial E^x}{\partial t} \right) dx \\ & + \left((\text{curl } \mathbf{B})^y - \frac{\partial E^y}{\partial t} \right) dy \\ & + \left((\text{curl } \mathbf{B})^z - \frac{\partial E^z}{\partial t} \right) dz \end{aligned} \quad [53]$$

In Minkowski space the expression $*d^*$ equals the codifferential. Therefore, the equation $d^*F = *d^*F = -j$ holds, with j given by $j^\mu = (\rho, \mathbf{J})$, which is equivalent to

$$\text{div } \mathbf{E} = \rho, \quad \text{curl } \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{J} \quad [54]$$

the inhomogeneous Maxwell equations.

Current Conservation

The electromagnetic 4-current is

$$j^\mu = \rho_0 u^\mu = (\rho_0 \gamma, \rho_0 \gamma \mathbf{v}) = (\rho, \mathbf{J}) \quad [55]$$

where ρ is the charge density and \mathbf{J} the current density. This corresponds to a 1-form

$$j = \rho dt - J^x dx - J^y dy - J^z dz \quad [56]$$

The Hodge dual is $*j = \sigma^3 - j^2 \wedge dt$, with the 3-form $\sigma^3 = \rho dx \wedge dy \wedge dz$, and the 2-form

$$j^2 = -J^x dy \wedge dz - J^y dz \wedge dx - J^z dx \wedge dy \quad [57]$$

From the Maxwell equation $d^*F = -j$, it follows that

$$(d^*)^2 F = -d^*j = 0 \quad [58]$$

that is

$$\begin{aligned} *d(*j) = & *d(\sigma^3 - j^2 \wedge dt) = *(d\sigma^3 - dj^2 \wedge dt) \\ = & * \left(\frac{\partial \rho}{\partial t} + \text{div } \mathbf{J} \right) dt \wedge dx \wedge dy \wedge dz \\ = & \frac{\partial \rho}{\partial t} + \text{div } \mathbf{J} = 0 \end{aligned} \quad [59]$$

This is the ‘‘continuity equation.’’

The total charge inside a volume V is $Q = \int_V \rho dV$, therefore

$$-\frac{dQ}{dt} = -\frac{d}{dt} \int_V \rho dV = \int_{\partial V} \mathbf{J} \cdot \mathbf{n} dS \quad [60]$$

where ∂V is the surface which encloses the volume V , dS is the surface element, and \mathbf{n} is the normal vector to this surface. This is current conservation.

The Gauge Potential

The ‘‘Poincaré lemma’’ tells us that $dF=0$ implies $F=dA$, with the 4-potential A :

$$A = \phi dt + A \quad [61]$$

and the vector potential $A = A_x dx + A_y dy + A_z dz$. From

$$\begin{aligned} F &= \mathcal{E} \wedge dt + \mathcal{B} = \left(d + dt \wedge \frac{\partial}{\partial t} \right) A \\ &= d\phi \wedge dt + dA + dt \wedge \frac{\partial A}{\partial t} \end{aligned} \quad [62]$$

it follows by comparing coefficients that

$$\mathcal{E} = d\phi - \frac{\partial A}{\partial t}, \quad \mathcal{B} = dA \quad [63]$$

In vector notation this is

$$E = \text{grad}\phi - \frac{\partial A}{\partial t}, \quad B = \text{curl}A \quad [64]$$

The 4-potential is determined up to a gauge function Λ :

$$A' = A + d\Lambda \quad [65]$$

This gauge freedom has no influence on the observable quantities E and B :

$$F' = dA' = dA + d^2\Lambda = dA = F \quad [66]$$

The Laplace operator is $\Delta = (d^* + d)^2 = dd^* + d^*d$, so when the 4-potential A fulfills the condition $d^*A = 0$, we have

$$\Delta A = d^*dA = d^*F = -j \quad [67]$$

the ‘‘classical wave equation.’’ The condition $d^*A = 0$ is called the ‘‘Lorentz gauge condition.’’ This condition can always be fulfilled by using the gauge freedom: $d^*(A + d\Lambda) = 0$ is fulfilled when $d^*d\Lambda = \Delta\Lambda = -d^*A$, where we have used the fact that $d^*\Lambda = 0$ for functions. That is to say, $d^*A = 0$ is fulfilled when Λ is a solution of the inhomogeneous wave equation.

Gauge Invariance

In quantum mechanics, the electron is described by a wave function which is determined up to a free phase. Indeed, at every point in space this phase can be chosen arbitrarily:

$$\begin{aligned} \psi(x) &\rightarrow \psi'(x) = \exp\{i\alpha(x)\}\psi(x) \\ \bar{\psi}(x) &\rightarrow \bar{\psi}'(x) = \bar{\psi}(x) \exp\{-i\alpha(x)\} \end{aligned} \quad [68]$$

with the only condition being that $\alpha(x)$ is a continuous function. The gauge transformation is

of the form $g = \exp\{i\alpha(x)\}$, with g an element of the abelian gauge group $G = U(1)$. The free action is

$$S_0 = \int \mathcal{L}_0 d^4x \quad [69]$$

with

$$\mathcal{L}_0 = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad [70]$$

the ‘‘Lagrange density.’’ This action is not invariant under gauge transformations:

$$\mathcal{L}_0 \rightarrow \mathcal{L}'_0 = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - (\partial_\mu \alpha) \bar{\psi} \gamma^\mu \psi \quad [71]$$

The undesired term can be compensated by the introduction of a gauge potential ω in a covariant derivative of ψ ,

$$D\psi = (d + \omega)\psi \quad [72]$$

which has the desired transformation property $D\psi \rightarrow \exp\{i\alpha\}D\psi$ when besides the transformation $\psi(x) \rightarrow \exp\{i\alpha(x)\}\psi(x)$ of the matter field the gauge potential simultaneously transforms according to the gauge transformation $\omega \rightarrow \omega - id\alpha$. The new Lagrange density is

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi = \mathcal{L}_0 + i\omega_\mu \bar{\psi}(x)\gamma^\mu \psi(x) \quad [73]$$

The substitution $\partial_\mu \rightarrow D_\mu$ is known to physicists; with $\omega = -iqA$ it is the ansatz of minimal coupling for taking into account electromagnetic effects: $\partial_\mu \rightarrow \partial_\mu - iqA_\mu$. The Lagrange density becomes in this notation $\mathcal{L} = \mathcal{L}_0 - A_\mu J^\mu$, where $J^\mu = -q\bar{\psi}\gamma^\mu\psi$.

The Lagrange density must now be completed by a kinetic term for the gauge potential and we get the complete electromagnetic Lagrange density

$$\mathcal{L} = \mathcal{L}_0 - A_\mu J^\mu - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad [74]$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. In the action this corresponds to

$$S = S_0 - \int_M A_\mu J^\mu \text{vol}^4 - \frac{1}{4} \int_M F_{\mu\nu} F^{\mu\nu} \text{vol}^4 \quad [75]$$

We get the field equations for the potential A by demanding that the variation of the action vanishes:

$$\delta S[A] = - \int_M \delta A_\mu J^\mu \text{vol}^4 - \frac{1}{4} \delta \int_M F_{\mu\nu} F^{\mu\nu} \text{vol}^4 \quad [76]$$

We write now

$$\int_M \delta A_\mu J^\mu \text{vol}^4 = (\delta A, j) \quad [77]$$

and

$$\begin{aligned} & \frac{1}{4} \delta \int_M F_{\mu\nu} F^{\mu\nu} \text{vol}^4 \\ &= \frac{1}{2} \delta \int_M F \wedge *F = \frac{1}{2} \delta(F, F) \\ &= (\delta dA, F) = (d\delta A, F) = (\delta A, d^*F) \end{aligned} \quad [78]$$

where we have exchanged the action of δ and d . Since this holds for arbitrary variations δA we find

$$d^*F = -j \quad [79]$$

the inhomogeneous Maxwell equation.

Nonabelian Gauge Theories

In $SU(N)$ gauge theory the elementary particles are taken to be members of symmetry multiplets. For example, in electroweak theory the left-handed electron and the neutrino are members of an $SU(2)$ doublet:

$$\psi = \begin{pmatrix} e^- \\ \nu \end{pmatrix} \quad [80]$$

A gauge transformation is

$$\psi'(x) = g^{-1}(x)\psi(x), \quad \bar{\psi}'(x) = \bar{\psi}(x)g(x) \quad [81]$$

with

$$g(x) = \exp \{ \Lambda(x) \} \quad [82]$$

where $g(x)$ is an element of the Lie group $SU(2)$ and Λ is an element of the Lie algebra $\mathfrak{su}(2)$. The Lie algebra is a vector space, and its elements may be expanded in terms of a basis:

$$\Lambda(x) = \Lambda^a(x)T_a \quad [83]$$

For $\mathfrak{su}(2)$ the basis elements are traceless and anti-Hermitian (see below), they are conventionally expressed in terms of the Pauli matrices,

$$T_a = \frac{\sigma_a}{2i} \quad [84]$$

with

$$\begin{aligned} \sigma_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & \sigma_2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ \sigma_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \quad [85]$$

They are conventionally normalized according to

$$\text{tr}(T_a T_b) = -\frac{1}{2} \delta_{ab} \quad [86]$$

The Dirac Lagrangian is not invariant with respect to local gauge transformations:

$$\begin{aligned} \mathcal{L}_0 &= \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \rightarrow \mathcal{L}'_0 \\ &= \mathcal{L}_0 + i\bar{\psi}\gamma^\mu (g\partial_\mu g^{-1})\psi \end{aligned} \quad [87]$$

We introduce the gauge potential

$$\omega_\mu(x) = \omega_\mu^a(x)T_a \quad [88]$$

with a gauge transformation

$$\omega_\mu \rightarrow \omega'_\mu = g^{-1}\omega_\mu g + g^{-1}\partial_\mu g \quad [89]$$

The Lagrange density is modified through a covariant derivative:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + \omega_\mu \quad [90]$$

The covariant derivative D_μ transforms according to

$$D_\mu \rightarrow D'_\mu = g^{-1}D_\mu g \quad [91]$$

and thus the modified Lagrange density

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi = \mathcal{L}_0 + i\bar{\psi}\gamma^\mu \omega_\mu \psi \quad [92]$$

is invariant with respect to local gauge transformations.

The extra term in the Lagrange density is conventionally written

$$-J_a^\mu A_\mu^a \quad [93]$$

with

$$A_\mu^a = -iq\omega_\mu^a \quad [94]$$

and

$$J_a^\mu = \bar{\psi}\gamma^\mu T_a \psi \quad [95]$$

In mathematical terminology ω is called a connection. The quantity A is the physicist's gauge potential. The connection is anti-Hermitian and the gauge potential Hermitian. The gauge potential also includes the coupling constant q . We will refer to both ω and A as the gauge potential, where the relation between them is given by eqn [94].

We can write the gauge potential as $A = A_\mu^a dx^\mu T_a$ or, in the $SU(2)$ case, as

$$A_\mu = A_\mu^1 T_1 + A_\mu^2 T_2 + A_\mu^3 T_3 \quad [96]$$

where we see explicitly that it involves three vector fields, which couple to the electroweak currents [95] with the single coupling constant q , and which will become after symmetry breaking the three vector bosons W_+ , W_- , Z_0 of the electroweak gauge theory. Actually, a mix of the neutral gauge boson and the photon will combine to yield the Z_0 boson, while the orthogonal mixture gives rise to the electromagnetic interaction, in an $SU(2) \times U(1)$ theory. At this stage,

the gauge bosons are all massless, their masses are generated by the ‘‘Higgs’ mechanism.’’

Lie-Algebra-Valued p -Forms

To describe nonabelian fields, we need Lie-algebra-valued p -forms:

$$\phi = T_a \phi^a \quad [97]$$

where T_a is a generator of the Lie algebra, the index a runs over the number of generators of the Lie algebra, and the ϕ^a are the usual scalar-valued p -forms. The composition in a Lie algebra is a Lie bracket, which is defined for two Lie-algebra-valued p -forms by

$$[\phi, \psi] := [T_a, T_b] \phi^a \wedge \psi^b \quad [98]$$

The Lie bracket in the algebra is

$$[T_a, T_b] = f_{ab}^c T_c \quad [99]$$

where f_{bc}^a are the structure constants. It follows from this that

$$[\psi, \phi] = [T_a, T_b] \psi^a \wedge \phi^b = -[T_b, T_a] \psi^a \wedge \phi^b \quad [100]$$

or

$$[\psi, \phi] = (-1)^{pq+1} [\phi, \psi] \quad [101]$$

when ϕ is a p -form and ψ is a q -form. In the special case that T_a is a matrix, also the product $T_a T_b$ is defined, and from this the product of two Lie-algebra-valued p -forms

$$\phi \wedge \psi = T_a \phi^a \wedge T_b \psi^b = T_a T_b \phi^a \wedge \psi^b \quad [102]$$

Now the Lie bracket is a commutator:

$$[T_a, T_b] = T_a T_b - T_b T_a \quad [103]$$

and

$$\begin{aligned} [\phi, \psi] &= [T_a, T_b] \phi^a \wedge \psi^b \\ &= T_a \phi^a \wedge T_b \psi^b - (-1)^{pq} T_b \psi^b \wedge T_a \phi^a \\ &= \phi \wedge \psi - (-1)^{pq} \psi \wedge \phi \end{aligned} \quad [104]$$

From this relation it follows that for ϕ and ψ odd p -forms

$$[\phi, \psi] = \phi \wedge \psi + \psi \wedge \phi \quad [105]$$

For ϕ an odd p -form

$$[\phi, \phi] = \phi \wedge \phi + \phi \wedge \phi = 2(\phi \wedge \phi) \quad [106]$$

The Gauge Potential and the Field Strength

The generalization of the abelian relationship between the gauge potential and the field strength, $F = dA$, is

$$\theta = d\omega + \frac{1}{2}[\omega, \omega] = d\theta + \omega \wedge \omega \quad [107]$$

where because ω is a 1-form we can use eqn [106]. The mathematician refers to θ as the curvature. The physicist writes, in analogy to eqn [94],

$$F = -i q \theta = \frac{1}{2} F_{\mu\nu}^a dx^\mu \wedge dx^\nu T_a \quad [108]$$

One obtains for the components

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - i q f_{bc}^a A_\mu^b A_\nu^c \quad [109]$$

A generalization of the gauge transformation of A , that is, $A' = A + d\Lambda$, is eqn [89]:

$$\omega' = g^{-1} \omega g + g^{-1} dg \quad [110]$$

A quantity ϕ with the transformation property

$$\phi' = g^{-1} \phi g \quad [111]$$

is called a ‘‘tensorial’’ quantity. The gauge potential ω is according to this definition nontensorial. Nevertheless the field strength is tensorial. Indeed

$$\begin{aligned} \theta' &= d(g^{-1} \omega g) + (dg^{-1}) \wedge dg \\ &\quad + \frac{1}{2} [g^{-1} \omega g + g^{-1} dg, g^{-1} \omega g + g^{-1} dg] \\ &= (dg^{-1}) \wedge \omega g + g^{-1} d\omega g - g^{-1} \omega \wedge dg + (dg^{-1}) \wedge dg \\ &\quad + \frac{1}{2} g^{-1} [\omega, \omega] g + \frac{1}{2} [g^{-1} \omega g, g^{-1} dg] \\ &\quad + \frac{1}{2} [g^{-1} dg, g^{-1} \omega g] + \frac{1}{2} [g^{-1} dg, g^{-1} dg] \\ &= g^{-1} \theta g + (dg^{-1}) \wedge \omega g - g^{-1} \omega \wedge dg + (dg^{-1}) \wedge dg \\ &\quad + g^{-1} \omega \wedge dg + g^{-1} dg \wedge g^{-1} \omega g + g^{-1} dg \wedge g^{-1} dg \\ &= g^{-1} \theta g \end{aligned} \quad [112]$$

where we have used the derivation of the relation $g^{-1} g = \text{Id}$ to get

$$dg^{-1} = -g^{-1} dg g^{-1} \quad [113]$$

In the abelian case, we had $dF = 0$. The non-abelian analog is

$$\begin{aligned} d\theta &= d\omega \wedge \omega - \omega \wedge d\omega \\ &= (\theta - \omega \wedge \omega) \wedge \omega - \omega \wedge (\theta - \omega \wedge \omega) \\ &= \theta \wedge \omega - \omega \wedge \theta \end{aligned} \quad [114]$$

or

$$d\theta + \omega \wedge \theta - \theta \wedge \omega = 0 \quad [115]$$

the Bianchi identity. It can also be written as

$$d\theta + \omega \wedge \theta - \theta \wedge \omega = d\theta + [\omega, \theta] = 0 \quad [116]$$

because from eqn [104]

$$\omega \wedge \theta + (-1)^{2-1}\theta \wedge \omega = [\omega, \theta] \quad [117]$$

The covariant derivative D is defined as

$$D\phi := d\phi + [\omega, \phi] \quad [118]$$

for ϕ a tensorial quantity. The covariant derivative takes tensorial p -forms into tensorial $(p+1)$ -forms:

$$\begin{aligned} D'\phi &= d(g^{-1}\phi g) + [g^{-1}\omega g + g^{-1}dg, g^{-1}\phi g] \\ &= dg^{-1} \wedge \phi g + g^{-1}d\phi g + (-1)^p g^{-1}\phi \wedge dg \\ &\quad + [g^{-1}\omega g, g^{-1}\phi g] + [g^{-1}dg, g^{-1}\phi g] \\ &= g^{-1}D\phi g + dg^{-1} \wedge \phi g + (-1)^p g^{-1}\phi \wedge dg \\ &\quad + g^{-1}dgg^{-1} \wedge \phi g - (-1)^p g^{-1}\phi \wedge dg \\ &= g^{-1}D\phi g \end{aligned} \quad [119]$$

We have thereby verified the transformation property of eqn [91].

The Gauge Group

From the gauge transformation $\psi' = g\psi$ the requirement $|\psi'|^2 = |\psi|^2$ leads to $g^\dagger g = 1$. That means that g belongs to the unitary Lie group $G = U(n)$, whose elements fulfill $g^\dagger = \bar{g}^T = g^{-1}$. For elements of the Lie algebra $\mathcal{G} = \mathfrak{u}(n)$ this implies

$$(e^X)^\dagger = e^{\bar{X}^T} = e^{-X} \quad [120]$$

or

$$X^\dagger = \bar{X}^T = -X \quad [121]$$

where \bar{X} is complex conjugation and X^T means transposition.

For elements of the Lie algebra we can define a scalar product (the Killing metric)

$$\langle X, Y \rangle := -\text{tr}(XY) = -X^\alpha_\beta X^\beta_\alpha \quad [122]$$

The scalar product is real:

$$\langle \bar{X}, \bar{Y} \rangle = -\bar{X}^\beta_\alpha \bar{Y}^\alpha_\beta = -X^\alpha_\beta X^\beta_\alpha = \langle X, Y \rangle \quad [123]$$

symmetric:

$$\langle X, Y \rangle = -\text{tr}(X, Y) = -\text{tr}(Y, X) = \langle Y, X \rangle \quad [124]$$

and positive definite:

$$\langle X, X \rangle = -X^\alpha_\beta X^\beta_\alpha = X^\alpha_\beta \bar{X}^\alpha_\beta = |X^\alpha_\beta|^2 \quad [125]$$

The scalar product is invariant under the action of G on \mathcal{G} : for $g \in G$

$$\begin{aligned} \langle gXg^{-1}, gYg^{-1} \rangle &= -\text{tr}(gXYg^{-1}) \\ &= -\text{tr}(X, Y) = \langle X, Y \rangle \end{aligned} \quad [126]$$

or for $X, Y, Z \in \mathcal{G}$

$$\langle e^{tX}Y e^{-tX}, e^{tX}Z e^{-tX} \rangle = \langle Y, Z \rangle \quad [127]$$

We take the derivative of this equation with respect to t at the value $t=0$ and get:

$$\langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle = 0 \quad [128]$$

We define an action of the algebra \mathcal{G} on itself: $ad(X): \mathcal{G} \rightarrow \mathcal{G}$

$$ad(X)Y = [X, Y] \quad [129]$$

We can then formulate our conclusion as follows: the action of \mathcal{G} on itself is anti-Hermitian:

$$\langle ad(X)Y, Z \rangle = -\langle Y, ad(X)Z \rangle \quad [130]$$

or

$$[ad(X)]^\dagger = -ad(X) \quad [131]$$

From $g^\dagger g = 1$ we have $|\det(g)|^2 = 1$. For the gauge group $G = SU(N)$ we require in addition $\det(g) = 1$. Since

$$\det(g) = \det(\exp(X)) = \exp(\text{tr}(X)) \quad [132]$$

the elements $X \in \mathfrak{su}(N)$ must be traceless. A basis of the vector space of traceless, anti-Hermitian (2×2) matrices is given by the Pauli matrices, eqn [85].

The Yang–Mills Action

The $SU(2)$ Yang–Mills action is, in analogy to the abelian case,

$$\begin{aligned} S &= -\frac{1}{4q^2} \int_M F^a_{\mu\nu} F^{a\mu\nu} \text{vol}^4 = \frac{1}{2q^2} \int_M \text{tr}(F_{\mu\nu} F^{\mu\nu}) \text{vol}^4 \\ &= \frac{1}{2q^2} \int_M \text{tr}(F \wedge *F) \end{aligned} \quad [133]$$

We have included the trace in our definition of the scalar product:

$$(\phi, \psi) := -\int_M \text{tr} \langle \phi_I \psi^I \rangle \text{vol}^n = -\int_M \text{tr}(\phi \wedge * \psi) \quad [134]$$

We then write eqn [133] as

$$S[\omega] = \frac{1}{2}(\theta, \theta) \quad [135]$$

taking into account the relation between θ and the field strength F , and indicating the dependence on

the gauge potential. Since θ is tensorial the action is invariant.

Now we calculate the variation von $S[\omega]$ with respect to a variation of the gauge potential:

$$\begin{aligned}\delta S[\omega] &= \frac{d}{dt} S[\omega(t)]|_{t=0} = \frac{1}{2} \delta(\theta, \theta) \\ &= \frac{1}{2} ((\delta\theta, \theta) + (\theta, \delta\theta)) \\ &= (\delta\theta, \theta) = \left(\delta \left(d\omega + \frac{1}{2} [\omega, \omega] \right), \theta \right) \\ &= \left(\delta d\omega + \frac{1}{2} [\delta\omega, \omega] + \frac{1}{2} [\omega, \delta\omega], \theta \right) \\ &= (d\delta\omega + [\omega, \delta\omega], \theta) \quad [136]\end{aligned}$$

where we have exchanged the order of δ and d . We remark that although ω is not a tensorial section, $\delta\omega$ is: for $\omega'_1 = g^{-1}\omega_1g + g^{-1}dg$ and $\omega'_2 = g^{-1}\omega_2g + g^{-1}dg$ is

$$\delta\omega = \omega'_1 - \omega'_2 = g^{-1}(\omega_1 - \omega_2)g \quad [137]$$

The quantity θ is in any case tensorial. Therefore, the covariant derivative is defined, and we have

$$D\delta\omega = d\delta\omega + [\omega, \delta\omega] \quad [138]$$

and

$$D\theta = d\theta + [\omega, \theta] \quad [139]$$

In general, the action of the covariant derivative on tensorial quantities can be written as $D = d + ad(\omega)$, where $ad(X)$ is the representation of the Lie algebra on itself introduced in the previous section. We now have

$$\delta S[\omega] = (D\delta\omega, \theta) = (\delta\omega, D^*\theta) = 0 \quad [140]$$

for an arbitrary variation $\delta\omega$. Therefore, $D^*\theta = 0$.

We have obtained

$$D^*\theta = 0 \quad [141]$$

the ‘‘Yang–Mills equations,’’ and

$$D\theta = 0 \quad [142]$$

the ‘‘Bianchi identities.’’ These are the generalizations of the Maxwell equations $d^*F = 0$ and $dF = 0$ in the absence of external sources. For the general case of interacting fermions, we write out the full action, in analogy to eqn [74], and obtain, in analogy to eqns [79] and [58],

$$D^*\theta = -J, \quad D^*J = 0 \quad [143]$$

We shall now derive, again for the pure gauge sector, coordinate expressions for the Yang–Mills equations. Consider the expression

$$\begin{aligned}\delta S[\omega] &= (D\delta\omega, \theta) = (\delta\omega, D^*\theta) \\ &= (d\delta\omega + [\omega, \delta\omega], \theta) \quad [144]\end{aligned}$$

The first term in the last expression is

$$(d\delta\omega, \theta) = (\delta\omega, d^*\theta) = -\text{tr} \int_M \delta\omega_\nu \{d^*\theta\}^\nu \text{vol}^4 \quad [145]$$

The second term can be computed using

$$\begin{aligned}[\omega, \delta\omega]_{\mu\nu} &= \{\omega \wedge \delta\omega + \delta\omega \wedge \omega\}(\partial_\mu, \partial_\nu) \\ &= \omega_\mu \delta\omega_\nu - \omega_\nu \delta\omega_\mu + \delta\omega_\mu \omega_\nu - \delta\omega_\nu \omega_\mu \quad [146]\end{aligned}$$

and hence

$$[\omega, \delta\omega]_{\mu\nu} \theta^{\mu\nu} = 2[\omega_\mu, \delta\omega_\nu] \theta^{\mu\nu} \quad [147]$$

because θ is antisymmetric, $\theta^{\mu\nu} = -\theta^{\nu\mu}$. Thus,

$$\begin{aligned}([\omega, \delta\omega], \theta) &= -\int_M \text{tr}([\omega, \delta\omega] \wedge * \theta) \\ &= -\frac{1}{2} \int_M \text{tr}([\omega, \delta\omega]_{\mu\nu} \theta^{\mu\nu}) \text{vol}^4 \\ &= -\int_M \text{tr}([\omega_\mu, \delta\omega_\nu] \theta^{\mu\nu}) \text{vol}^4 \\ &= \int_M \langle [\omega_\mu, \delta\omega_\nu], \theta^{\mu\nu} \rangle \text{vol}^4 \quad [148]\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the scalar product in \mathcal{G} . From eqn [128] this equals

$$\begin{aligned}& -\int_M \langle \delta\omega_\nu, [\omega_\mu, \theta^{\mu\nu}] \rangle \text{vol}^4 \\ &= \int_M \text{tr}(\delta\omega_\nu [\omega_\mu, \theta^{\mu\nu}]) \text{vol}^4 \quad [149]\end{aligned}$$

Combining this with eqn [144] gives

$$\begin{aligned}(\delta\omega, D^*\theta) &= -\int_M \text{tr}(\delta\omega_\nu \{ (d^*\theta)^\nu - [\omega_\mu, \theta^{\mu\nu}] \}) \text{vol}^4 \\ &= (\delta\omega, \{ (d^*\theta)^\nu - [\omega_\mu, \theta^{\mu\nu}] \}) \quad [150]\end{aligned}$$

We can now insert the coordinate expression for

$$(d\theta)^\nu = -\partial_\mu \theta^{\mu\nu} \quad [151]$$

Finally, the coordinate expressions of the Yang–Mills equations $D^*\theta = 0$ are

$$(D^*\theta)^\nu = -\{ \partial_\mu \theta^{\mu\nu} + [\omega_\mu, \theta^{\mu\nu}] \} = 0 \quad [152]$$

The Analogy with Electromagnetism

The Yang–Mills equation and the Bianchi identity in the absence of external sources are

$$\partial_\nu F^{\mu\nu} - iq[A_\nu, F^{\mu\nu}] = 0 \quad [153]$$

and

$$\begin{aligned}\partial_\mu F_{\nu\tau} + \partial_\tau F_{\mu\nu} + \partial_\nu F_{\tau\mu} - iq\{[A_\mu, F_{\nu\tau}] \\ + [A_\tau, F_{\mu\nu}] + [A_\nu, F_{\tau\mu}]\} = 0 \quad [154]\end{aligned}$$

We shall write these equations in terms of the fields

$$F^{i0} = E^i, \quad i = 1, 2, 3 \quad [155]$$

$$F^{12} = B^3, \quad F^{31} = B^2, \quad F^{23} = B^1 \quad [156]$$

where the \mathbf{E} and \mathbf{B} vectors may be thought of as “electric” and “magnetic” fields, even though they have Lie-algebra indices, $F^{i0} = (F^a)^{i0} T_a$, etc. In the context of the SU(3) theory, they are referred to as the “chromoelectric” and “chromomagnetic” fields, respectively.

The Yang–Mills equations with $\mu = 0$ are

$$\partial_i F^{i0} - iq[A_i, F^{i0}] = 0 \quad [157]$$

with $i = 1, 2, 3$ a spatial index. In vector notation this is

$$\operatorname{div} \mathbf{E} = iq(\mathbf{A} \cdot \mathbf{E} - \mathbf{E} \cdot \mathbf{A}) \quad [158]$$

This is the analog of Gauss’s equation. Even though we started out without external sources, $iq(\mathbf{A} \cdot \mathbf{E} - \mathbf{E} \cdot \mathbf{A})$ plays the role of a “charge density.” The Yang–Mills field \mathbf{E} and the potential \mathbf{A} combine to act as a source for the Yang–Mills field. This is an essential feature of nonabelian gauge theories in which they differ from the abelian case, due to the fact that the commutator $[\mathbf{A}, \mathbf{E}]$ is nonvanishing.

Now consider the Yang–Mills equations with a spatial index $\mu = i$:

$$\partial_0 F^{i0} + \partial_j F^{ij} - iq[A_0, F^{i0}] - iq[A_j, F^{ij}] = 0 \quad [159]$$

In vector notation this is

$$\operatorname{curl} \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} - iq(A_0 \mathbf{E} - \mathbf{E} A_0) + iq(\mathbf{A} \times \mathbf{B} + \mathbf{B} \times \mathbf{A}) \quad [160]$$

replacing the Ampere–Maxwell law. Note that there are two extra contributions to the “current” other than the displacement current.

The analogs of the laws of Faraday and of the absence of magnetic monopoles are derived similarly from the Bianchi identities. The results are

$$\operatorname{curl} \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = iq\{(\mathbf{A} \times \mathbf{E} + \mathbf{E} \times \mathbf{A}) + (A_0 \mathbf{B} - \mathbf{B} A_0)\} \quad [161]$$

and

$$\operatorname{div} \mathbf{B} = iq(\mathbf{A} \cdot \mathbf{B} - \mathbf{B} \cdot \mathbf{A}) \quad [162]$$

Further Remarks

The foundations of the mathematics of differential forms were laid down by Poincaré (1953). They were applied to the description of electrodynamics

already by Cartan (1923). A modern presentation of differential forms and the manifolds on which they are defined is given in Abraham *et al.* (1983). A recent treatment of electrodynamics in this approach is Hehl and Obukhov (2003). Weyl’s argument is in his paper of 1929.

Nonabelian gauge theories today explain the electromagnetic, the strong and weak nuclear interactions. The original paper is that of Yang and Mills (1954). Glashow, Salam, and Weinberg (1980) saw the way to apply it to the weak interactions by using spontaneous symmetry breaking to generate the masses through the use of the Higgs’ (1964) mechanism. t’Hooft and Veltman (1972) showed that the resulting quantum field theory was renormalizable. The strong interactions were recognized as the nonabelian gauge theory with gauge group SU(3) by Gell-Mann (1972). For a modern treatment which puts nonabelian gauge theories in the context of differential geometry, see Frankel (1987).

See also: Dirac Fields in Gravitation and Nonabelian Gauge Theory; Electroweak Theory; Measure on Loop Spaces; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Electrodynamics and its Precision Tests.

Further Reading

- Abraham A, Marsden J, and Ratiu T (1983) *Manifolds, Tensor Analysis, and Applications*. MA: Addison-Wesley.
- Cartan É (1923) *On manifolds with an Affine Connection and the Theory of General Relativity*. English translation of the French original 1923/1924 (Bibliopolis, Napoli 1986).
- Frankel T (1987) *The Geometry of Physics, An Introduction*. Cambridge University Press.
- Gell-Mann M (1972) Quarks: developments in the quark theory of hadrons. *Acta Physica Austriaca Suppl.* IV: 733.
- Glashow SL (1980) Towards a unified theory: threads in a tapestry. *Reviews of Modern Physics* 52: 539.
- Hehl FW and Obukhov YN (2003) *Foundations of Classical Electrodynamics*. Boston: Birkhäuser.
- Higgs PW (1964) Broken symmetries and the masses of gauge bosons. *Physical Review Letters* 13: 508.
- t’Hooft G and Veltman M (1972) Regularization and renormalization of gauge fields. *Nuclear Physics B* 44: 189.
- Poincaré H (1953) *Oeuvre*. Paris: Gauthier-Villars.
- Salam A (1980) Gauge unification of fundamental forces. *Reviews of Modern Physics* 52: 525.
- Weinberg SM (1980) Conceptual foundations of the unified theory of weak and electromagnetic interactions. *Reviews of Modern Physics* 52: 515.
- Weyl H (1929) Elektron und gravitation. *Zeitschrift fuer Physik* 56: 330.
- Yang CN and Mills RL (1954) Construction of isotopic spin and isotopic gauge invariance. *Physical Review* 96: 191.

Abelian Higgs Vortices

J M Speight, University of Leeds, Leeds, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

For the purpose of this article, vortices are topological solitons arising in field theories in $(2 + 1)$ -dimensional spacetime when a complex-valued field ϕ is allowed to acquire winding at infinity, meaning that the phase of $\phi(t, \mathbf{x})$, as \mathbf{x} traverses a large circle in the spatial plane, changes by $2\pi n$, where n is a nonzero integer. Such winding cannot be removed by any continuous deformation of ϕ (hence “topological”) and traps a considerable amount of energy which tends to coalesce into smooth, stable lumps with highly particle-like characteristics (hence “solitons”). Clearly, the universe is $(3 + 1)$ dimensional. Nonetheless, planar field theories are of physical interest for two main reasons. First, the theory may arise by dimensional reduction of a $(3 + 1)$ -dimensional model under the assumption of translation invariance in one direction. Vortices are then transverse slices through straight tube-like objects variously interpreted as magnetic flux tubes in a superconductor or cosmic strings. Second, a crucial ingredient of the standard model of particle physics is spontaneous breaking of gauge symmetry by a Higgs field. As well as endowing the fundamental gauge bosons and chiral fermions with mass, this mechanism can potentially generate various types of topological solitons (monopoles, strings, and domain walls) whose structure and interactions one would like to understand. Vortices in $(2 + 1)$ dimensions are interesting in this regard because they arise in the simplest field theory exhibiting the Higgs mechanism, the abelian Higgs model (AHM). They are thus a useful theoretical laboratory in which to test ideas which may ultimately find application in more realistic theories. This article describes the properties of abelian Higgs vortices and explains how, using a mixture of numerical and analytical techniques, a good understanding of their dynamical interactions has been obtained.

The Abelian Higgs Model

Throughout this article spacetime will be \mathbb{R}^{2+1} endowed with the Minkowski metric with signature $(+, -, -)$, and Cartesian coordinates $x^\mu, \mu = 0, 1, 2$, with $x^0 = t$ (the speed of light $c = 1$). A spacetime point will be denoted x , its spatial part by $\mathbf{x} = (x^1, x^2)$. Latin indices j, k, \dots range over 1, 2, and repeated indices (Latin or Greek) are summed over.

We sometimes use polar coordinates in the spatial plane, $\mathbf{x} = r(\cos\theta, \sin\theta)$, and sometimes a complex coordinate $z = x^1 + ix^2 = re^{i\theta}$. Occasionally, it is convenient to think of \mathbb{R}^{2+1} as a subspace of \mathbb{R}^{3+1} and denote by \mathbf{k} the unit vector in the (fictitious) third spatial direction. The complex scalar Higgs field is denoted ϕ , and the electromagnetic gauge potential A_μ , best thought of as the components of a 1-form $A = A_\mu dx^\mu$. $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor which, in \mathbb{R}^{2+1} , has only three independent components, identified with the magnetic field $B = F_{12}$ and electric field $(E_1, E_2) = (F_{01}, F_{02})$. The gauge-covariant derivative is $D_\mu\phi = \partial_\mu\phi - ieA_\mu\phi$, e being the electric charge of the Higgs. Under a $U(1)$ gauge transformation,

$$\phi \mapsto e^{i\Lambda}\phi, \quad A_\mu \mapsto A_\mu + e^{-1}\partial_\mu\Lambda \quad [1]$$

$\Lambda: \mathbb{R}^{2+1} \rightarrow \mathbb{R}$ being any smooth function, $F_{\mu\nu}$ and $|\phi|$ remain invariant, while $D_\mu\phi \mapsto e^{i\Lambda}D_\mu\phi$. Only gauge-invariant quantities are physically observable (classically).

With these conventions, the AHM has Lagrangian density

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{\alpha}{2}D_\mu\phi\overline{D^\mu\phi} - \frac{\lambda}{8}(\nu^2 - |\phi|^2)^2 \quad [2]$$

which is manifestly gauge invariant. By rescaling ϕ, A_μ, \mathbf{x} and the unit of action, we can (and henceforth will) assume that $e = \nu = \alpha = 1$. The only parameter which cannot be scaled away is $\lambda > 0$. Its value greatly influences the model’s behavior.

The field equations, obtained by demanding that $\phi(x), A_\mu(x)$ be a local extremal of the action $S = \int \mathcal{L} d^3x$, are

$$\begin{aligned} D_\mu D^\mu\phi + \frac{\lambda}{2}(1 - |\phi|^2)\phi &= 0 \\ \partial^\mu F_{\mu\nu} + \frac{i}{2}(\phi\overline{D_\nu\phi} - \overline{\phi}D_\nu\phi) &= 0 \end{aligned} \quad [3]$$

This is a coupled set of nonlinear second-order PDEs. Of particular interest are solutions which have finite total energy. Energy is not a Lorentz-invariant quantity. To define it we must choose an inertial frame and, having broken Lorentz invariance, it is convenient to work in a temporal gauge, for which $A_0 \equiv 0$ (which may be obtained by a gauge transformation with $\Lambda(t, \mathbf{x}) = \int_0^t A_0(t', \mathbf{x}) dt'$, after which only time-independent gauge transformations are permitted). The potential energy of a field is then

$$\begin{aligned} E &= \frac{1}{2} \int \left(B^2 + D_i\phi\overline{D_i\phi} + \frac{\lambda}{4}(1 - |\phi|^2)^2 \right) dx^1 dx^2 \\ &= E_{\text{mag}} + E_{\text{grad}} + E_{\text{self}} \end{aligned} \quad [4]$$

while its kinetic energy is

$$E_{\text{kin}} = \frac{1}{2} \int \left(|\partial_0 A|^2 + \partial_0 \phi \overline{\partial_0 \phi} \right) dx^1 dx^2 \quad [5]$$

If ϕ, A satisfy the field equations then the total energy $E_{\text{tot}} = E_{\text{kin}} + E$ is independent of t . By Derrick's theorem, static solutions have $E_{\text{mag}} \equiv E_{\text{self}}$ (Manton and Sutcliffe 2004, pp. 82–87).

Configurations with finite energy have quantized total magnetic flux. To see this, note that E finite implies $|\phi| \rightarrow 1$ as $r \rightarrow \infty$, so $\phi \sim e^{i\chi(r, \theta)}$ at large r for some real (in general, multivalued) function χ . The winding number of ϕ is its winding around a circle of large radius R , that is, the integer $n = (\chi(R, 2\pi) - \chi(R, 0))/2\pi$. Although the phase of ϕ is clearly gauge dependent, n is not, because to change this, a gauge transformation $e^{i\Lambda}: \mathbb{R}^2 \rightarrow \text{U}(1)$ would itself need nonzero winding around the circle, contradicting smoothness of $e^{i\Lambda}$. The model is invariant under spatial reflexions, under which $n \mapsto -n$, so we will assume (unless noted otherwise) that $n \geq 0$. Finiteness of E also implies that $D\phi = d\phi - iA\phi \rightarrow 0$, so $A \sim -id\phi/\phi \sim d\chi$ as $r \rightarrow \infty$ (note $\phi \neq 0$ for large r). Hence, the total magnetic flux is

$$\int_{\mathbb{R}^2} B d^2x = \lim_{R \rightarrow \infty} \oint_{S_R} A = \lim_{R \rightarrow \infty} \int_0^{2\pi} \partial_\theta \chi d\theta = 2\pi n \quad [6]$$

where $S_R = \{x: |x| = R\}$ and we have used Stokes's theorem. The above argument uses only generic properties of E , namely that finite E_{self} requires $|\phi|$ to assume a nonzero constant value as $r \rightarrow \infty$. So flux quantization is a robust feature of this type of model. As presented, the argument is somewhat formal, but it can be made mathematically rigorous at the cost of gauge-fixing technicalities (Manton and Sutcliffe 2004, pp. 164–166). Note that if $n \neq 0$ then, by continuity, $\phi(x)$ must vanish at some $x \in \mathbb{R}^2$, and one expects a lump of energy density to be associated with each such x since $\phi = 0$ maximizes the integrand of E_{self} .

Radially Symmetric Vortices

The model supports static solutions within the radially symmetric ansatz $\phi = \sigma(r)e^{in\theta}, A = a(r) d\theta$, which reduces the field equations to a coupled pair of nonlinear ODEs:

$$\begin{aligned} \frac{d^2\sigma}{dr^2} + \frac{1}{r} \frac{d\sigma}{dr} - \frac{1}{r^2} (n-a)^2 \sigma + \frac{\lambda}{2} (1-\sigma^2) \sigma &= 0 \\ \frac{d^2a}{dr^2} - \frac{1}{r} \frac{da}{dr} + (n-a)\sigma^2 &= 0 \end{aligned} \quad [7]$$

Finite energy requires $\lim_{r \rightarrow \infty} \sigma(r) = 1, \lim_{r \rightarrow \infty} a(r) = n$ while smoothness requires $\sigma(r) \sim \text{const}_1 r^n, a(r) \sim$

$\text{const}_2 r^2$ as $r \rightarrow 0$. It is known that solutions to this system, which we shall call n -vortices, exist for all n, λ , though no explicit formulas for them are known. They may be found numerically, and are depicted in Figure 1. Note that σ and a always rise monotonically to their vacuum values, and B always falls monotonically to 0, as r increases. These solutions have their magnetic flux concentrated in a single, symmetric lump, a flux tube in the \mathbb{R}^{3+1} picture. In contrast, the total energy density (integrand of E in [4]) is nonmonotonic for $n \geq 2$, being peaked on a ring whose radius grows with n . This is a common feature of planar solitons.

The large r asymptotics of n -vortices are well understood. For $\lambda \leq 4$ one may linearize [7] about $\sigma = 1, a = n$, yielding

$$\sigma(r) \sim 1 + \frac{q_n}{2\pi} K_0(\sqrt{\lambda}r) \quad [8]$$

$$a(r) \sim n + \frac{m_n}{2\pi} r K_1(r) \quad [9]$$

where q_n, m_n are unknown constants and K_α denotes the modified Bessel's function. For $\lambda > 4$ linearization is no longer well justified, and the asymptotic behaviour of σ (though not a) is quite different (Manton and Sutcliffe 2004, pp. 174–175). We shall not consider this rather extreme regime further. Note that

$$K_\alpha(r) \sim \sqrt{\frac{\pi}{2r}} e^{-r} \quad \text{as } r \rightarrow \infty \quad [10]$$

for all α , so both σ and a approach their vacuum values exponentially fast, but with different decay lengths: $1/\sqrt{\lambda}$ for σ , 1 for a . This can be seen in Figure 1a. The constants q_n and m_n depend on λ and must be inferred by comparing the numerical solutions with [8], [9]; $q = q_1$ and $m = m_1$ will receive a physical interpretation shortly.

The 1-vortex (henceforth just ‘‘vortex’’) is stable for all λ , but n -vortices with $n \geq 2$ are unstable to break up into n separate vortices if $\lambda > 1$. We shall say that the AHM is type I if $\lambda < 1$, type II if $\lambda > 1$, and critically coupled if $\lambda = 1$, based on this distinction. Let E_n denote the energy of an n -vortex. Figure 2 shows the energy per vortex E_n/n plotted against n for $\lambda = 0.5, 1$, and 2. It decreases with n for $\lambda = 0.5$, indicating that it is energetically favorable for isolated vortices to coalesce into higher winding lumps. For $\lambda = 2$, by contrast, E_n/n increases with n indicating that it is energetically favorable for n -vortices to fission into their constituent vortex parts. The case $\lambda = 1$ balances between these behaviors: E_n/n is independent of n . In fact, the energy of a collection of vortices is independent of their positions in this case.

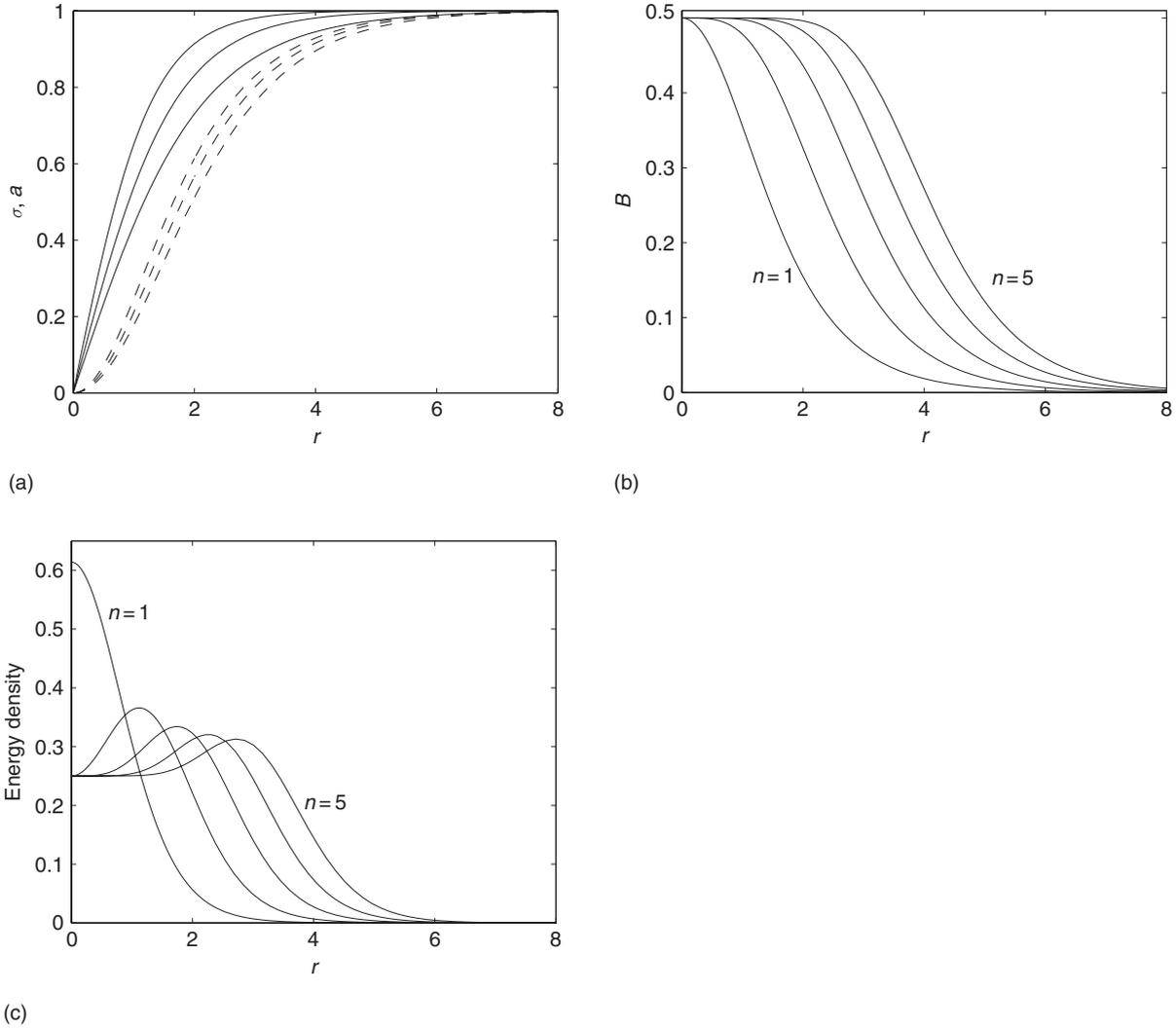


Figure 1 Static, radially symmetric n -vortices: (a) the 1-vortex profile functions $\sigma(r)$ (solid curve) and $a(r)$ (dashed curve) for $\lambda = 2, 1$, and $1/2$, left to right; (b) the magnetic field B ; and (c) the energy density of n -vortices, $n = 1$ to 5, left to right, for $\lambda = 1$.

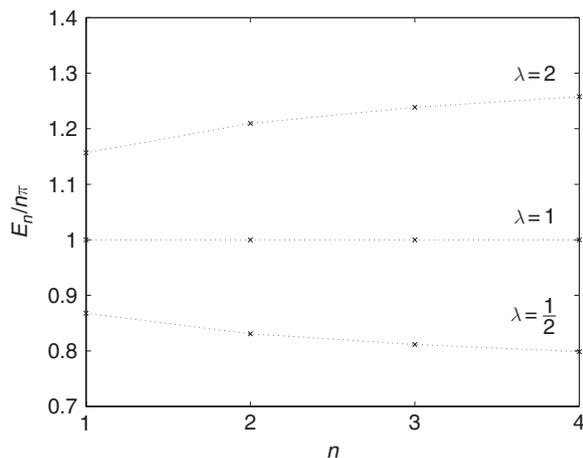


Figure 2 The energy per unit winding E_n/n of radially symmetric n -vortices for $\lambda = 1/2, 1$, and 2.

Interaction Energy

A precise understanding of the type III dichotomy can be obtained using the 2-vortex interaction energy $E_{\text{int}}(s)$ introduced by Jacobs and Rebbi. This is defined to be the minimum of E over all $n=2$ configurations for which $\phi(\mathbf{x})=0$ at some pair of points $\mathbf{x}_1, \mathbf{x}_2$ distance s apart. One interprets $\mathbf{x}_1, \mathbf{x}_2$ as the vortex positions. E_{int} can only depend on their separation $s = |\mathbf{x}_1 - \mathbf{x}_2|$, by translation and rotation invariance. **Figure 3** presents graphs of $E_{\text{int}}(s)$ generated by a lattice minimization algorithm. For $\lambda < 1$, vortices uniformly attract one another, so a vortex pair has least energy when coincident. For $\lambda > 1$, vortices uniformly repel, always lowering their energy by moving further apart. The graph for $\lambda = 1$ would be a horizontal line, $E_{\text{int}}(s) = 2\pi$.

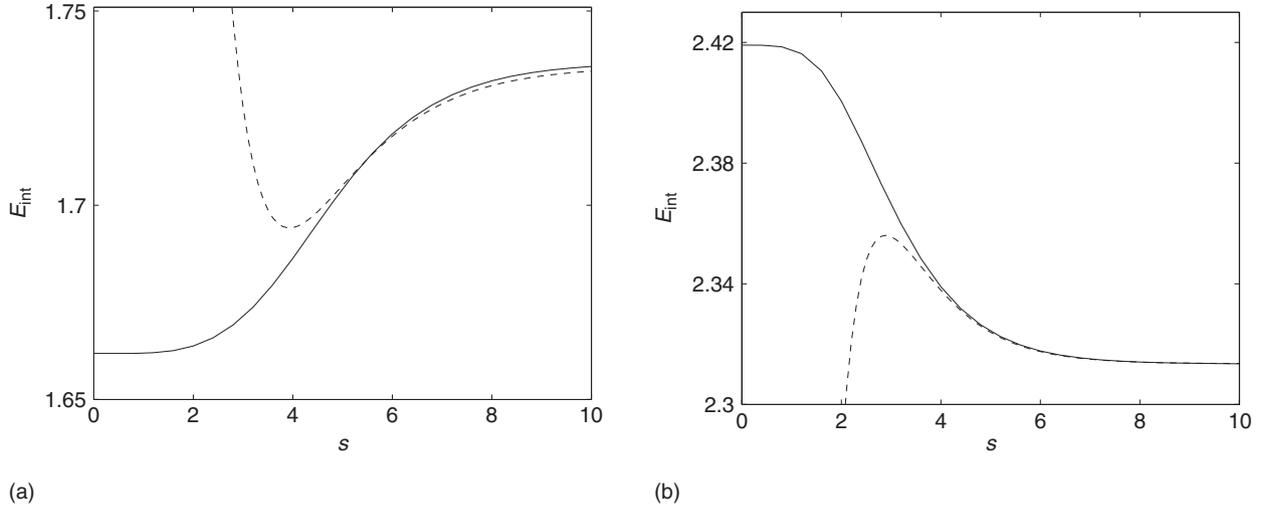


Figure 3 The 2-vortex interaction energy $E_{\text{int}}(s)$ as a function of vortex separation (solid curve), in comparison with its asymptotic form $E_{\text{int}}^{\infty}(s)$ (dashed curve) for (a) $\lambda=1/2$ and (b) $\lambda=2$.

The large s behavior of $E_{\text{int}}(s)$ is known, and can be understood in two ways (Manton and Sutcliffe 2004, pp. 177–181). Speight, adapting ideas of Manton on asymptotic monopole interactions, observed that, in the real ϕ gauge ($\phi \mapsto e^{-i\theta}\phi$, $A \mapsto A - d\theta$), the difference between the vortex and the vacuum $\phi=1, A=0$ at large r ,

$$\psi = \phi - 1 \sim \frac{q}{2\pi} K_0(\sqrt{\lambda}r) \quad [11]$$

$$(A_0, \mathbf{A}) \sim \frac{m}{2\pi} (0, \mathbf{k} \times \nabla K_0(r)) \quad [12]$$

is identical to the solution of a linear Klein–Gordon–Proca theory,

$$(\partial_\mu \partial^\mu + \lambda)\psi = \kappa, \quad (\partial_\mu \partial^\mu + 1)A_\nu = j_\nu \quad [13]$$

in the presence of a composite point source,

$$\kappa = q\delta(\mathbf{x}), \quad (j_0, \mathbf{j}) = m(0, \mathbf{k} \times \nabla\delta(\mathbf{x})) \quad [14]$$

located at the vortex position. Viewed from afar, therefore, a vortex looks like a point particle carrying both a scalar monopole charge q and a magnetic dipole moment m , a “point vortex,” inducing a real scalar field of mass $\sqrt{\lambda}$ (the Higgs particle) and a vector boson field of mass 1 (the “photon”). If physics is to be model independent, therefore, the interaction energy of a pair of well-separated vortices should approach that of the corresponding pair of point vortices as the separation grows. Computing the latter is an easy exercise in classical linear field theory, yielding

$$E_{\text{int}}(s) \sim E_{\text{int}}^{\infty}(s) = 2E_1 - \frac{q^2}{2\pi} K_0(\sqrt{\lambda}s) + \frac{m^2}{2\pi} K_0(s) \quad [15]$$

Bettencourt and Rivers obtained the same formula by a more direct superposition ansatz approach, though they did not give the constants q, m a physical interpretation.

The force between a well-separated vortex pair, $-E_{\text{int}}'(s)$, consists of the mutual attraction of identical scalar monopoles, of range $1/\sqrt{\lambda}$, and the mutual repulsion of identical magnetic dipoles, of range 1. If $\lambda < 1$, scalar attraction dominates at large s so vortices attract. If $\lambda > 1$, magnetic repulsion dominates and they repel. If $\lambda=1$ then $q \equiv m$, as we shall see, so the forces cancel exactly. Figure 3 shows both E_{int} and E_{int}^{∞} for $\lambda=0.5, 2$. The agreement is good for s large, but breaks down for $s < 4$, as one expects. Vortices are not point particles, as in the linear model, and when they lie close together the overlap of their cores produces significant effects.

The same method predicts the interaction energy between an n_1 -vortex and an n_2 -vortex at large separation. We just replace $2E_1$ by $E_{n_1} + E_{n_2}$, q^2 by $q_{n_1}q_{n_2}$, and m^2 by $m_{n_1}m_{n_2}$. In particular, an antivortex ((-1) -vortex) has $E_{-1} = E_1, q_{-1} = q_1 = q$, and $m_{-1} = -m_1 = -m$, so the interaction energy for a vortex–antivortex pair is

$$E_{\text{int}}^{v\bar{v}}(s) \sim 2E_1 - \frac{q^2}{2\pi} K_0(\sqrt{\lambda}r) - \frac{m^2}{2\pi} K_0(r) \quad [16]$$

which is uniformly attractive. It would be pleasing if q_n, m_n could be deduced easily from q, m . One might guess $q_n = |n|q, m_n = nm$, in analogy with monopoles. Unfortunately, this is false: q_n, m_n grow approximately exponentially with $|n|$.

Vortex Scattering

The AHM being Lorentz invariant, one can obtain time-dependent solutions wherein a single n -vortex travels at constant velocity, with speed $0 < v < 1$ and $E_{\text{tot}} = (1 - v^2)^{-1/2} E_n$, by Lorentz boosting the static solutions described above. Of more dynamical interest are solutions in which two or more vortices undergo relative motion. The simplest problem is vortex scattering. Two vortices, initially well separated, are propelled towards one another. In the center-of-mass (COM) frame they have, as $t \rightarrow -\infty$, equal speed v , and approach one another along parallel lines distance b (the impact parameter) apart, see **Figure 4**. If $b = 0$, they approach head-on. Assuming they do not capture one another, they interact and, as $t \rightarrow \infty$, recede along parallel straight lines having been deflected through an angle Θ (the scattering angle). If scattering is elastic, the exit lines also lie b apart and each vortex travels at speed v as $t \rightarrow \infty$. The dependence of Θ on v, b , and λ has been studied through lattice simulations by several authors, perhaps most comprehensively by [Myers, Rebbi, and Strilka \(1992\)](#). We shall now describe their results.

Note first that vortex scattering is actually inelastic: vortices recede with speed $< v$ because some of their initial kinetic energy is dispersed by the collision as small-amplitude traveling waves (“radiation”). This energy loss can be as high as 80% in very fast collisions at small b . At small v the energy loss is tiny, but can still have important consequences for type I vortices: if v is very small, they start with only just enough energy to escape their mutual attraction. In undergoing a small b collision they can lose enough of this energy to become trapped in an oscillating bound state. In this case they do not truly scatter and Θ is ill-defined. [Myers *et al.*](#) find that $v \geq 0.2$ suffices to avoid

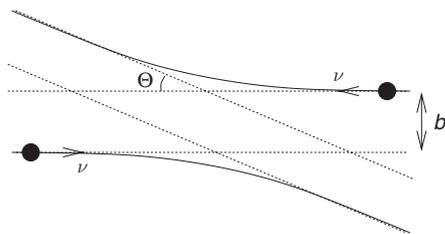


Figure 4 The geometry of vortex scattering.

capture when $\lambda = 1/2$. Since type I vortices attract, one might expect Θ to be always negative, indicating that the vortices deflect towards one another. In fact, as **Figure 5a** shows, this happens only for small v and large b . Another naive expectation is that $\Theta = 0$ or $\Theta = 180^\circ$ when $b = 0$ (either vortices pass through one another or ricochet backwards in a head-on collision). In fact $\Theta = 90^\circ$, the only other possibility allowed by reflexion symmetry of the initial data. **Figure 6** depicts snapshots of such a scattering process at modest v . The vortices deform each other as they get close until, at the moment of coincidence, they are close to the static 2-vortex ring. They then break apart along a line perpendicular to their line of approach. One may consider them to have exchanged half-vortices, so that each emergent vortex is a mixture of the incoming vortices. This rather surprising phenomenon was actually predicted by Ruback in advance of any numerical simulations and turns out to be a generic feature of planar topological solitons.

Consider now the type II case ($\lambda = 2$, **Figure 5b**). Here, $\Theta > 0$ for all v, b as one expects of particles that repel each other. Head-on scattering is more interesting now since two regimes emerge: for $v > v_{\text{crit}} \approx 0.3$, one has the surprising 90° scattering already described, while for $v < v_{\text{crit}}$ the vortices bounce backwards, $\Theta = 180^\circ$. This is easily explained. In order to undergo 90° head-on scattering, the vortices must become coincident (otherwise reflexion symmetry is violated), hence must have initial energy at least E_2 . For $v < v_{\text{crit}}$, where

$$\frac{2E_1}{\sqrt{1 - v_{\text{crit}}^2}} = E_2 \quad [17]$$

they have too little energy, so come to a halt before coincidence, then recede from one another. The solution v_{crit} of [17] depends on λ and is plotted in **Figure 7**. For v slightly above v_{crit} , we see that, in contrast to the type I case, $\Theta(b)$ is not monotonic: maximum deflection occurs at nonzero b .

The point vortex formalism yields a simple model of type II vortex scattering which is remarkably successful at small v . One writes down the Lagrangian for two identical (nonrelativistic) point particles of mass E_1 moving along trajectories $\mathbf{x}_1(t), \mathbf{x}_2(t)$ under the influence of the repulsive potential E_{int}^∞ ,

$$L = \frac{1}{2} E_1 (|\dot{\mathbf{x}}_1|^2 + |\dot{\mathbf{x}}_2|^2) - E_{\text{int}}^\infty (|\mathbf{x}_1 - \mathbf{x}_2|) \quad [18]$$

Energy and angular momentum conservation reduce $\Theta(v, b)$ to an integral over one variable ($s = |\mathbf{x}_1 - \mathbf{x}_2|$) which is easily computed numerically. To illustrate, **Figure 5b** shows the result for $\lambda = 2, v = 0.1$ in comparison with the lattice simulations of

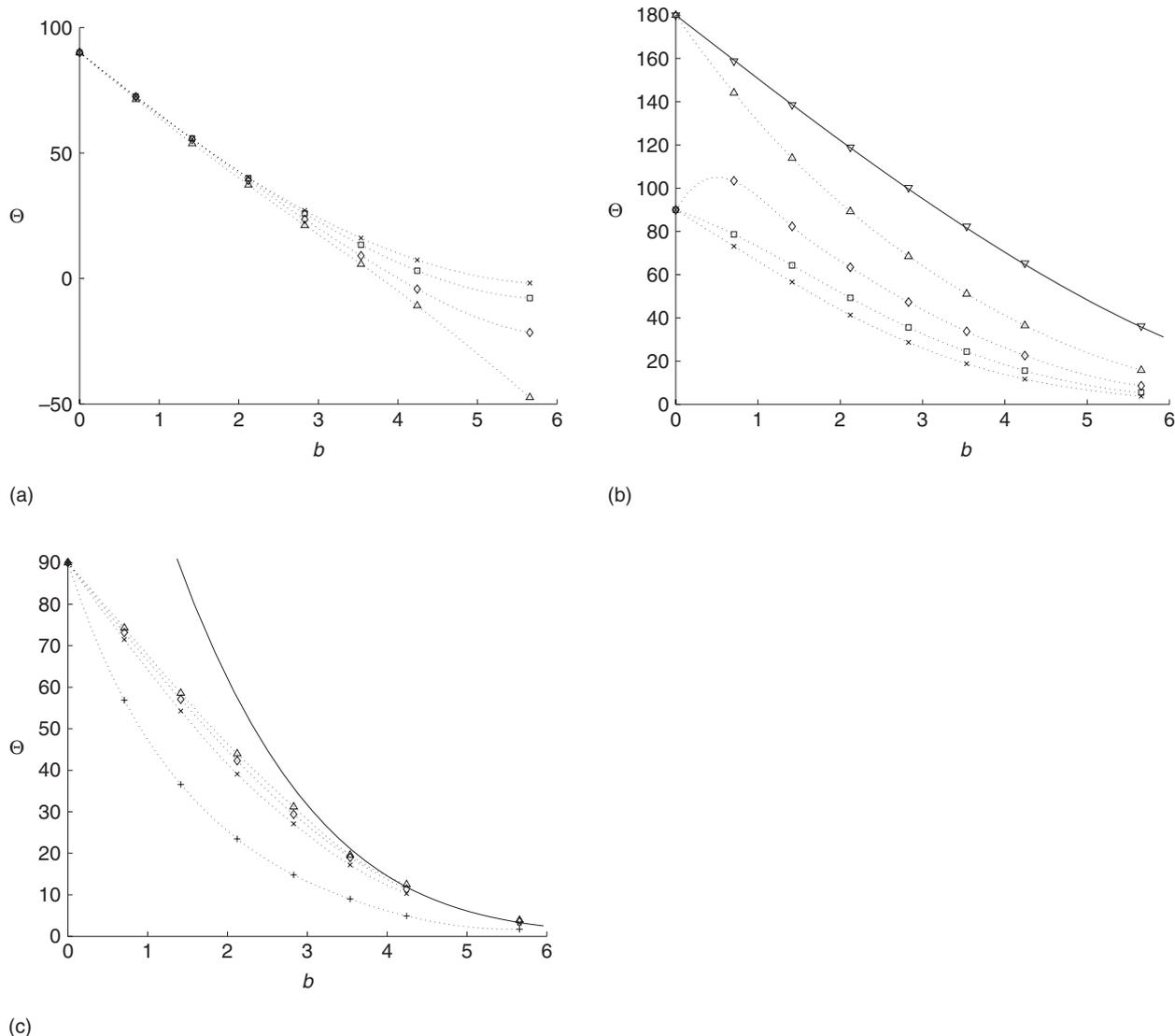


Figure 5 The 2-vortex scattering angle Θ as a function of impact parameter b for $\nu=0.1$ (∇), $\nu=0.2$ (Δ), $\nu=0.3$ (\diamond), $\nu=0.4$ (\approx), $\nu=0.5$ (\times), and $\nu=0.9$ ($+$), as computed by Myers *et al.* (1992): (a) $\lambda=1/2$; (b) $\lambda=2$; (c) $\lambda=1$. The dotted curves are merely guides to the eye. The solid curves in (b), (c) were computed using the point vortex model. Note that Myers *et al.* use different normalizations, so $b = \sqrt{2}b_{MRS}$ and $\lambda = \lambda_{MRS}/2$.

Myers *et al.* The agreement is almost perfect. For large ν the approximation breaks down not only because relativistic corrections become significant, but also because small b collisions then probe the small $|\mathbf{x}_1 - \mathbf{x}_2|$ region where vortex core overlap effects become important. For the same reason, the point vortex model is less useful for type I scattering. Here there is no repulsion to keep the vortices well separated, so its validity is restricted to the small ν , large b regime.

Critical coupling is theoretically the most interesting regime, where most analytic progress has been made. Since $E_{\text{int}} \equiv E_{\text{int}}^\infty \equiv 0$, one might expect vortex scattering to be trivial ($\Theta(\nu, b) \equiv 0$), but this is quite wrong, as shown in Figure 5c. In particular,

$\Theta(\nu, 0) = 90^\circ$ for all ν , just as in the large ν type I and type II cases. The point is that scalar attraction and magnetic repulsion of vortices are mediated by fields with different Lorentz transformation properties. While they cancel for static vortices, there is no reason to expect them to cancel for vortices in relative motion.

Critical Coupling

The AHM with $\lambda=1$ has many remarkable properties, at which we have so far only hinted. These all stem from Bogomol'nyi's crucial observation (Manton and Sutcliffe 2004, pp. 197–202) that the potential energy in this case can be rewritten as

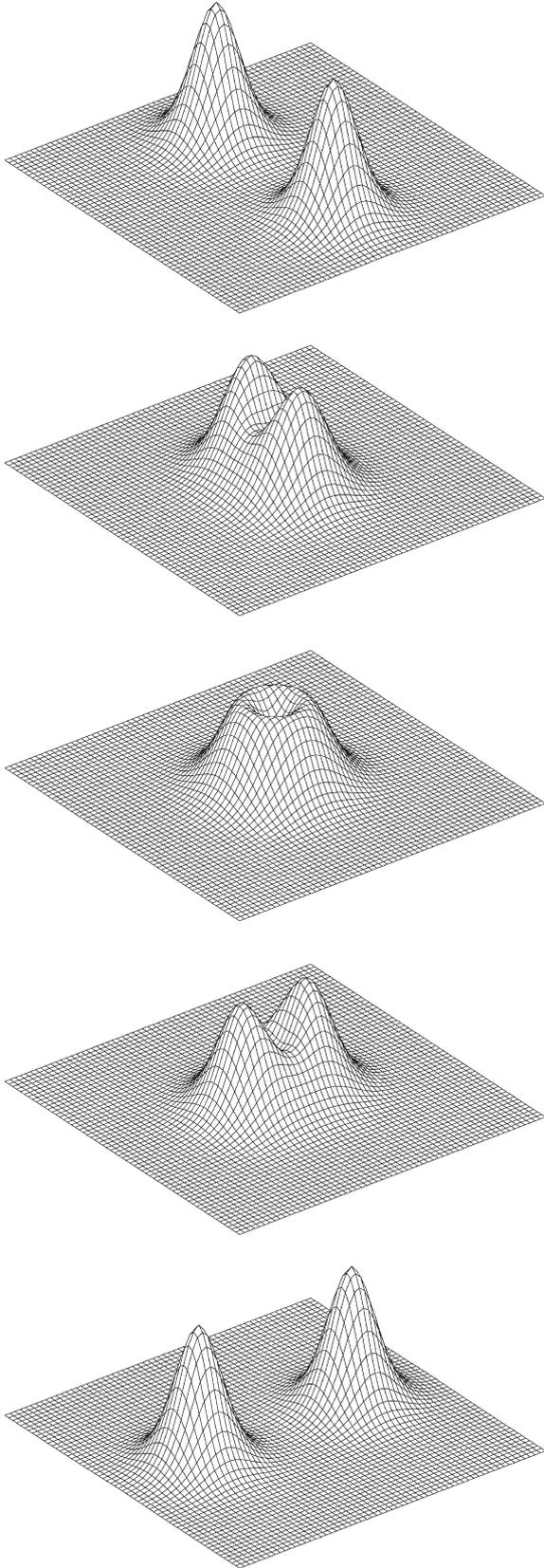


Figure 6 Snapshots of the energy density during a head-on collision of vortices. This 90° scattering phenomenon is a generic feature of planar topological soliton dynamics.

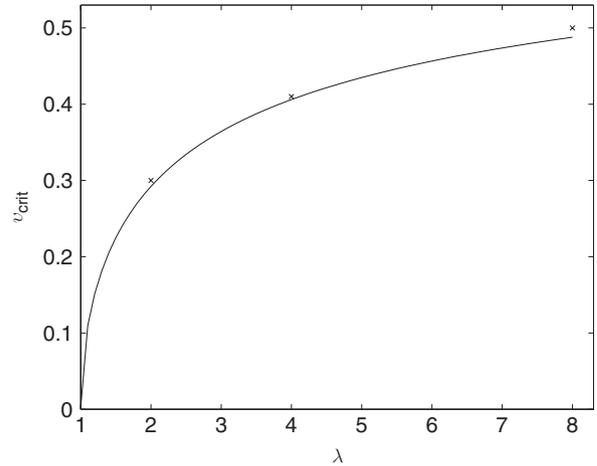


Figure 7 The critical velocity for 90° head-on scattering of type II vortices v_{crit} as a function of λ , as predicted by equation [17] (solid curve), in comparison with the results of Myers *et al.* (1992), (crosses).

$$E = \frac{1}{2} \int \left\{ \left(B - \frac{1}{2}(1 - |\phi|^2) \right)^2 + |D_1\phi + iD_2\phi|^2 + B \right\} d^2\mathbf{x} - i \int_{\mathbb{R}^2} d(\bar{\phi}D\phi) \quad [19]$$

The last integral vanishes by Stokes's theorem, so $E \geq \pi n$ by flux quantization [6], and $E = \pi n$ if and only if

$$(D_1 + iD_2)\phi = 0 \quad [20]$$

$$\frac{1}{2}(1 - |\phi|^2) = B \quad [21]$$

Note that system [20], [21] is first order, in contrast to the second-order field equations [3]. No explicit solutions of [20], [21] are known. However, Taubes has proved that for each unordered list $[z_1, z_2, \dots, z_n]$ of n points in \mathbb{C} , not necessarily distinct, there exists a solution of [20], [21], unique up to gauge transformations, with $\phi(z_1) = \phi(z_2) = \dots = \phi(z_n) = 0$ and ϕ nonvanishing elsewhere, the zero at z_r having the same multiplicity as z_r has in the list. Note that the list is unordered: a solution is uniquely determined by the positions and multiplicities of the zeroes of ϕ , but the order in which we label these is irrelevant. The solution minimizes E within the class \mathcal{C}_n of winding n configurations, so is automatically a stable static solution of the model.

Equation [20] applied to the symmetric n -vortex, $\phi = \sigma(r) e^{in\theta}$, $A = a(r) d\theta$ implies $a(r) = n - r\sigma'(r)/\sigma(r)$. Comparing with [8], [9], it follows that $q_n = m_n$ when $\lambda = 1$ as previously claimed, since $K_1 = -K'_0$. Tong has conjectured, based on a string duality argument, that $q_1 = -2\pi 8^{1/4}$. This is consistent with current numerics but has no direct derivation so far.

Taubes's theorem shows that this n -vortex is just one point, corresponding to the list $[0, 0, \dots, 0]$, in a $2n$ -dimensional space of static multivortex solutions called the moduli space \mathbf{M}_n . This space may be visualized as the flat, finite-dimensional valley bottom in \mathcal{C}_n on which E attains its minimum value, πn . Points in \mathbf{M}_n are in one-to-one correspondence with distinct unordered lists $[z_1, z_2, \dots, z_n]$, which are themselves in one-to-one correspondence with points in \mathbb{C}^n , as follows. To each list, we assign the unique monic polynomial whose roots are z_r ,

$$\begin{aligned} p(z) &= (z - z_1)(z - z_2) \cdots (z - z_n) \\ &= a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n \end{aligned} \quad [22]$$

This polynomial is uniquely determined by its coefficients $(a_0, a_1, \dots, a_{n-1}) \in \mathbb{C}^n$, which give good global coordinates on $\mathbf{M}_n \cong \mathbb{C}^n$. The zeros z_r of ϕ may be used as local coordinates on \mathbf{M}_n , away from Δ , the subset of \mathbf{M}_n on which two or more of the zeros z_r coincide, but are not good global coordinates.

Let $(\phi, \mathbf{A})_a$ denote the static solution corresponding to $\mathbf{a} \in \mathbb{C}^n$. If the zeros z_r are all at least s apart, Taubes showed the solution is just a linear superposition of 1-vortices located at z_r , up to corrections exponentially small in s . Imagine these constituent vortices are pushed with small initial velocities. Then $(\phi(t), \mathbf{A}(t))$ must remain close to the valley bottom \mathbf{M}_n , since departing from it costs kinetic energy, of which there is little. Manton has suggested, therefore, that the dynamics is well approximated by the constrained variational problem wherein $(\phi(t), \mathbf{A}(t)) = (\phi, \mathbf{A})_{a(t)} \in \mathbf{M}_n$ for all t . Since the action $S = \int \mathcal{L} d^3x = \int (E_{\text{kin}} - E) dt$, and $E = \pi n$, constant, on \mathbf{M}_n , this constrained problem amounts to Lagrangian mechanics on configuration space \mathbf{M}_n with Lagrangian $L = E_{\text{kin}}|_{\mathbf{M}_n}$. Now E_{kin} is real, positive, and quadratic in time derivatives of ϕ, \mathbf{A} , so

$$L = \frac{1}{2} \gamma_{rs}(\mathbf{a}) \dot{a}_r \dot{a}_s \quad [23]$$

γ_{rs} forming the entries of a positive-definite $n \times n$ Hermitian matrix ($\gamma_{sr} \equiv \overline{\gamma_{rs}}$). Since $(\phi, \mathbf{A})_a$ is not known explicitly, neither are $\gamma_{rs}(\mathbf{a})$. Observe, however, that L is the Lagrangian for geodesic motion in \mathbf{M}_n with respect to the Riemannian metric

$$\gamma = \gamma_{rs}(\mathbf{a}) da_r d\bar{a}_s \quad [24]$$

Manton originally proposed this geodesic approximation for monopoles, but it is now standard for all topological solitons of Bogomol'nyi type (where one has a moduli space of static multisolitons saturating a topological lower bound on E). Note that geodesics are independent of initial speed, which agrees with Myers *et al.*: [Figure 5c](#) shows that $\Theta(v, b)$

is approximately independent of v for $v \leq 0.5$. Further, [Stuart \(1994\)](#) has proved that, for initial speeds of order ϵ , small, the fields stay (pointwise) ϵ^2 close to their geodesic approximant for times of order ϵ^{-1} .

On symmetry grounds, two vortex dynamics in the COM frame reduces to geodesic motion in $\mathbf{M}_2^0 \cong \mathbb{C}$, the subspace of centered 2-vortices ($a_1 = 0$, so $z_1 = -z_2$), with induced metric

$$\gamma^0 = G(|a_0|) da_0 d\bar{a}_0 \quad [25]$$

G being some positive function. Note that $a_0 = z_1 z_2$, so the intervortex distance $|z_1 - z_2| = 2|z_1| = 2|a_0|^{1/2}$. The line $a_0 = \beta \in \mathbb{R}$, traversed with β increasing, say, is geodesic in \mathbf{M}_2^0 . The vortex positions (roots of $z^2 + a_0$) are $\pm\sqrt{|\beta|}$ for $\beta \leq 0$ and $\pm i\sqrt{\beta}$ for $\beta > 0$. This describes perfectly the 90° scattering phenomenon: two vortices approach head-on along the x^1 axis, coincide to form a 2-vortex ring, then break apart along the x^2 axis, as in [Figure 6](#). This behavior occurs because $a_0 = z_1 z_2$, rather than $z_1 - z_2$, is the correct global coordinate on \mathbf{M}_2^0 , since vortices are classically indistinguishable.

Samols found a useful formula ([Manton and Sutcliffe 2004](#), pp. 205–215) for γ in terms of the behavior of $|\phi_a|$ close to its zeros, using which he devised an efficient numerical scheme to evaluate $G(|a_0|)$, and computed $\Theta(b)$ in detail, finding excellent agreement with lattice simulations at low speeds. He also studied the quantum scattering of vortices, approximating the quantum state by a wave function Ψ on \mathbf{M}_n evolving according to the natural Schrödinger equation for quantum geodesic motion,

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{1}{2} \hbar^2 \Delta_\gamma \Psi \quad [26]$$

where Δ_γ is the Laplace–Beltrami operator on (\mathbf{M}_n, γ) . This technique, introduced for monopoles by Gibbons and Manton, is now standard for solitons of Bogomol'nyi type.

By analyzing the forces between moving point vortices at $\lambda=1$, [Manton and Speight \(2003\)](#) showed that, as the vortex separations become uniformly large, the metric on \mathbf{M}_n approaches

$$\begin{aligned} \gamma^\infty &= \pi \sum_r \left[dz_r d\bar{z}_r - \frac{q^2}{4\pi} \sum_{s \neq r} K_0(|z_r - z_s|) \right. \\ &\quad \left. \times (dz_r - dz_s)(d\bar{z}_r - d\bar{z}_s) \right] \end{aligned} \quad [27]$$

This formula can also be obtained by a method of matched asymptotic expansions. We can use [\[27\]](#) to study 2-vortex scattering for large b , when the

vortices remain well separated. (Note that γ^∞ is not positive definite if any $|z_r - z_s|$ becomes too small.) The results are good, provided $\nu \leq 0.5$ and $b \geq 3$ (see Figure 5c).

Other Developments

The (critically coupled) AHM on a compact physical space Σ is of considerable theoretical and physical interest. Bradlow showed that $M_n(\Sigma)$ is empty unless $V = \text{Area}(\Sigma) \geq 4\pi n$, so there is a limit to how many vortices a space of finite area can accommodate (Manton and Sutcliffe 2004, pp. 227–230). Manton has analyzed the thermodynamics of a gas of vortices by studying the statistical mechanics of geodesic flow on $M_n(\Sigma)$. In this context, spatial compactness is a technical device to allow nonzero vortex density n/V for finite n , without confining the fields to a finite box, which would destroy the Bogomol’nyi properties. In the limit of interest, $n, V \rightarrow \infty$ with n/V fixed, the thermodynamical properties turn out to depend on Σ only through V , so $\Sigma = S^2$ and $\Sigma = T^2$ give equivalent results, for example. The equation of state of the gas is ($P = \text{pressure}$, $T = \text{temperature}$)

$$P = \frac{nT}{V - 4\pi n} \quad [28]$$

which is similar, at low density n/V , to that of a gas of hard disks of area 2π . The crucial step in deriving [28] is to find the volume of $M_n(\Sigma)$ which, despite there being no formula for γ , may be computed exactly by remarkable indirect arguments (Manton and Sutcliffe 2004, pp. 231–234).

The static AHM coincides with the Ginzburg–Landau model of superconductivity, which has precisely the same type I/II classification. Here the “Higgs” field represents the wave function of a condensate of Cooper pairs, usually (but not always) electrons. There has been a parallel development of the static model by condensed matter theorists, therefore; see Fossheim and Sudbo (2004), for example. In fact the vortex was actually first discovered by Abrikosov in the condensed matter context. One important difference is that type I superconductors do not support vortex solutions in an external magnetic field B_{ext} because the critical $|B_{\text{ext}}|$ required to create a single vortex is greater than the critical $|B_{\text{ext}}|$ required to destroy the condensate completely ($\phi \equiv 0$). Type II superconductors do support vortices, and there are such superconductors with $\lambda \approx 1$, but the vortex dynamics we have described is not relevant to these systems. In this context there is an obvious preferred

reference frame (the rest frame of the superconductor) so it is unsurprising that the Lorentz-invariant AHM is inappropriate. Insofar as vortices move at all, they seem to obey a first-order (in time) dynamical system, in contrast to the second-order AHM. Manton has devised a first-order system which may have relevance to superconductivity, by replacing E_{kin} with a Chern–Simons–Schrödinger functional (Manton and Sutcliffe 2004, pp. 193–197). Rather than attracting or repelling, vortices now tend to orbit one another at constant separation. There is again a moduli space approximation to slow vortex dynamics for $\lambda \approx 1$, but it has a Hamiltonian-mechanical rather than Riemannian-geometric flavor.

Finally, an interesting simplification of the AHM, which arises, for example, as a phenomenological model of liquid helium-4, is obtained if we discard the gauge field A_μ , or equivalently set the electric charge of ϕ to $e = 0$. There is now no type I/II classification, since λ may be absorbed by rescaling. The resulting model, which has only global $U(1)$ phase symmetry, supports n -vortices $\phi = \sigma(r)e^{in\theta}$ for all n , but these are not exponentially spatially localized,

$$\sigma(r) = 1 - \frac{n^2}{\lambda r^2} - \frac{n^2(8 + n^2)}{2\lambda^2 r^4} + O(r^{-6}) \quad [29]$$

and cannot have finite E by Derrick’s theorem. They are unstable for $|n| > 1$, and 1-vortices uniformly repel one another. They can be given an interesting first-order dynamics (the Gross–Pitaevski equation).

Abbreviations

A_μ	electromagnetic gauge potential
b	impact parameter
D_μ	gauge-covariant derivative
E	potential energy
E_{kin}	kinetic energy
$F_{\mu\nu}$	electromagnetic field strength tensor
L	Lagrangian
\mathcal{L}	Lagrangian density
S	action
ϕ	Higgs field
Θ	scattering angle

See also: Fractional Quantum Hall Effect; Ginzburg–Landau Equation; High T_c Superconductor Theory; Integrable Systems: Overview; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Fields with Topological Defects; Solitons and Other Extended Field Configurations; Symmetry Breaking in Field Theory; Topological Defects and Their Homotopy Classification; Variational Techniques for Ginzburg–Landau Energies.

Further Reading

- Atiyah M and Hitchin N (1988) *The Geometry and Dynamics of Magnetic Monopoles*. Princeton: Princeton University Press.
- Fosshem K and Sudbo A (2004) *Superconductivity: Physics and Applications*. Hoboken NJ: Wiley.
- Jaffe A and Taubes C (1980) *Vortices and Monopoles: Structure of Static Gauge Theories*. Boston: Birkhäuser.
- Nakahara M (1990) *Geometry, Topology and Physics*. Bristol: Adam-Hilger.
- Manton NS and Speight JM (2003) Asymptotic interactions of critically coupled vortices. *Communications in Mathematical Physics* 236: 535–555.

- Manton NS and Sutcliffe PM (2004) *Topological Solitons*. Cambridge: Cambridge University Press.
- Myers E, Rebbi C, and Strilka R (1992) Study of the interaction and scattering of vortices in the abelian Higgs (or Ginzburg-Landau) model. *Physical Review* 45: 1355–1364.
- Rajaraman R (1989) *Solitons and Instantons*. Amsterdam: North-Holland.
- Stuart D (1994) Dynamics of abelian Higgs vortices in the near Bogomolny regime. *Communications in Mathematical Physics* 159: 51–91.
- Vilenkin A and Shellard EPS (1994) *Cosmic Strings and Other Topological Defects*. Cambridge: Cambridge University Press.

Adiabatic Piston

Ch Gruber, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

A Lesne, Université P.-M. Curie, Paris VI, Paris, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Macroscopic Problem

The “adiabatic piston” is an old problem of thermodynamics which has had a long and controversial history. It is the simplest example concerning the time evolution of an adiabatic wall, that is, a wall which does not conduct heat. The system consists of a gas in a cylinder divided by an adiabatic wall (the piston). Initially, the piston is held fixed by a clamp and the two gases are in thermal equilibrium characterized by (p^\pm, T^\pm, N^\pm) , where the index $-/+$ refers to the gas on the left/right side of the piston and (p, T, N) denote the pressure, the temperature, and the number of particles (**Figure 1**). Since the piston is adiabatic, the whole system remains in equilibrium even if $T^- \neq T^+$. At time $t = 0$, the clamp is removed and the piston is let free to move without any friction in the cylinder. The

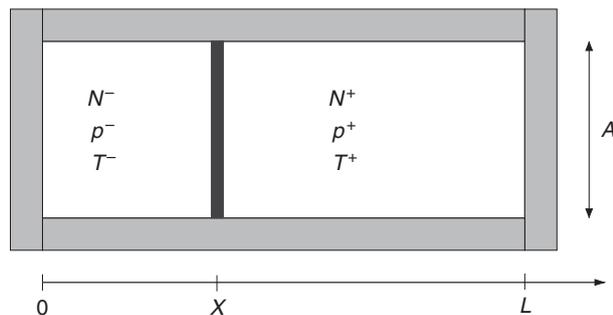


Figure 1 The adiabatic piston problem.

question is to find the final state, that is, the final position X_f of the piston and the parameters (p_f^\pm, T_f^\pm) of the gases.

In the late 1950s, using the two laws of equilibrium thermodynamics (i.e., thermostatics), Landau and Lifshitz concluded that the adiabatic piston will evolve toward a final state where $p^-/T^- = p^+/T^+$. Later, Callen (1963) and others realized that the maximum entropy condition implies that the system will reach mechanical equilibrium where the pressures are equal $p_f^- = p_f^+$; however, nothing could be said concerning the final position X_f or the final temperatures T_f^\pm which should depend explicitly on the viscosity of the fluids. It thus became a controversial problem since one was forced to accept that the two laws of thermostatics are not sufficient to predict the final state as soon as adiabatic movable walls are involved (see early references in Gruber (1999)).

Experimentally, the adiabatic piston was used already before 1924 to measure the ratio c_p/c_v of the specific heats of gases. In 2000, new measurements have shown that one has to distinguish between two regimes, corresponding to weak damping or strong damping, with very different properties, for example, for weak damping the frequency of oscillations corresponds to adiabatic oscillations, whereas for strong damping it corresponds to isothermal oscillations.

Microscopic Problem

The “adiabatic piston” was first considered from a microscopic point of view by Lebowitz who introduced in 1959 a simple model to study heat conduction. In this model, the gas consists of point particles of mass m making purely elastic collisions on the wall of the cylinder and on the piston. Furthermore, the gas is very dilute so that the

equation of state $p = nk_B T$ is satisfied at equilibrium, where n is the density of particles in the gas and k_B the Boltzmann constant. The adiabatic piston is taken as a heavy particle of mass $M \gg m$ without any internal degree of freedom. Using this same model Feynman (1965) gave a qualitative analysis in *Lectures in Physics*. He argued intuitively but correctly that the system should converge first toward a state of mechanical equilibrium where $p^- = p^+$ and then very slowly toward thermal equilibrium. This approach toward thermal equilibrium is associated with the “wiggles” of the piston induced by the random collisions with the atoms of the gas. Of course, this stochastic behavior is not part of thermodynamics and the evolution beyond the mechanical equilibrium cannot appear in the macroscopical framework assuming that the piston does not conduct heat.

From a microscopical point of view, one is confronted with two different problems: the approach toward mechanical equilibrium in the absence of any *a priori* friction (where the entropy of both gases should increase) and, on a different timescale, the approach toward thermal equilibrium (where the entropy of one gas should decrease but the total entropy increase).

The conceptual difficulties of the problem beyond mechanical equilibrium come from the following intuitive reasoning. When the piston moves toward the hotter gas, the atoms of the hotter gas gain energy, whereas those of the cooler gas lose energy. When the piston moves toward the cooler side, it is the opposite. Since on an average the hotter side should cool down and the cold side should warm up, we are led to conclude that on an average the piston should move toward the colder side. On the other hand, from $p = nk_B T$, the piston should move toward the warmer side to maintain pressure balance.

In 1996, Crosignani, Di Porto, and Segev introduced a kinetic model to obtain equations describing the adiabatic approach toward mechanical equilibrium. Starting with the microscopical model introduced by Lebowitz, Gruber, Piasecki, and Frachebourg, later joined by Lesne and Pache, initiated in 1998 a systematic investigation of the adiabatic piston within the framework of statistical mechanics, together with a large number of numerical simulations. This analysis was based on the fact that m/M is a very small parameter to investigate expansions in powers of m/M (see Gruber and Piasecki (1999) and Gruber *et al.* (2003) and reference therein). An approach using dynamical system methods was then developed by Lebowitz *et al.* (2000) and Chernov *et al.* (2002). An

extension to hard-disk particles was analyzed at the same time by Kestemont *et al.* (2000). Recently, several other authors have contributed to this subject.

The general picture which emerges from all the investigations is the following. For an infinite cylinder, starting with mechanical equilibrium $p^- = p^+ = p$, the piston evolves to a stationary stochastic state with nonzero velocity toward the warmer side

$$\langle V \rangle = \frac{m}{M} \sqrt{\frac{\pi k_B}{8m}} (\sqrt{T^+} - \sqrt{T^-}) + o\left(\frac{m}{M}\right) \quad [1]$$

with relaxation time

$$\tau = \frac{M}{A} \sqrt{\frac{\pi k_B}{8m}} \frac{1}{p} \left(\frac{1}{\sqrt{T^-}} + \frac{1}{\sqrt{T^+}} \right)^{-1} \quad [2]$$

where M/A is the mass per unit area of the piston. In this state the piston has a temperature $T_P = \sqrt{T^+ T^-}$ and there is a heat flux

$$j_Q = (\sqrt{T^-} - \sqrt{T^+}) \frac{m}{M} \sqrt{\frac{8k_B}{m\pi}} p + o\left(\frac{m}{M}\right) \quad [3]$$

$(p^- = p^+ = p)$

For a finite cylinder and $p^+ \neq p^-$, the evolution proceeds in four different stages. The first two are deterministic and adiabatic. They correspond to the thermodynamic evolution of the (macroscopic) adiabatic piston. The last two stages, which go beyond thermodynamics, are stochastic with heat transfer across the piston. More precisely:

1. In the first stage whose duration is the time needed for the shock wave to bounce back on the piston, the evolution corresponds to the case of the infinite cylinder (with $p^- \neq p^+$). If $R = Nm/M > 10$, the piston will be able to reach and maintain a constant velocity

$$V = (p^- - p^+) \sqrt{\frac{\pi k_B}{8m}} \frac{\sqrt{T^- T^+}}{p^+ \sqrt{T^-} + p^- \sqrt{T^+}} + o\left(\frac{m}{M}\right) \quad [4]$$

for $|p^- - p^+| \ll 1$

2. In the second stage the evolution toward mechanical equilibrium is either weakly or strongly damped depending on R . If $R < 1$, the evolution is very weakly damped, the dynamics takes place on a timescale $t' = \sqrt{R}t$, and the effect of the collisions on the piston is to introduce an external potential $\phi(X) = c_1/X^2 + c_2/(L - X)^2$. On the other hand, if $R > 4$, the evolution is strongly damped (with two oscillations only) and depends neither on M nor on R .

3. After mechanical equilibrium has been reached, the third stage is a stochastic approach toward thermal equilibrium associated with heat transfer across the piston. This evolution is very slow and exhibits a scaling property with respect to $t' = mt/M$.
4. After thermal equilibrium has been reached ($T^- = T^+$, $p^- = p^+$), in a fourth stage the gas will evolve very slowly toward a state with Maxwellian distribution of velocities, induced by the collision with the stochastic piston.

The general conclusion is thus that a wall which is adiabatic when fixed will become a heat conductor under a stochastic motion. However, it should be stressed that the time required to reach thermal equilibrium will be several orders of magnitude larger than the age of the universe for a macroscopical piston and such a wall could not reasonably be called a heat conductor. However, for mesoscopic systems, the effect of stochasticity may lead to very interesting properties, as shown by [Van den Broeck *et al.* \(2004\)](#) in their investigations of Brownian (or biological) motors.

Microscopical Model

The system consists of two fluids separated by an “adiabatic” piston inside a cylinder with x -axis, length L , and area A . The fluids are made of N^\pm identical light particles of mass m . The piston is a heavy flat disk, without any internal degree of freedom, of mass $M \gg m$, orthogonal to the x -axis, and velocity parallel to this x -axis. If the piston is fixed at some position X_0 , and if the two fluids are in thermal equilibrium characterized by $(p_0^\pm, T_0^\pm, N^\pm)$, then they will remain in equilibrium forever even if $T_0^+ \neq T_0^-$: it is thus an “adiabatic piston” in the sense of thermodynamics. At a certain time $t=0$, the piston is let free to move and the problem is to study the time evolution. To define the dynamics, we consider that the system is purely Hamiltonian, that is, the particles and the piston move without any friction according to the laws of mechanics. In particular, the collisions between the particles and the walls of the cylinder, or the piston, are purely elastic and the total energy of the system is conserved. In most studies, one considers that the particles are point particles making purely elastic collisions. Since the piston is bound to move only in the x -direction, the velocity components of the particles in the transverse directions play no role in this problem. Moreover, since there is no coupling between the components in the x - and transverse directions, one can simplify the model further by assuming that all probability distributions are

independent of the transverse coordinates. We are thus led to a formally one-dimensional problem (except for normalizations). Therefore, in this review, we consider that the particles are noninteracting and all velocities are parallel to the x -axis. From the collision law, if v and V denote the velocities of a particle and the piston before a collision, then under the collision on the piston:

$$\begin{aligned} v \rightarrow v' &= 2V - v + \alpha(v - V) \\ V \rightarrow V' &= V + \alpha(v - V) \end{aligned} \quad [5]$$

where

$$\alpha = \frac{2m}{M + m} \quad [6]$$

Similarly, under a collision of a particle with the boundary at $x=0$ or $x=L$:

$$v \rightarrow v' = -v \quad [7]$$

Let us mention that more general models have also been considered, for example, the case where the two fluids are made of point particles with different masses m^\pm , or two-dimensional models where the particles are hard disks. However, no significant differences appear in these more general models and we restrict this article to the simplest case.

One can study different situations: $L = \infty$, L finite, and $L \rightarrow \infty$. Furthermore, taking first M and A finite, one can investigate several limits.

1. Thermodynamic limit for the piston only. In this limit, L is fixed (finite or infinite) and $A \rightarrow \infty, M \rightarrow \infty$, keeping constant the initial densities n^\pm of the fluid and the parameter

$$\gamma = \frac{2mA}{M + m} = \alpha A \sim 2m \frac{A}{M} \quad [8]$$

If L is finite, this means that $N^\pm \rightarrow \infty$ while keeping constant the parameters

$$R^\pm = \frac{mN^\pm}{M} = \frac{M_{\text{gas}}^\pm}{M} \quad [9]$$

2. Thermodynamic limit for the whole system, where $L \rightarrow \infty$ and $A \sim L^2, N^\pm \sim L^3$. In this limit, space and time variables are rescaled according to $x' = x/L$ and $t' = t/L$. This limit can be considered as a limiting case of (1) where $R^\pm \sim \sqrt{A} \rightarrow \infty$ (and time is scaled).
3. Continuum limit where L and M are fixed and $N^\pm \rightarrow \infty, m \rightarrow 0$ keeping M_{gas}^\pm constant, that is, $R^\pm = cte$.

The case L infinite and the limit (1) have been investigated using statistical mechanics (Liouville or

Boltzmann's equations). On the other hand, the limit (2) has been studied using dynamical system methods, reducing first the system to a billiard in an $(N^+ + N^- + 1)$ -dimensional polyhedron. The limit (3) has been introduced to derive hydrodynamical equations for the fluids.

In this article, we present the approach based on statistical mechanics. Although not as rigorous as (2) on a mathematical level, it yields more informations on the approach toward mechanical and thermal equilibrium. Moreover, it indicates what are the open problems which should be mathematically solved. In all investigations, advantage is taken of the fact that m/M is very small and one introduces the small parameter

$$\epsilon = \sqrt{m/M} \ll 1 \quad [10]$$

Let us note that ϵ measures the ratio of thermal velocities for the piston and a fluid particle, whereas $\alpha \sim \epsilon^2$ measures the ratio of velocity changes during a collision.

Starting Point: Exact Equations

Using the statistical point of view, the time evolution is given by Liouville's equation for the probability distribution on the whole phase space for $(N^+ + N^- + 1)$ particles, with L, A, N^\pm , and M finite. Initially ($t \leq 0$), the piston is fixed at $(X_0, V_0 = 0)$ and the fluids are in thermal equilibrium with homogeneous densities n_0^\pm , velocity distributions $\varphi_0^\pm(v) = \varphi_0^\pm(-v)$, and temperatures

$$T_0^\pm = m \int_{-\infty}^{\infty} dv n_0^\pm \varphi_0^\pm(v) v^2 \quad [11]$$

Integrating out the irrelevant degrees of freedom, the Liouville's equation yields the equations for the distribution $\rho^\pm(x, v; t)$ of the right and left particles:

$$\partial_t \rho^\pm(x, v; t) + v \partial_x \rho^\pm(x, v; t) = I^\pm(x, v; t) \quad [12]$$

The collision term $I^\pm(x, v; t)$ is a functional of $\rho_{\pm, P}(X, v; X, V; t)$, the two-point correlation function for a right (resp. left) particle at $(x = X, v)$ and the piston at (X, V) . Similarly, one obtains for the velocity distribution of the piston:

$$\begin{aligned} \partial_t \Phi(V; t) = & A \int_{-\infty}^{\infty} (V - v) [\theta(V - v) \rho_{\text{surf}}^-(v'; V'; t) \\ & + \theta(v - V) \rho_{\text{surf}}^-(v; V; t)] dv \\ & - A \int_{-\infty}^{\infty} (V - v) [\theta(v - V) \rho_{\text{surf}}^+(v'; V'; t) \\ & + \theta(V - v) \rho_{\text{surf}}^+(v; V; t)] dv \end{aligned} \quad [13]$$

where (v', V') are given by eqn [5] and

$$\rho_{\text{surf}}^\pm(v; V; t) = \int_{-\infty}^{\infty} dX \rho_{\pm, P}(X, v; X, V; t) \quad [14]$$

We thus have to solve eqns [12]–[13] with initial conditions

$$\begin{aligned} \rho^-(x, v; t = 0) &= n_0^- \varphi_0^-(v) \theta(x) \theta(X_0 - x) \\ \rho^+(x, v; t = 0) &= n_0^+ \varphi_0^+(v) \theta(L - x) \theta(x - X_0) \\ \Phi(V; t = 0) &= \delta(V) \end{aligned} \quad [15]$$

Using the fact that $\alpha = 2m/(M + m) \ll 1$, we can rewrite eqn [13] as a formal series in powers of α :

$$\partial_t \Phi(V; t) = \gamma \sum_{k=1}^{\infty} \frac{(-1)^k \alpha^{k-1}}{k!} \left(\frac{\partial}{\partial V} \right)^k \tilde{F}_{k+1}(V; t) \quad [16]$$

$$\begin{aligned} \tilde{F}_k(V; t) = & \int_V^{\infty} (v - V)^k \rho_{\text{surf}}^-(v; V; t) dv \\ & - \int_{-\infty}^V (v - V)^k \rho_{\text{surf}}^+(v; V; t) dv \end{aligned} \quad [17]$$

from which one obtains the equations for the moments of the piston velocity:

$$\begin{aligned} \frac{1}{\gamma} \frac{d \langle V^n \rangle}{dt} &= \sum_{k=1}^n \alpha^{k-1} \frac{n!}{k!(n-k)!} \int_{-\infty}^{\infty} dV V^{n-k} \tilde{F}_{k+1}(V; t) \end{aligned} \quad [18]$$

However, we do not know the two-point correlation functions.

If the length of the cylinder is infinite, the condition $M \gg m$ implies that the probability for a particle to make more than one collision on the piston is negligible. Alternatively, one could choose initial distributions $\varphi_0^\pm(v)$ which are zero for $|v| < v_{\text{min}}$, where v_{min} is taken such that the probability of a recollision is strictly zero. Therefore, if $L = \infty$, one can consider that before a collision on the piston the particles are distributed with $\varphi_0^\pm(v)$ for all t , and the two-point correlation functions factorize, that is,

$$\begin{aligned} \rho_{\text{surf}}^-(v; V; t) &= \rho_{\text{surf}}^-(v; t) \Phi(V; t), \quad \text{if } v > V \\ \rho_{\text{surf}}^+(v; V; t) &= \rho_{\text{surf}}^+(v; t) \Phi(V; t), \quad \text{if } v < V \end{aligned} \quad [19]$$

where for $L = \infty$, $\rho_{\text{surf}}^\pm(v; t) = n_0^\pm \varphi_0^\pm(v)$ and thus the conditions to obtain eqn [18] are satisfied.

If L is finite, one can show that the factorization property (eqn [19]) is an exact relation in the thermodynamic limit for the piston ($A \rightarrow \infty$, $M/A = cte$). For finite L and finite A , we introduce

Assumption 1 (Factorization condition). Before a collision the two-point correlation functions have the factorization property (eqn [19]) to first order in α .

Under the factorization condition, we have

$$\tilde{F}_k(V; t) = F_k(V; t)\Phi(V; t) \quad [20]$$

with

$$\begin{aligned} F_k(V; t) &= \int_V^\infty dv (v - V)^k \rho_{\text{surf}}^-(v; t) \\ &\quad - \int_{-\infty}^V dv (v - V)^k \rho_{\text{surf}}^+(v; t) \\ &= F_k^-(V; t) - F_k^+(V; t) \end{aligned} \quad [21]$$

and from eqn [18]

$$\left(\frac{M}{A}\right) \frac{d}{dt} \langle V \rangle = M\alpha \langle F_2(V; t) \rangle_\Phi \quad [22]$$

$$\left(\frac{M}{A}\right) \frac{d}{dt} \langle V^2 \rangle = M\alpha [\langle VF_2(V; t) \rangle_\Phi + \alpha \langle F_3(V; t) \rangle_\Phi] \quad [23]$$

Introducing $\bar{V} = \langle V \rangle_\Phi$ then from eqns [12] and [20], it follows that the (kinetic) energies satisfy

$$\begin{aligned} \frac{d}{dt} \left(\frac{\langle E^\pm \rangle}{A}\right) &= \pm M\alpha \left[\langle F_2^\pm(V; t) \rangle_\Phi \bar{V} \right. \\ &\quad \left. + \langle (V - \bar{V}) F_2^\pm(V; t) \rangle_\Phi \right. \\ &\quad \left. + \frac{\alpha}{2} \langle F_3^\pm(V; t) \rangle_\Phi \right] \end{aligned} \quad [24]$$

which implies conservation of energy.

From the first law of thermodynamics,

$$\frac{d}{dt} \left(\frac{\langle E^\pm \rangle}{A}\right) = \frac{1}{A} \left[P_W^{P \rightarrow \pm} + P_Q^{P \rightarrow \pm} \right] \quad [25]$$

where $P_W^{P \rightarrow \pm}$ and $P_Q^{P \rightarrow \pm}$ denote the work- and heat-power transmitted by the piston to the fluid, we conclude from eqns [22] and [25] that the heat flux is

$$\begin{aligned} \frac{1}{A} P_Q^{P \rightarrow \pm} &= \pm M\alpha \left[\langle (V - \bar{V}) F_2^\pm(V; t) \rangle_\Phi \right. \\ &\quad \left. + \frac{\alpha}{2} \langle F_3^\pm(V; t) \rangle_\Phi \right] \end{aligned} \quad [26]$$

Since $\alpha \ll 1$, it is interesting to introduce the irreducible moments

$$\Delta_r = \langle (V - \bar{V})^r \rangle_\Phi \quad [27]$$

and the expansion around $\bar{V} = \langle V \rangle_t$,

$$F_n^\pm(V; t) = \sum_{r=0}^{\infty} \frac{1}{r!} F_n^{(r, \pm)}(\bar{V})(V - \bar{V})^r \quad [28]$$

from which one obtains equations for $d\Delta_r/dt$. In particular, using the identities

$$F_3^{(r+1, \pm)} = -3F_2^{(r, \pm)}, \quad F_2^{(r+2, \pm)} = 2F_0^{(r, \pm)} \quad [29]$$

in [22] and [24], we have

$$\begin{aligned} \langle F_2^\pm(V; t) \rangle_\Phi &= F_2^\pm(\bar{V}; t) \\ &\quad + \sum_{r \geq 0} \frac{2}{(2+r)!} F_0^{(r, \pm)} \Delta_{2+r} \end{aligned} \quad [30]$$

$$\begin{aligned} \frac{d}{dt} \left(\frac{\langle E^\pm \rangle}{A}\right) &= \pm M\alpha \left[\langle F_2^\pm(V; t) \rangle_\Phi \bar{V} \right. \\ &\quad \left. + \frac{\alpha}{2} F_3^\pm(\bar{V}; t) + \frac{1}{2} \sum_{r \geq 2} \frac{1}{r!} (2r - 3\alpha) \right. \\ &\quad \left. \times F_2^{(r-1, \pm)}(\bar{V}; t) \Delta_r \right] \end{aligned} \quad [31]$$

Depending on the questions or approximations one wants to study, either the distribution $\Phi(V; t)$ or the moments $\langle V^n \rangle_t$ will be the interesting objects. Finally, with the condition [19], one can take eqn [12] for $x \neq X_t$ and impose the boundary conditions at $x = X_t$:

$$\begin{aligned} \rho^-(X_t, v; t) &= \rho^-(X_t, v'; t), \quad \text{if } v < V_t \\ \rho^+(X_t, v; t) &= \rho^+(X_t, v'; t), \quad \text{if } v > V_t \end{aligned} \quad [32]$$

and similarly for $x=0$ and $x=L$ with $v' = -v$.

Let us note that this factorization condition is of the same nature as the molecular chaos assumption introduced in kinetic theory, and with this condition eqn [13] yields the Boltzmann equation for this model.

In the following, to obtain explicit results as a function of the initial temperatures T_0^\pm , we take Maxwellian distributions $\varphi_0^\pm(v)$ and initial conditions $(p_0^\pm, T_0^\pm, n_0^\pm)$ such that the velocity of the piston remains small (i.e., $|\langle V \rangle_t| \ll |\langle v^\pm \rangle_0|$).

Distribution $\Phi(V; t)$ for the Infinite Cylinder ($L = \infty$)

To lowest order in $\epsilon = \sqrt{m/M}$, and assuming $|1 - p^+/p^-|$ is of order ϵ , one obtains from eqn [16] the usual Fokker-Planck equation whose solution gives

$$\Phi_0(V; t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Delta(t)} \exp - \left(\frac{(V - \bar{V}(t))^2}{2\Delta^2(t)} \right) \quad [33]$$

with

$$\begin{aligned}\bar{V}(t) &= (p^- - p^+) \sqrt{\frac{\pi k_B}{8m}} \left[\frac{p^+}{\sqrt{T^+}} + \frac{p^-}{\sqrt{T^-}} \right]^{-1} (1 - e^{-\lambda t}) \\ \lambda &= \frac{A}{M} \sqrt{\frac{8m}{\pi k_B}} \left[\frac{p^+}{\sqrt{T^+}} + \frac{p^-}{\sqrt{T^-}} \right] \\ \Delta^2(t) &= \frac{k_B}{M} \sqrt{T^- T^+} \frac{p^+ \sqrt{T^+} + p^- \sqrt{T^-}}{p^+ \sqrt{T^-} + p^- \sqrt{T^+}} (1 - e^{-2\lambda t})\end{aligned}\quad [34]$$

where we have dropped the index “zero” on the variable T^\pm, n^\pm and used the equation of state $p^\pm = n^\pm k_B T^\pm$.

In conclusion, in the thermodynamic limit for the piston ($M \rightarrow \infty, M/A$ fixed), eqn [33] shows that the evolution is deterministic, that is, $\Phi(V; t) = \delta(V - \bar{V}(t))$, where the velocity $\bar{V}(t)$ of the piston tends exponentially fast toward stationary value $V_{\text{stat}} = \bar{V}(\infty)$ with relaxation time $\tau = \lambda^{-1}$.

Let us note that for $p^+ = p^-$, we have $\bar{V}(t) \equiv 0$ and the evolution [33] is identical to the Ornstein–Uhlenbeck process of thermalization of the Brownian particle starting with zero velocity and friction coefficient λ . The analysis of [16] to first order in ϵ yields then

$$\Phi(V; t) = \left[1 + \epsilon \sum_{k=0}^3 a_k(t) (V - \bar{V}(t))^k \right] \Phi_0(V; t) \quad [35]$$

where $a_k(t)$ can be explicitly calculated and $a_0(t) = -\Delta^2(t) a_2(t)$ because of the normalization condition. Moreover, $a_2(t) \sim (p^- - p^+)$, that is, $a_2(t) = 0$ if $p^- = p^+$. From [35], one obtains

$$\begin{aligned}\langle V \rangle_t &= \sqrt{\frac{\pi k_B}{8m}} \frac{\sqrt{T^- T^+}}{p^+ \sqrt{T^-} + p^- \sqrt{T^+}} \\ &\times \left\{ (p^- - p^+) (1 - e^{-\lambda t}) \right. \\ &+ (p^- - p^+) \frac{2\pi}{8} \frac{(p^- T^+ - p^+ T^-)}{(p^+ \sqrt{T^-} + p^- \sqrt{T^+})^2} \\ &\times (1 - 2\lambda t e^{-\lambda t} - e^{-2\lambda t}) \\ &+ \frac{m}{M} \frac{1}{\sqrt{T^- T^+}} (p^- T^+ - p^+ T^-) \\ &\left. \times \left(\frac{p^+ \sqrt{T^+} + p^- \sqrt{T^-}}{p^+ \sqrt{T^-} + p^- \sqrt{T^+}} \right) (1 - e^{-\lambda t})^2 \right\} \quad [36]\end{aligned}$$

and

$$\langle V^2 \rangle_t - \langle V \rangle_t^2 = \Delta^2(t) \left[1 + \sqrt{\frac{m}{M}} 2\Delta^2(t) a_2(t) \right] \quad [37]$$

From eqn [36], we now conclude that for equal pressures $p^- = p^+$, the piston will evolve stochastically to a stationary state with nonzero velocity toward the warmer side

$$\left. \begin{aligned}\langle V \rangle_{\text{stat}} &= \frac{m}{M} \sqrt{\frac{\pi k_B}{8m}} (\sqrt{T^+} - \sqrt{T^-}) \\ \langle V^2 \rangle_{\text{stat}} - \langle V \rangle_{\text{stat}}^2 &= \frac{k_B}{M} \sqrt{T^- T^+}\end{aligned}\right\} \text{if } p^- = p^+ \quad [38]$$

Let us remark that we have established eqn [35] under the condition that $|1 - p^+/p^-| = \mathcal{O}(\epsilon)$, but as we see in the next section, the stationary value V_{stat} obtained from eqn [36] remains valid whenever $|(1 - p^+/p^-)(1 - \sqrt{T^+/T^-})| \ll 1$.

Moments $\langle V^n \rangle_t$: Thermodynamic Limit for the Piston

General Equations: Adiabatic Evolution

In the thermodynamic limit $M \rightarrow \infty, \alpha \rightarrow 0, \gamma = \alpha A$ is fixed and eqn [16] reduces to

$$\partial_t \Phi(V; t) = -\gamma \frac{\partial}{\partial V} \tilde{F}_2(V; t) \quad [39]$$

Integrating [39] with initial condition $\Phi(V; t=0) = \delta(V)$ yields

$$\Phi(V, t) = \delta(V - \bar{V}(t)), \text{ that is, } \langle V^n \rangle_t = \langle V \rangle_t^n \quad [40]$$

where

$$\frac{d}{dt} V(t) = \gamma F_2(V(t); t), \quad V(t=0) = 0 \quad [41]$$

Moreover,

$$\tilde{F}_2(V; t) = F_2(V; t) \Phi(V; t) \quad [42]$$

and

$$\begin{aligned}\rho_{\pm, P}(X, v; X, V; t) &= \rho^\pm(x, v; t) \delta(X - X(t)) \\ &\times \delta(V - V(t))\end{aligned} \quad [43]$$

where $dX(t)/dt = V(t), X(t=0) = X_0$.

In conclusion, as already mentioned, in this limit the factorization condition (eqn [19]) is an exact relation. Let us note that $\rho_{\text{surf}}^\pm(v; t) = \rho_{\text{surf}}^\pm(2V - v; t)$ if $v > V(t)$ (on the right) or $v < V(t)$ (on the left). Let us also remark that $2mF_2^\pm(V(t); t)$ represents the effective pressure from the right/left exerted on the piston. Moreover, since for any distribution $\rho_{\text{surf}}^\pm(v; t)$, the functions $F_2^-(V; t)$ and $-F_2^+(V; t)$ are monotonically decreasing, we can introduce the decomposition

$$p_{\text{surf}}^\pm = 2mF_2^\pm(V; t) = \hat{p}^\pm \pm \left(\frac{M}{A} \right) \lambda^\pm(V; t) V \quad [44]$$

where the static pressure at the surface is $\hat{p}^\pm(t) = p_{\text{surf}}^\pm(V=0; t)$ and the friction coefficients

$\lambda^\pm(V; t)$ are strictly positive. The evolution [41] is thus of the form

$$\frac{d}{dt} V(t) = \frac{A}{M} (\hat{p}^- - \hat{p}^+) - \lambda(V) V \quad [45]$$

It involves the difference of static pressure and the friction coefficient $\lambda(V) = \lambda^-(V) + \lambda^+(V)$. Finally, from eqn [12], we obtain the evolution of the (kinetic) energy per unit area for the fluids in the left and right compartments:

$$\frac{d}{dt} \left(\frac{\langle E^\pm \rangle}{A} \right) = \pm 2m F_2^\pm(V; t) V \quad [46]$$

Therefore, from [40] and [46], and the first law of thermodynamics, we recover the conclusions obtained in the previous section, that is, in the thermodynamic limit for the piston, the evolution (eqns [41], [12], and [35]) is deterministic and adiabatic (i.e., in [46] only work and no heat is involved).

Infinite Cylinder ($L = \infty, M = \infty$)

As already discussed, for $L = \infty$ we can neglect the recollisions. Therefore, in F_2^\pm the distribution $\rho^\pm(v; t)$ can be replaced by $n_0^\pm \varphi_0^\pm(v)$ and $F_2^\pm(V)$ is independent of t . In this case, the evolution of the piston is simply given by the ordinary differential equation

$$\frac{d}{dt} V(t) = \frac{A}{M} 2m F_2(V), \quad V(t=0) = 0 \quad [47]$$

where $F_2(V)$ is a strictly decreasing function of V . If $p_0^+ = p_0^-$, then $V(t) = 0$, that is, the piston remains at rest and the two fluids remain in their original thermal equilibrium. If $p_0^+ \neq p_0^-$, that is, $n_0^+ k_B T_0^+ \neq n_0^- k_B T_0^-$, the piston will evolve monotonically to a stationary state with constant velocity V_{stat} solution of $F_2(V_{\text{stat}}) = 0$. From [34], it follows that V_{stat} is a function of $n_0^+/n_0^-, T_0^-, T_0^+$ but does not depend on the value M/A . Moreover, the approach to this stationary state is exponentially fast with relaxation time $\tau_0 = 1/\lambda(V=0)$. For Maxwellian distributions $\varphi_0^\pm(v)$, V_{stat} is a solution of

$$k_B (n_0^- T_0^- - n_0^+ T_0^+) - V_{\text{stat}} \sqrt{\frac{8k_B m}{\pi}} \left(n_0^- \sqrt{T_0^-} - n_0^+ \sqrt{T_0^+} \right) + V_{\text{stat}}^2 m (n_0^- - n_0^+) + \mathcal{O}(V_{\text{stat}}^3) = 0 \quad [48]$$

Moreover,

$$\tau_0^{-1} = \frac{A}{M} \sqrt{\frac{8k_B m}{\pi}} \left(n_0^- \sqrt{T_0^-} + n_0^+ \sqrt{T_0^+} \right) \quad [49]$$

which implies that the relaxation time will be very small either if $M/A \ll 1$, or if $n_0^\pm = \xi \tilde{n}_0^\pm$ with $\xi \gg 1$. In this case, the piston acquires almost immediately

its final velocity V_{stat} and one can solve eqn [12] to obtain the evolution of the fluids.

Finite Cylinder ($L < \infty, M = \infty$)

For finite L , introducing the average temperature in the fluids

$$T_{\text{av}}^\pm = \frac{2\langle E^\pm \rangle_t}{k_B N^\pm} \quad [50]$$

we have to solve [41] and [46], that is,

$$\begin{aligned} \frac{d}{dt} V(t) &= \frac{A}{M} 2m [F_2^-(V; t) - F_2^+(V; t)] \\ k_B \frac{d}{dt} T_{\text{av}}^\pm &= \pm 4m \frac{A}{N^\pm} F_2^\pm(V; t) V \end{aligned} \quad [51]$$

where $F_2^\pm(V; t)$ is a functional of $\rho_{\text{surf}}^\pm(v; t)$ which we decompose as

$$F_2^\pm(V; t) = \hat{n}^\pm(t) k_B \hat{T}^\pm(t) \pm \left(\frac{M}{A} \right) \lambda^\pm(V; t) V \quad [52]$$

with

$$\begin{aligned} \hat{n}^-(t) &= \int_0^\infty dv \rho_{\text{surf}}^-(v; t) \\ \hat{n}^+(t) &= \int_{-\infty}^0 dv \rho_{\text{surf}}^+(v; t) \end{aligned} \quad [53]$$

and

$$\hat{n}^\pm k_B \hat{T}^\pm = \hat{p}^\pm \quad [54]$$

For a time interval $\tau_1 = L\sqrt{m/k_B T}$ which is the time for the shock wave to bounce back, the piston will evolve as already discussed. In particular, if R^\pm is sufficiently large, then after a time $\tau_0 = \mathcal{O}((R^\pm)^{-1})$ the piston will reach the velocity \bar{V} given by $F_2(\bar{V}, t) = 0$ (eqn [47]). For $t > \tau_1$, $F_2^\pm(V; t)$ depends explicitly on time. For R^\pm sufficiently large, we can expect that for all t the velocity $V(t)$ will be a functional of $\rho_{\text{surf}}^\pm(v; t)$ given by $F_2[V(t); \rho_{\text{surf}}^\pm(\cdot; t)] = 0$, and thus the problem is to solve eqn [12] with the boundary condition (eqn [32]). Since $V(t)$ so defined is independent of M/A , the evolution will be independent of M/A if R^\pm is sufficiently large. This conclusion, which we cannot prove rigorously, will be confirmed by numerical simulations.

To give a qualitative discussion of the evolution for arbitrary values of R^\pm , we shall use the following assumption already introduced in the experimental measurement of c_p/c_v .

Assumption 2 (Average assumption). The surface coefficients $\hat{n}^\pm(t)$ and $\hat{T}^\pm(t)$ (eqns [52]–[53]) coincide to order 1 in α with the average value of the density and temperature in the fluids, that is,

$$\hat{n}^- = \frac{N^-}{AX(t)}, \quad \hat{n}^+ = \frac{N^+}{A(L-X(t))}$$

$$\hat{T}^\pm = T_{\text{av}}^\pm(t) \quad [55]$$

We still need an expression for the friction coefficients. From

$$F_2^\pm(V; t) = \hat{p}^\pm(t) - 4mVF_1^\pm(V=0; t) + mV^2\hat{n}^\pm(t) + \mathcal{O}(V^3) \quad [56]$$

then, assuming that to first order in α , $F_1^\pm(V=0; t)$ is the same function of $\hat{T}^\pm(t)$ as for Maxwellian distributions, we have

$$\lambda^\pm(V) = \left(\frac{A}{M}\right) m\hat{n}^\pm \left[\sqrt{\frac{8k_B\hat{T}^\pm}{m\pi}} \pm V \right] + \mathcal{O}(V^2) \quad [57]$$

Therefore, choosing initial condition such that $V(t)$ is small for all time, eqn [51] yields

$$\sqrt{\hat{T}^-X} - \sqrt{\hat{T}^+(L-X)} = C = \sqrt{\hat{T}_0^-X_0} - \sqrt{\hat{T}_0^+(L-X_0)} \quad [58]$$

We thus obtain the equilibrium point for the adiabatic evolution ($M=\infty$):

$$\left(\frac{N^-}{A}\right) T_f^- = \frac{2E_0 X_f}{Ak_B L} \quad [59]$$

$$\left(\frac{N^+}{A}\right) T_f^+ = \frac{2E_0}{Ak_B} \left(1 - \frac{X_f}{L}\right) \quad [60]$$

where

$$\frac{2E_0}{Ak_B} = \left(\frac{N^-}{A}\right) T_0^- + \left(\frac{N^+}{A}\right) T_0^+ \quad [61]$$

and

$$\sqrt{\left(\frac{A}{N^-}\right) X_f^3} - \sqrt{\left(\frac{A}{N^+}\right) (L-X_f)^3} = \sqrt{\frac{AL}{2E_0 k_B}} C \quad [62]$$

Solving [58]–[62] gives the equilibrium state (X_f, T_f^\pm) , which is a state of mechanical equilibrium $p_f^- = p_f^+$, but not thermal equilibrium $T_f^- \neq T_f^+$. Moreover, this equilibrium state does not depend on M . Having obtained the equilibrium point, we can then investigate the evolution close to the equilibrium point. Linearizing eqn [51] around (X_f, T_f^\pm) yields

$$\frac{d}{dt} V = k_B \left[\left(\frac{N^-}{M}\right) \frac{T_f^- X_f^2}{X^3} - \left(\frac{N^+}{M}\right) \frac{T_f^+ (L-X_f)^2}{(L-X)^3} \right] - \lambda(V=0)V \quad [63]$$

In other words, the effect of collisions on the piston is to induce an external potential of the form $[c_1|X|^{-2} + c_2(L-X)^{-2}]$ and a friction force. It is a damped harmonic oscillator with

$$\omega_0^2 = 6 \left(\frac{E_0}{M}\right) \frac{1}{X_f(L-X_f)}$$

$$\lambda = 4 \sqrt{\frac{1}{\pi}} \sqrt{\frac{E_0}{ML}} \left[\sqrt{\frac{R^-}{X_f}} + \sqrt{\frac{R^+}{(L-X_f)}} \right] \quad [64]$$

(recall that $R^\pm = mN^\pm/M$). For the case $N^- = N^+$ to be considered in the simulations, eqn [64] implies that the motion is weakly damped if

$$R < R_{\text{max}} = \frac{3\pi}{2} \left[\sqrt{\frac{X_f}{L}} + \sqrt{1 - \frac{X_f}{L}} \right]^{-2} \quad [65]$$

with period

$$\tau = \frac{2\pi}{\omega_0} \frac{1}{\sqrt{R - R_{\text{max}}}} \quad [66]$$

and strongly damped if $R > R_{\text{max}}$, in agreement with experimental observations.

Moments $\langle V^n \rangle_t$: Piston with Finite Mass

Equation to First Order in $\alpha = 2m/(M+m)$

If the mass of the piston is finite with $M \gg m$, then the irreducible moments Δ_r are of the order $\alpha^{[(r+1)/2]}$ where $[(r+1)/2]$ is the integral part of $(r+1)/2$. If the factorization condition [19] is satisfied, to first order in α we have

$$\langle V^n \rangle_t = V^n(t) + \frac{n(n-1)}{2} V^{n-2}(t) \Delta_2(t) \quad [67]$$

where $V(t) = \langle V \rangle_t$ and $\Delta_2(t) = \langle V^2 \rangle_t - \langle V \rangle_t^2$ are solutions of

$$\frac{1}{\gamma} \frac{d}{dt} V(t) = F_2 + \Delta_2 F_0$$

$$\frac{1}{\gamma} \frac{d}{dt} \Delta_2(t) = -4\Delta_2 F_1 + \alpha F_3 \quad [68]$$

$$\frac{1}{\gamma} \frac{d}{dt} \langle E^\pm \rangle_t = \pm \{ M[F_2^\pm + \Delta_2 F_0^\pm] V + (M/2)[4\Delta_2 F_1^\pm - \alpha F_3^\pm] \}$$

and $\Delta_2 \doteq k_B T_P/M$ defines the temperature of the piston.

Infinite Cylinder: Heat Transfer

For the infinite cylinder, the factorization assumption is an exact relation and in this case the functions $F_k(V; t)$ are independent of t . The solution

of the autonomous system [68] with $F_k = F_k(V)$ shows that the piston evolves to a stationary state with velocity \bar{V} given by

$$F_2(\bar{V}) + \frac{\alpha F_3(\bar{V})F_0(\bar{V})}{4 F_1(\bar{V})} = 0 \quad [69]$$

The temperature of the piston is

$$\bar{\Delta}_2 = \frac{k_B T_P}{M} = \frac{\alpha F_3(\bar{V})}{4 F_1(\bar{V})} \quad [70]$$

and the heat flux from the piston to the fluid is

$$\frac{1}{A} P_Q^{P \rightarrow -} = \frac{m^2}{2M} \left[\frac{F_3^+ F_1^- - F_3^- F_1^+}{F_1^- - F_1^+} \right] \quad [71]$$

If we choose initial conditions such that $|V(t)| \ll 1$ for all t , and Maxwellian distributions $\varphi^\pm(v)$, the solutions $V(t)$, $\Delta_2(t)$ coincide with the solutions previously obtained (eqns [36] and [37]) and

$$\begin{aligned} \frac{1}{A} P_Q^{P \rightarrow -} &= (T^+ - T^-) \times \frac{m}{M} \sqrt{\frac{8k_B}{m\pi}} \\ &\times \frac{p^- p^+}{(p^+ \sqrt{T^-} + p^- \sqrt{T^+})} \end{aligned} \quad [72]$$

In conclusion, to first order in m/M , there is a heat flux from the warm side to the cold one proportional to $(T^+ - T^-)$, induced by the stochastic motion of the piston.

Finite Cylinder ($L < \infty, M < \infty$)

Singular character of the perturbation approach Whereas the leading order is actually the “thermodynamic behavior” $M = \infty$ in the first two stages of the evolution (fast relaxation toward mechanical equilibrium), the fluctuations of order $\mathcal{O}(\alpha)$ rule the slow relaxation toward thermal equilibrium. It is thus obvious that a naive perturbation approach cannot give access to “both” regimes. This difficulty is reminiscent of the boundary-layer problems encountered in hydrodynamics, and the perturbation method to be used here is the exact temporal analog of the matched perturbative expansion method developed for these boundary layers. The idea is to implement two different perturbation approaches:

1. one at short times, with time variable t describing the fast dynamics ruling the fast relaxation toward mechanical equilibrium; and
2. one for longer times, with a rescaled time variable $\tau = \alpha t$.

The second perturbation approach above is supplemented with a “slaving principle,” expressing that at each time of the slow evolution, that is, at fixed τ , the still present fast dynamics has reached a local asymptotic state, slaved to the values of the slow

observables. The initial conditions are set on the first-stage solution. The initial conditions of the second regime match the asymptotic behavior of the first-stage solution (“matching condition”).

The slaving principle is implemented by interpreting an evolution equation of the form

$$\frac{da}{dt} \equiv \alpha \frac{da}{d\tau} = A(\tau, a), \quad A = \mathcal{O}(1) \quad [73]$$

as follows: it indicates that a is in fact a fast quantity relaxing at short times ($\ll \tau$) toward a stationary state $a_{\text{eq}}(\tau)$ slaved to the slow evolution and determined by the condition

$$A[\tau, a_{\text{eq}}(\tau)] = 0 \quad [74]$$

(at lowest order in α , actually $A[\tau, a_{\text{eq}}(\tau)] = \mathcal{O}(\alpha)$ which prescribes the leading order of $a_{\text{eq}}(\tau)$); the following-order terms can be arbitrarily fixed as long as only the first order of perturbation is implemented. Physically, such a condition arises to express that an instantaneous mechanical equilibrium takes place at each time τ of the slow relaxation to thermal equilibrium.

Equations for the fluctuation-induced evolution of the system Following this procedure, we arrive at explicit expressions for the rescaled quantities (of order $\mathcal{O}(1)$) $\tilde{V} = V/\alpha$, $\tilde{\Delta}_2 = \Delta_2/\alpha$, and $\tilde{\Pi} = (p^- - p^+)/\alpha$:

$$\begin{aligned} \tilde{V} &= \frac{m}{3} \left(\frac{AL}{E_0} \right) \left(\frac{F_3^- F_1^+ - F_3^+ F_1^-}{F_1} \right) + \mathcal{O}(\alpha) \\ \frac{\tilde{\Pi}}{2m} &= \frac{2m}{3} \left(\frac{AL}{E_0} \right) (F_3^- F_1^+ - F_3^+ F_1^-) \\ &\quad - \frac{F_3 F_1}{4F_1} + \mathcal{O}(\alpha) \\ \tilde{\Delta}_2 &= \frac{F_3}{4F_1} + \mathcal{O}(\alpha) \end{aligned} \quad [75]$$

We then introduce a (dimensionless) rescaled position for the piston

$$\xi = \frac{1}{2} - \frac{X}{L} \in \left[-\frac{1}{2}, \frac{1}{2} \right] \quad [76]$$

which satisfies

$$\frac{d\xi}{d\tau} = -k_B(T^- - T^+) \left(\frac{2A}{3E_0} \right) \frac{F_1^- F_1^+}{F_1} \quad [77]$$

To discuss eqn [77], a third assumption has to be introduced.

Assumption 3 (Maxwellian Identities). In the regime when $V = \mathcal{O}(\alpha)$, the relations between the functionals F_1, F_2 , and F_3 are the same at lowest order in α as if the distributions $\rho_{\text{surf}}^\pm(v; V; t)$ were Maxwellian in v :

$$\begin{aligned}
 F_1^\pm(V) &\approx \mp \rho^\pm \sqrt{\frac{k_B T^\pm}{2m\pi}} \\
 F_3^\pm(V) &\approx \left(\frac{2k_B T^\pm}{m}\right) F_1^\pm(V) - VF_2^\pm(V)
 \end{aligned}
 \tag{78}$$

Using these identities and the (dimensionless) rescaled time

$$s = \tau \frac{2}{3L} \sqrt{\frac{k_B}{m\pi}} \sqrt{\frac{2(N^-T_0^- + N^+T_0^+)}{N}}
 \tag{79}$$

where $N = N^+ + N^-$, we obtain a deterministic equation describing the piston motion (Gruber *et al.* 2003):

$$\begin{aligned}
 \frac{d\xi}{ds} &= - \left[\sqrt{\frac{N}{2N^+}} (1 + 2\xi) - \sqrt{\frac{N}{2N^-}} (1 - 2\xi) \right] \\
 \xi(0) &= \frac{1}{2} - \frac{X_{\text{ad}}}{L}
 \end{aligned}
 \tag{80}$$

where X_{ad} is the piston position at the end of the adiabatic regime (i.e., X_f , eqn [62]). The meaningful observables straightforwardly follow from the solution $\xi(s)$:

$$\begin{aligned}
 X(s) &= L \left(\frac{1}{2} - \xi(s) \right) \\
 T^\pm(s) &= [1 \pm 2\xi(s)] \left(\frac{N^-T_0^- + N^+T_0^+}{2N^\pm} \right)
 \end{aligned}
 \tag{81}$$

The first-order perturbation analysis using a single rescaled time $t_1 = \alpha t_0$ is valid in the regime when $V = \mathcal{O}(\alpha)$ and it gives access to the relaxation toward

thermal equilibrium up to a temperature difference $T^+ - T^- = \mathcal{O}(\alpha)$. For the sake of technical completeness (rather than physical relevance, since the above first-order analysis is enough to get the observable, meaningful behavior), let us mention that the perturbation analysis can be carried over at higher orders; using further rescaled times $t_2 = \alpha^2 t_0, \dots, t_n = \alpha^n t_0$, it would allow us to control the evolution up to a temperature difference $|T^+ - T^-| = \mathcal{O}(\alpha^n)$; however, one could expect that the factorization condition does not hold at higher orders.

Numerical Simulations

As we have seen, the results were established under the condition that m/M is a small parameter. Moreover for finite systems ($L < \infty, M < \infty$), it was assumed that before collisions and to first order in m/M , the factorization and the average assumptions are satisfied. The numerical simulations are thus essential to check the validity of these assumptions, to determine the range of acceptable values m/M for the perturbation expansion, to investigate the thermodynamic limit, and to guide the intuition.

In all simulation, we have taken $k_B = 1, m = 1, T^- = 1$ and usually $T^+ = 10$. For L finite, we have taken $L = 60, X_0 = 10, A = 10^5$, and $N^+ = N^- = N/2$, that is, $p^- = R(M/A)(1/10)$ and $p^+ = 2p^-$. The number of particles N was varied from a few hundreds to one or several millions; the mass M of the piston from 1 to 10^5 . We give below some of the results which have been obtained for $L = \infty$ (Figures 2 and 3)

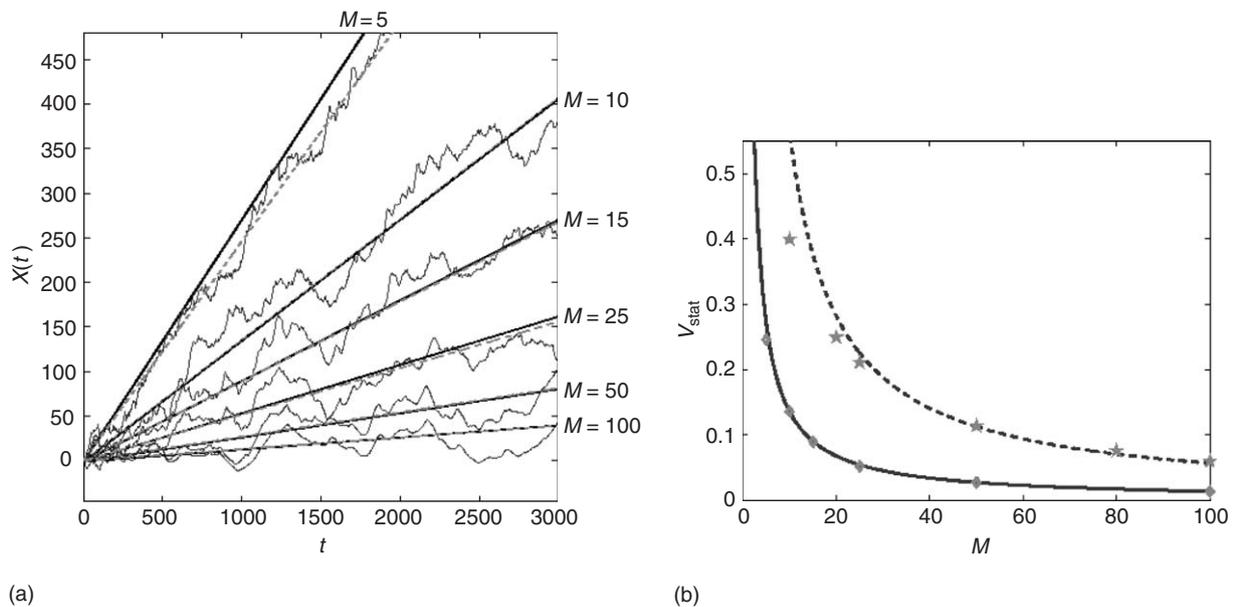


Figure 2 Evolution of the piston for $L = \infty$, and $p^- = p^+ = 1$ as observed in simulations (stochastic line in (a), dots in (b)) compared with prediction: (a) position $X(t)$ for $T^+ = 10$; and (b) stationary velocity for $T^+ = 10$ (continuous line) and $T^+ = 100$ (dotted line), as a function of M .

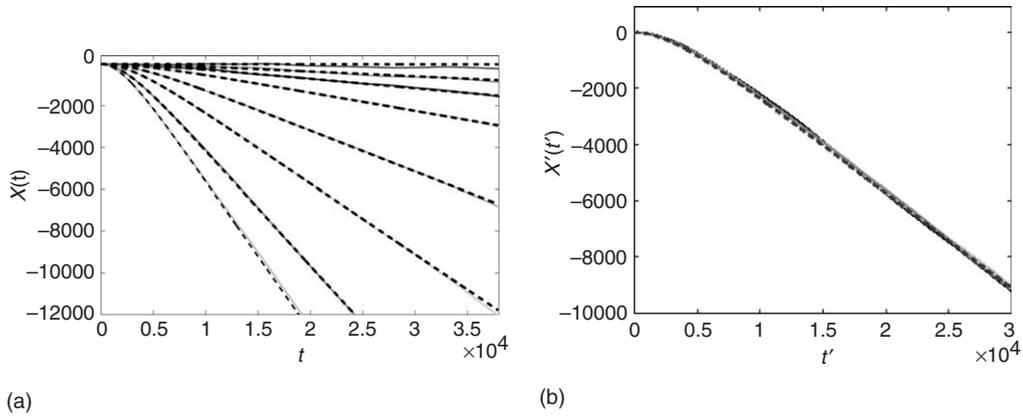


Figure 3 Evolution of the piston for $L = \infty, M = 10^4$, and $p^+ \neq p^-$ as observed in simulations (continuous line) compared with predictions (dotted line): (a) $p^- = 1, p^+ = p^- + \Delta p$, from top to bottom $\Delta p/p^- = 0.05, 0.1, 0.2, 1, 2, 3$; and (b) $p^- = \zeta, p^+ = 2\zeta, \Delta p/p^- = 1; X' = \zeta X, t' = \zeta t, \zeta = 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4$.

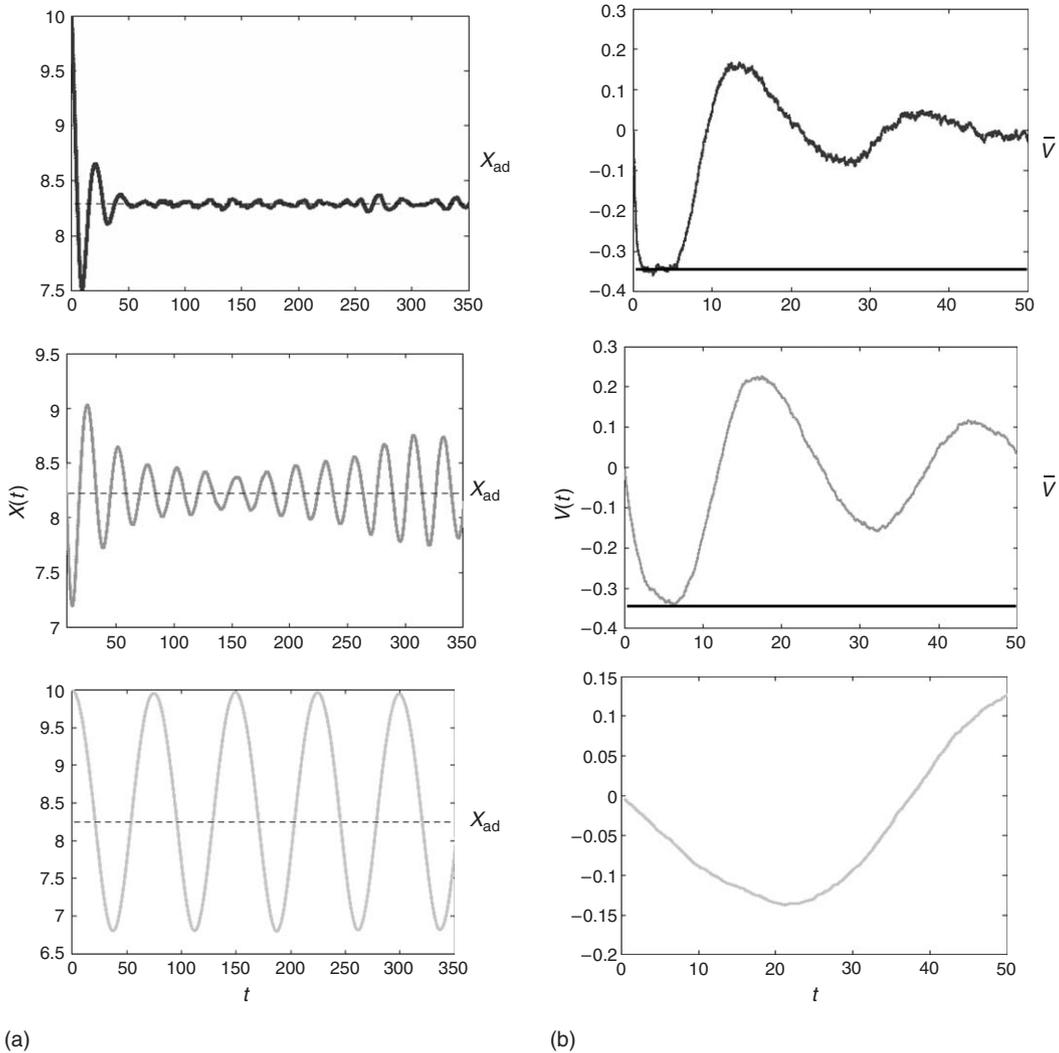


Figure 4 “Deterministic” evolution toward mechanical equilibrium for $L < \infty, M = 10^5$: (a) position $X(t)$; one finds $X_{ad}^{sim} = 8.3$ whereas $X_{ad}^{th} = 8.42$ and (b) velocity $V(t)$; one finds $\bar{V}^{sim} = -0.343$ whereas $\bar{V}^{th} = -0.3433$. From top to bottom: $R = 12$: strong damping, independent of R and M for $R > 4$ and $M > 10^3$. $R = 2$: critical damping. $R = 0.1$: weak damping; damping coefficient increases with R and $\omega_0 \sim \sqrt{R}$ for $R < 1$ but is independent of M for $M > 10^3$.

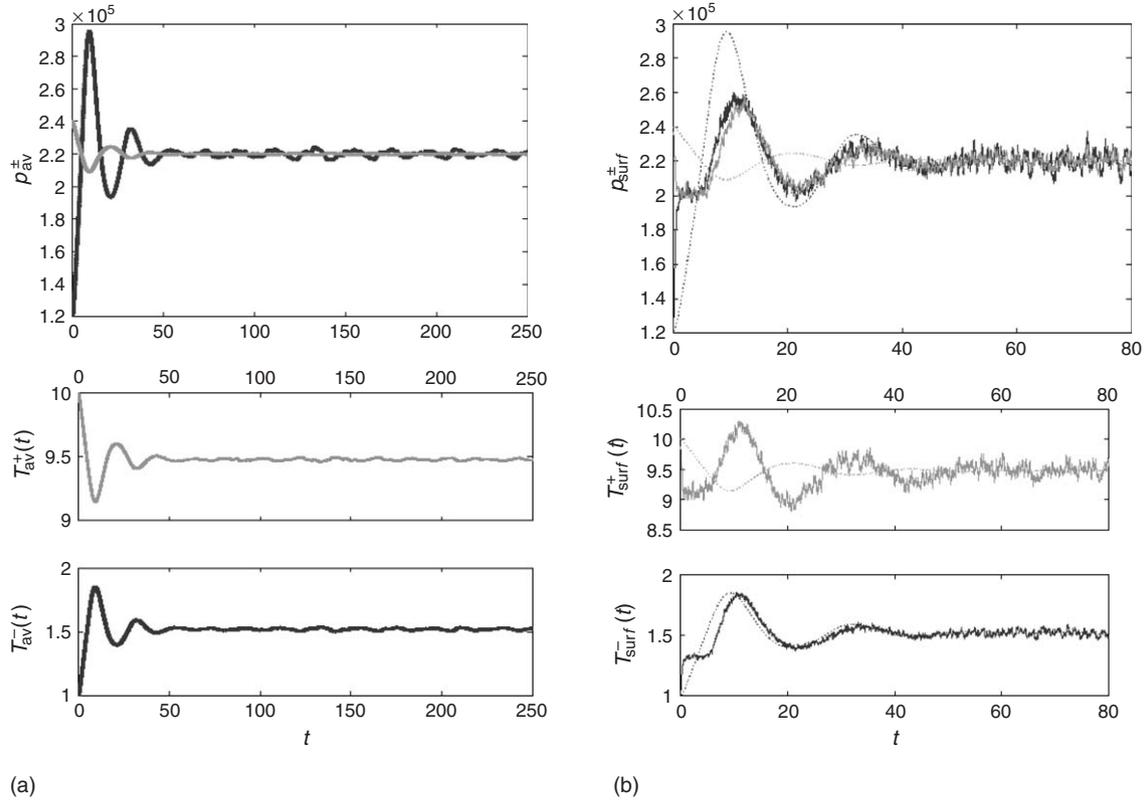


Figure 5 Same conditions as **Figure 4**, $R=12$: (a) average pressure and temperature in the fluid: $p_{av}^{\pm}(t) = 2E^{\pm}n^{\pm}/N^{\pm}$, $T_{av}^{\pm} = E^{\pm}/N^{\pm}k_B$ and (b) pressure and temperature at the surface of the piston. Prediction: $T_{ad}^{-} = 1.54$, $T_{ad}^{+} = 9.46$, $p_{ad}^{-} = p_{ad}^{+} = 2.2$. Simulations: $T_{ad}^{-} = 1.52$, $T_{ad}^{+} = 9.48$, $p_{ad}^{-} = p_{ad}^{+} = 2.2$.

and for $L < \infty$ approach to mechanical equilibrium (**Figures 4–6**) and to thermal equilibrium (**Figures 7 and 8**).

Conclusions and Open Problems

In this article, the adiabatic piston has been investigated to first order in the small parameter m/M , but no attempt has been made to control the remainder terms. For an infinite cylinder, no other assumptions were necessary and the numerical simulations (**Figures 2 and 3**) are in perfect agreement with the theoretical prediction in particular for the stationary velocity V_{stat} , the friction coefficient $\lambda(V)$, and the relaxation time τ .

For a finite cylinder ($L < \infty$) and in the thermodynamic limit ($M = \infty$), we were forced to introduce the average assumption to obtain a set of autonomous equations. As we have seen when initially $p^- \neq p^+$, this limiting case also describes the evolution to lowest order during the first two stages characterized by a time of the order $t_1 = L\sqrt{m/k_B T}$, where the evolution is adiabatic and deterministic. In the first stage, that is, before the shock wave bounces back on the piston, the simulations confirm the theoretical

predictions. In particular, they show that if $R > 4$, the piston will be able to reach and maintain for some time the velocity V_{stat} , whereas this will not be the case for $R < 1$ (**Figure 4b**). In the second stage of the evolution, the simulations (**Figure 4**) exhibit damped oscillations toward mechanical equilibrium which are in very good agreement with the predictions for the final state (X_{ad}, T_{ad}^{\pm}) , the frequency of oscillations and the existence of weak and strong damping depending on $R < 1$ or $R > 4$. Moreover, the general behavior of the evolution observed in the simulations as a function of the parameters was as predicted. However, the damping coefficient of these oscillations is wrong by one or several orders of magnitude. To understand this discrepancy, we note that using the average assumption we have related the damping to the friction coefficient. However, the simulations clearly show that those two dissipative effects have totally different origins. Indeed, as one can see with $L = \infty$, friction is associated with the fact that the density of the gas in front and in the back of the piston is not the same as in the bulk, and this generates a shock wave that propagates in the fluid. For finite L , when $R > 4$, the stationary velocity V_{stat} is reached and the effect of friction is

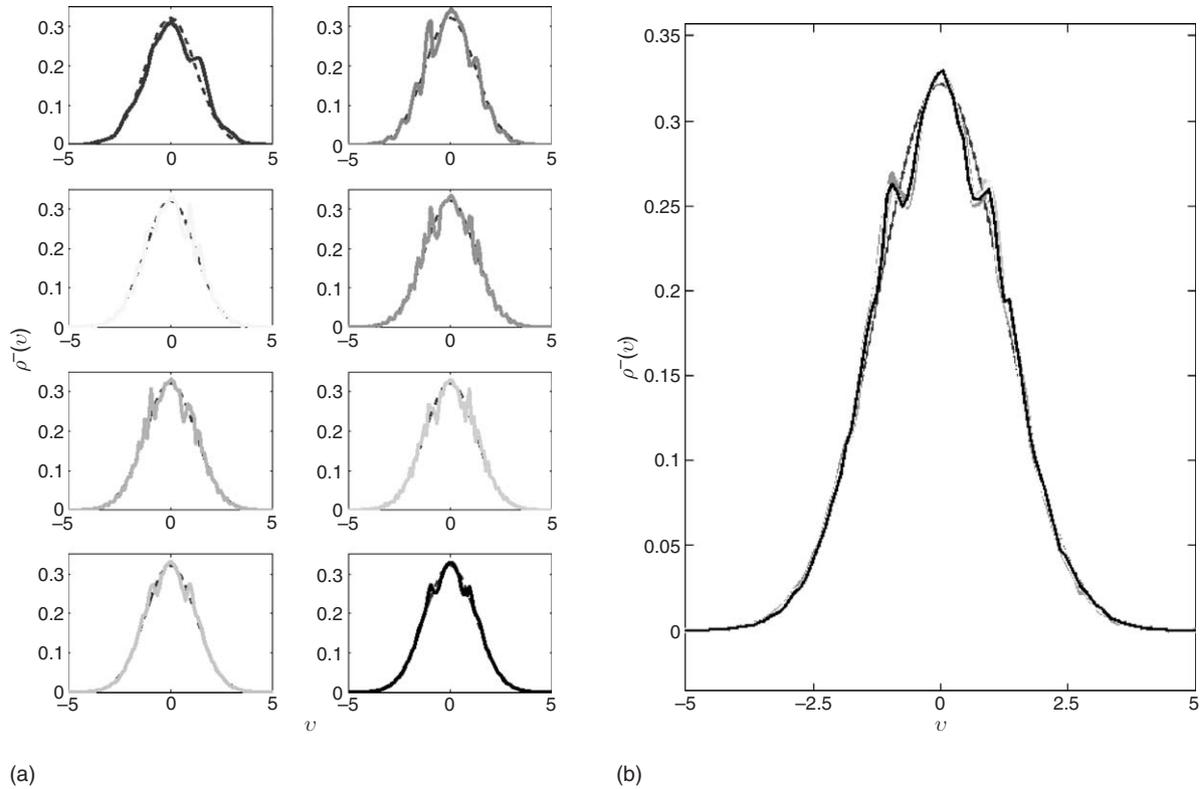


Figure 6 Velocity distribution in the left compartment. Same conditions as **Figure 4**, $R = 12$. Dotted line corresponds to Maxwellian with $T^- = 1.52$: (a) $t = 12, 24, 36, 48, 60, 92, 144, 240$ from top to bottom and (b) $t = 276-460$.

to transfer in this first stage more and more energy to the fluid on one side and vice versa on the other side. However, to stop the piston and reverse its motion, only a certain amount of the transferred energy is necessary and the rest remains as dissipated energy in the fluid leading to a strong damping. On the other hand, for $R < 1$, the value V_{stat} is never reached and all the energy transferred is necessary to revert the

motion. In this case very little dissipation is involved and the damping will be very small. This indicates that the mechanism responsible for damping is associated with shock waves bouncing back and forth and the average assumption, which corresponds to a homogeneity condition throughout the gas, cannot describe the situation. In fact, the simulations (**Figure 5b**) indicate that the average assumption does

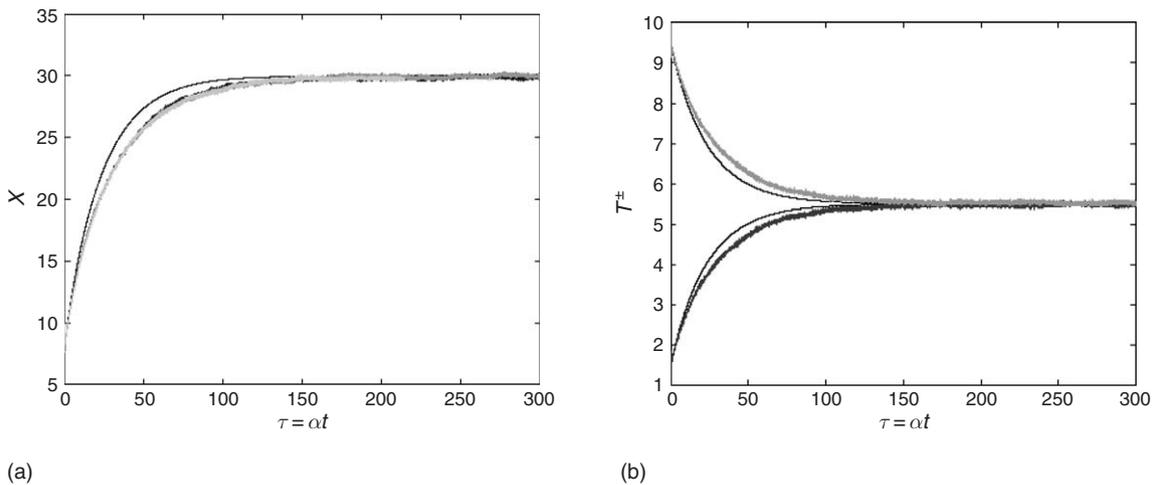


Figure 7 Approach to thermal equilibrium, $N^\pm = 3 \times 10^4$. The smooth curves correspond to the predictions, the stochastic curves to simulations: (a) position $X(\tau)$, $\tau = \alpha t$, no visible difference for $M = 100, 200, 1000$ and (b) average temperatures $T^\pm(\tau)$, $\tau = \alpha t$, $M = 200$.

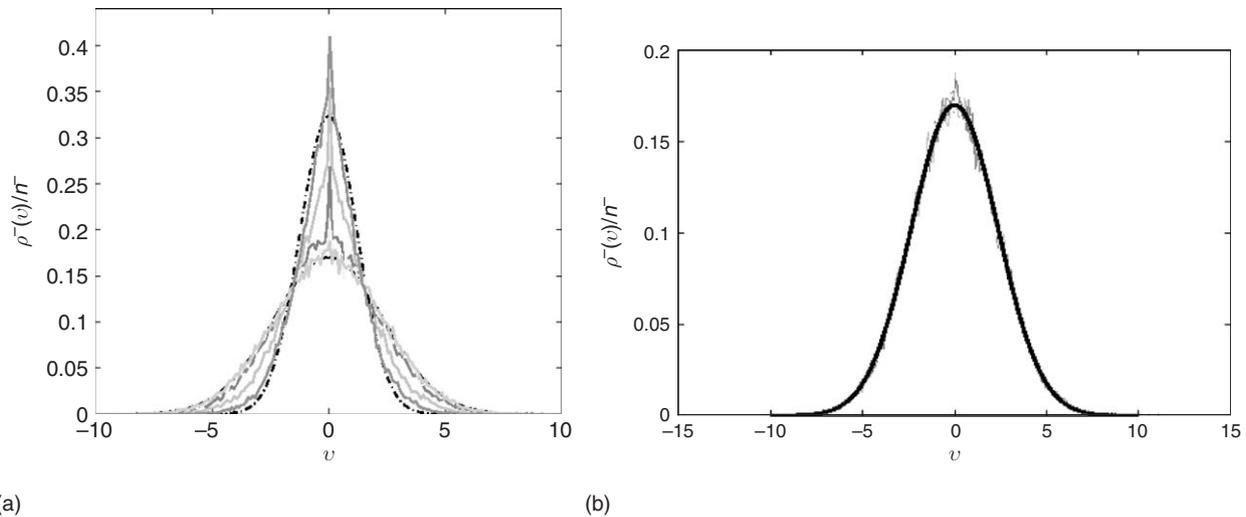


Figure 8 Approach to thermal equilibrium from $T_{\text{ad}}^- = 1.54$ (dotted line in(a)) to $T_f^- = 5.5$ (heavy line in (b)). Velocity distribution function on the left for $M = 200$, $N^\pm = 5 \times 10^4$. (a) $\tau = \alpha t = 2, 4, 14, 48, 92, 144$ and (b) approach to Maxwellian distribution for $\tau > 445$.

not hold in this second stage. In conclusion, one is forced to admit that to describe correctly the adiabatic evolution, it is necessary to study the coupling between the motion of the piston and the hydrodynamic equations of the gas. Preliminary investigations have been initiated, but this is still one of the major open problems. Another problem would be to study the evolution in the case of interacting particles. However, investigations with hard disks suggest that no new effects should appear. To investigate adiabatic evolution, a simpler version of the adiabatic piston problem, without any controversy, has been introduced: this is the model of a standard piston with a constant force acting on it.

In the third stage, that is, the very slow approach to thermal equilibrium, another assumption was necessary, namely the factorization condition. The simulations (Figure 7) show a very good agreement with the prediction, and in particular the scaling property with $t' = t/M$ is perfectly verified. It appears that the small discrepancy between simulations and theoretical predictions could be due to the fact that, to compute explicitly the coefficients in the equations of motion, we have taken Maxwellian relations for the velocities of the gas particles, which is clearly not the case (Figure 8a).

The fourth stage of the evolution, that is, the approach to Maxwellian distributions (Figure 8b), is still another major open problem. Some preliminary studies have been conducted, where one investigates the stability and the evolution of the system when initially the two gases are in the same equilibrium state, but characterized by a distribution function which is not Maxwellian.

Finally, let us mention that the relation between the piston problem and the second law of thermodynamics is one more major problem. The question of entropy production out of equilibrium, and the validity of the second law, are still highly controversial. Again, preliminary results can be found in the literature. Among other things, this question has led to a model of heat conductivity gases, which reproduces the correct behavior (Gruber and Lesne 2005).

See also: Billiards in Bounded Convex Domains; Boltzmann Equation (Classical and Quantum); Hamiltonian Fluid Dynamics; Multiscale Approaches; Nonequilibrium Statistical Mechanics (Stationary); Overview; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach.

Further Reading

- Callen HB (1963) *Thermodynamics*. New York: Wiley. (Appendix C. See also Callen HB (1985) *Thermodynamics and Thermostatistics*, 2nd edn., pp. 51 and 53. New York: Wiley.)
- Chernov N, Sinai YaG, and Lebowitz JL (2002) Scaling dynamic of a massive piston in a cube filled with ideal gas: exact results. *Journal of Statistical Physics* 109: 529–548.
- Feynman RP (1965) *Lectures in Physics I*. New York: Addison-Wesley.
- Gruber Ch (1999) Thermodynamics of systems with internal adiabatic constraints: time evolution of the adiabatic piston. *European Journal of Physics* 20: 259–266.
- Gruber Ch and Lesne A (2005) Hamiltonian model of heat conductivity and Fourier law. *Physica A* 351: 358.
- Gruber Ch, Pache S, and Lesne A (2003) Two-time-scale relaxation towards thermal equilibrium of the enigmatic piston. *Journal of Statistical Physics* 112: 1199–1228.

Gruber Ch and Piasecki J (1999) Stationary motion of the adiabatic piston. *Physica A* 268: 412–442.
 Kestemont E, Van den Broeck C, and MalekMM (2000) The “adiabatic” piston: and yet it moves. *Europhysics Letters* 49: 143.
 Lebowitz JL, Piasecki J, and Sinai YaG (2000) Scaling dynamics of a massive piston in an ideal gas. In: Szász D (ed.) *Hard*

Ball Systems and the Lorentz Gas, Encyclopedia of Mathematical Sciences Series, vol. 101, pp. 217–227. Berlin: Springer.

Van den Broeck C, Meurs P, and Kawai R (2004) From Maxwell demon to Brownian motor. *New Journal of Physics* 7: 10.

AdS/CFT Correspondence

C P Herzog, University of California at Santa Barbara, Santa Barbara, CA, USA

I R Klebanov, Princeton University, Princeton, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The anti-de Sitter/conformal field theory (AdS/CFT) correspondence is a conjectured equivalence between a quantum field theory in d spacetime dimensions with conformal scaling symmetry and a quantum theory of gravity in $(d + 1)$ -dimensional anti-de Sitter space. The most promising approaches to quantizing gravity involve superstring theories, which are most easily defined in 10 spacetime dimensions, or M -theory which is defined in 11 spacetime dimensions. Hence, the AdS/CFT correspondences based on superstrings typically involve backgrounds of the form $\text{AdS}_{d+1} \times Y_{9-d}$ while those based on M -theory involve backgrounds of the form $\text{AdS}_{d+1} \times Y_{10-d}$, where Y are compact spaces.

The examples of the AdS/CFT correspondence discussed in this article are dualities between (super)conformal nonabelian gauge theories and superstrings on $\text{AdS}_5 \times Y_5$, where Y_5 is a five-dimensional Einstein space (i.e., a space whose Ricci tensor is proportional to the metric, $R_{ij} = 4g_{ij}$). In particular, the most basic (and maximally supersymmetric) such duality relates $\mathcal{N} = 4$ $SU(N)$ super Yang–Mills (SYM) and type IIB superstring in the curved background $\text{AdS}_5 \times S^5$.

There exist special limits where this duality is more tractable than in the general case. If we take the large- N limit while keeping the 't Hooft coupling $\lambda = g_{\text{YM}}^2 N$ fixed (g_{YM} is the Yang–Mills coupling strength), then each Feynman graph of the gauge theory carries a topological factor N^χ , where χ is the Euler characteristic of the graph. The graphs of spherical topology (often called “planar”), to be identified with string tree diagrams, are weighted by N^2 ; the graphs of toroidal topology, to be identified

with string one-loop diagrams, by N^0 , etc. This counting corresponds to the closed-string coupling constant of order N^{-1} . Thus, in the large- N limit the gauge theory becomes “planar,” and the dual string theory becomes classical. For small $g_{\text{YM}}^2 N$, the gauge theory can be studied perturbatively; in this regime the dual string theory has not been very useful because the background becomes highly curved. The real power of the AdS/CFT duality, which already has made it a very useful tool, lies in the fact that, when the gauge theory becomes strongly coupled, the curvature in the dual description becomes small; therefore, classical supergravity provides a systematic starting point for approximating the string theory.

There is a strong motivation for an improved understanding of dualities of this type. In one direction, generalizations of this duality provide the tantalizing hope of a better understanding of quantum chromodynamics (QCD); QCD is a non-abelian gauge theory that describes the strong interactions of mesons, baryons, and glueballs, and has a conformal symmetry which is broken by quantum effects. In the other direction, AdS/CFT suggests that quantum gravity may be understandable as a gauge theory. Understanding the confinement of quarks and gluons that takes place in low-energy QCD and quantizing gravity are well acknowledged to be two of the most important outstanding problems of theoretical physics.

Some Geometrical Preliminaries

The d -dimensional sphere of radius L , S^d , may be defined by a constraint

$$\sum_{i=1}^{d+1} (X^i)^2 = L^2 \quad [1]$$

on $d + 1$ real coordinates X^i . It is a positively curved maximally symmetric space with symmetry group $SO(d + 1)$. We will denote the round metric on S^d of unit radius by $d\Omega_d^2$.

The d -dimensional anti-de Sitter space, AdS_d , may be defined by a constraint

$$(X^0)^2 + (X^d)^2 - \sum_{i=1}^{d-1} (X^i)^2 = L^2 \quad [2]$$

This constraint shows that the symmetry group of AdS_d is $\text{SO}(2, d-1)$. AdS_d is a negatively curved maximally symmetric space, that is, its curvature tensor is related to the metric by

$$R_{abcd} = -\frac{1}{L^2} [g_{ac}g_{bd} - g_{ad}g_{bc}] \quad [3]$$

Its metric may be written as

$$ds_{\text{AdS}}^2 = L^2 \left(-(y^2 + 1)dt^2 + \frac{dy^2}{y^2 + 1} + y^2 d\Omega_{d-2}^2 \right) \quad [4]$$

where the radial coordinate $y \in [0, \infty)$, and t is defined on a circle of length 2π . This space has closed timelike curves; to eliminate them, we will work with the universal covering space where $t \in (-\infty, \infty)$. The boundary of AdS_d , which plays an important role in the AdS/CFT correspondence, is located at infinite y . There exists a subspace of AdS_d called the Poincaré wedge, with the metric

$$ds^2 = \frac{L^2}{z^2} \left(dz^2 - (dx^0)^2 + \sum_{i=1}^{d-2} (dx^i)^2 \right) \quad [5]$$

where $z \in [0, \infty)$.

A Euclidean continuation of AdS_d is the Lobachevsky space (hyperboloid), L_d . It is obtained by reversing the sign of $(X^d)^2$, dt^2 , and $(dx^0)^2$ in [2], [4], and [5], respectively. After this Euclidean continuation, the metrics [4] and [5] become equivalent; both of them cover the entire L_d . Another equivalent way of writing the metric is

$$ds_L^2 = L^2 \left(d\rho^2 + \sinh^2 \rho d\Omega_{d-1}^2 \right) \quad [6]$$

which shows that the boundary at infinite ρ has the topology of S^{d-1} . In terms of the Euclideanized metric [5], the boundary consists of the \mathbf{R}^{d-1} at $z=0$, and a single point at $z=\infty$.

The Geometry of Dirichlet Branes

Our path toward formulating the $\text{AdS}_5/\text{CFT}_4$ correspondence requires introduction of Dirichlet branes, or D-branes for short. They are soliton-like “membranes” of various internal dimensionalities contained in type II superstring theories. A Dirichlet p -brane (or Dp brane) is a $(p+1)$ -dimensional hyperplane in $(9+1)$ -dimensional spacetime where strings are allowed to end. A D-brane is much like a

topological defect: upon touching a D-brane, a closed string can open up and turn into an open string whose ends are free to move along the D-brane. For the endpoints of such a string the $p+1$ longitudinal coordinates satisfy the conventional free (Neumann) boundary conditions, while the $9-p$ coordinates transverse to the Dp brane have the fixed (Dirichlet) boundary conditions, hence the origin of the term “Dirichlet brane.” The Dp brane preserves half of the bulk supersymmetries and carries an elementary unit of charge with respect to the $(p+1)$ -form gauge potential from the Ramond–Ramond (RR) sector of type II superstring.

For this article, the most important property of D-branes is that they realize gauge theories on their world volume. The massless spectrum of open strings living on a Dp brane is that of a maximally supersymmetric $U(1)$ gauge theory in $p+1$ dimensions. The $9-p$ massless scalar fields present in this supermultiplet are the expected Goldstone modes associated with the transverse oscillations of the Dp brane, while the photons and fermions provide the unique supersymmetric completion. If we consider N parallel D-branes, then there are N^2 different species of open strings because they can begin and end on any of the D-branes. N^2 is the dimension of the adjoint representation of $U(N)$, and indeed we find the maximally supersymmetric $U(N)$ gauge theory in this setting.

The relative separations of the Dp branes in the $9-p$ transverse dimensions are determined by the expectation values of the scalar fields. We will be interested in the case where all scalar expectation values vanish, so that the N Dp branes are stacked on top of each other. If N is large, then this stack is a heavy object embedded into a theory of closed strings which contains gravity. Naturally, this macroscopic object will curve space: it may be described by some classical metric and other background fields including the RR $(p+2)$ -form field strength. Thus, we have two very different descriptions of the stack of Dp branes: one in terms of the $U(N)$ supersymmetric gauge theory on its world volume, and the other in terms of the classical RR charged p -brane background of the type II closed superstring theory. The relation between these two descriptions is at the heart of the connections between gauge fields and strings that are the subject of this article.

Coincident D3 Branes

Gauge theories in $3+1$ dimensions play an important role in physics, and as explained above, parallel D3 branes realize a $(3+1)$ -dimensional $U(N)$ SYM

theory. Let us compare a stack of D3 branes with the RR-charged black 3-brane classical solution where the metric assumes the form

$$ds^2 = H^{-1/2}(r) \left[-f(r)(dx^0)^2 + (dx^i)^2 \right] + H^{1/2}(r) \left[f^{-1}(r)dr^2 + r^2 d\Omega_5^2 \right] \quad [7]$$

where $i = 1, 2, 3$ and

$$H(r) = 1 + \frac{L^4}{r^4}, \quad f(r) = 1 - \frac{r_0^4}{r^4}$$

The solution also contains an RR self-dual 5-form field strength

$$F = dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \wedge d(H^{-1}) + 4L^4 \text{vol}(S^5) \quad [8]$$

so that the Einstein equation of type IIB supergravity, $R_{\mu\nu} = F_{\mu\alpha\beta\gamma\delta} F_{\nu}{}^{\alpha\beta\gamma\delta} / 96$, is satisfied.

In the extremal limit $r_0 \rightarrow 0$, the 3-brane metric becomes

$$ds^2 = \left(1 + \frac{L^4}{r^4} \right)^{-1/2} \left(-(dx^0)^2 + (dx^i)^2 \right) + \left(1 + \frac{L^4}{r^4} \right)^{1/2} (dr^2 + r^2 d\Omega_5^2) \quad [9]$$

Just like the stack of parallel, ground-state D3 branes, the extremal solution preserves 16 of the 32 supersymmetries present in the type IIB theory. Introducing $z = L^2/r$, one notes that the limiting form of [9] as $r \rightarrow 0$ factorizes into the direct product of two smooth spaces, the Poincaré wedge [5] of AdS₅, and S⁵, with equal radii of curvature L . The 3-brane geometry may thus be viewed as a semi-infinite throat of radius L which, for $r \gg L$, opens up into flat (9 + 1)-dimensional space. Thus, for L much larger than the string length scale, $\sqrt{\alpha'}$, the entire 3-brane geometry has small curvatures everywhere and is appropriately described by the supergravity approximation to type IIB string theory.

The relation between L and $\sqrt{\alpha'}$ may be found by equating the gravitational tension of the extremal 3-brane classical solution to N times the tension of a single D3 brane:

$$\frac{2}{\kappa^2} L^4 \text{vol}(S^5) = N \frac{\sqrt{\pi}}{\kappa} \quad [10]$$

where $\text{vol}(S^5) = \pi^3$ is the volume of a unit 5-sphere, and $\kappa = \sqrt{8\pi G}$ is the ten-dimensional gravitational constant. It follows that

$$L^4 = \frac{\kappa}{2\pi^{5/2}} N = g_{\text{YM}}^2 N \alpha'^2 \quad [11]$$

where we used the standard relations $\kappa = 8\pi^{7/2} g_{st} \alpha'^2$ and $g_{\text{YM}}^2 = 4\pi g_{st}$ [10]. Thus, the size of the throat in string units is $\lambda^{1/4}$. This remarkable emergence of the 't Hooft coupling from gravitational considerations is at the heart of the success of the AdS/CFT correspondence. Moreover, the requirement $L \gg \sqrt{\alpha'}$ translates into $\lambda \gg 1$: the gravitational approach is valid when the 't Hooft coupling is very strong and the perturbative field-theoretic methods are not applicable.

Example: Thermal Gauge Theory from Near-Extremal D3 Branes

An important black hole observable is the Bekenstein–Hawking (BH) entropy, which is proportional to the area of the event horizon. For the 3-brane solution [7], the horizon is located at $r = r_0$. For $r_0 > 0$ the 3-brane carries some excess energy E above its extremal value, and the BH entropy is also non-vanishing. The Hawking temperature is then defined by $T^{-1} = \partial S_{\text{BH}} / \partial E$.

Setting $r_0 \ll L$ in [9], we obtain a near-extremal 3-brane geometry, whose Hawking temperature is found to be $T = r_0 / (\pi L^2)$. The eight-dimensional “area” of the horizon is

$$A_b = (r_0/L)^3 V_3 L^5 \text{vol}(S^5) = \pi^6 L^8 T^3 V_3 \quad [12]$$

where V_3 is the spatial volume of the D3 brane (i.e., the volume of the x^1, x^2, x^3 coordinates). Therefore, the BH entropy is

$$S_{\text{BH}} = \frac{2\pi A_b}{\kappa^2} = \frac{\pi^2}{2} N^2 V_3 T^3 \quad [13]$$

This gravitational entropy of a near-extremal 3-brane of Hawking temperature T is to be identified with the entropy of $\mathcal{N} = 4$ supersymmetric U(N) gauge theory (which lives on N coincident D3 branes) heated up to the same temperature.

The entropy of a free U(N) $\mathcal{N} = 4$ supermultiplet – which consists of the gauge field, $6N^2$ massless scalars, and $4N^2$ Weyl fermions – can be calculated using the standard statistical mechanics of a massless gas (the blackbody problem), and the answer is

$$S_0 = \frac{2\pi^2}{3} N^2 V_3 T^3 \quad [14]$$

It is remarkable that the 3-brane geometry captures the T^3 scaling characteristic of a conformal field theory (CFT) (in a CFT this scaling is guaranteed by the extensivity of the entropy and the absence of dimensionful parameters). Also, the N^2 scaling indicates the presence of $O(N^2)$ unconfined degrees

of freedom, which is exactly what we expect in the $\mathcal{N}=4$ supersymmetric $U(N)$ gauge theory. But what is the explanation of the relative factor of $3/4$ between S_{BH} and S_0 ? In fact, this factor is not a contradiction but rather a prediction about the strongly coupled $\mathcal{N}=4$ SYM theory at finite temperature. As we argued above, the supergravity calculation of the BH entropy, [13], is relevant to the $\lambda \rightarrow \infty$ limit of the $\mathcal{N}=4$ $SU(N)$ gauge theory, while the free-field calculation, [14], applies to the $\lambda \rightarrow 0$ limit. Thus, the relative factor of $3/4$ is not a discrepancy: it relates two different limits of the theory. Indeed, on general field-theoretic grounds, we expect that in the 't Hooft large- N limit, the entropy is given by

$$S = \frac{2\pi^2}{3} N^2 f(\lambda) V_3 T^3 \quad [15]$$

The function f is certainly not constant: perturbative calculations valid for small $\lambda = g_{\text{YM}}^2 N$ give

$$f(\lambda) = 1 - \frac{3}{2\pi^2} \lambda + \frac{3 + \sqrt{2}}{\pi^3} \lambda^{3/2} + \dots \quad [16]$$

Thus, the BH entropy in supergravity, [13], is translated into the prediction that

$$\lim_{\lambda \rightarrow \infty} f(\lambda) = \frac{3}{4} \quad [17]$$

The Essentials of the AdS/CFT Correspondence

The AdS/CFT correspondence asserts a detailed map between the physics of type IIB string theory in the throat of the classical 3-brane geometry, that is, the region $r \ll L$, and the gauge theory living on a stack of D3 branes. As already noted, in this limit $r \ll L$, the extremal D3 brane geometry factors into a direct product of $\text{AdS}_5 \times S^5$. Moreover, the gauge theory on this stack of D3 branes is the maximally supersymmetric $\mathcal{N}=4$ SYM.

Since the horizon of the near-extremal 3-brane lies in the region $r \ll L$, the entropy calculation could have been carried out directly in the throat limit, where $H(r)$ is replaced by L^4/r^4 . Another way to motivate the identification of the gauge theory with the throat is to think about the absorption of massless particles. In the D-brane description, a particle incident from asymptotic infinity is converted into an excitation of the stack of D-branes, that is, into an excitation of the gauge theory on the world volume. In the supergravity description, a

particle incident from the asymptotic (large r) region tunnels into the $r \ll L$ region and produces an excitation of the throat. The fact that the two different descriptions of the absorption process give identical cross sections supports the identification of excitations of $\text{AdS}_5 \times S^5$ with the excited states of the $\mathcal{N}=4$ SYM theory.

Maldacena (1998) motivated this correspondence by thinking about the low-energy ($\alpha' \rightarrow 0$) limit of the string theory. On the D3 brane side, in this low-energy limit, the interaction between the D3 branes and the closed strings propagating in the bulk vanishes, leaving a pure $\mathcal{N}=4$ SYM theory on the D3 branes decoupled from type IIB superstrings in flat space. Around the classical 3-brane solutions, there are two types of low-energy excitations. The first type propagate in the bulk region, $r \gg L$, and have a cross section for absorption by the throat which vanishes as the cube of their energy. The second type are localized in the throat, $r \leq L$, and find it harder to tunnel into the asymptotically flat region as their energy is taken smaller. Thus, both the D3 branes and the classical 3-brane solution have two decoupled components in the low-energy limit, and in both cases, one of these components is type IIB superstrings in flat space. Maldacena conjectured an equivalence between the other two components.

Immediate support for this identification comes from symmetry considerations. The isometry group of AdS_5 is $SO(2,4)$, and this is also the conformal group in $3+1$ dimensions. In addition, we have the isometries of S^5 which form $SU(4) \sim SO(6)$. This group is identical to the R-symmetry of the $\mathcal{N}=4$ SYM theory. After including the fermionic generators required by supersymmetry, the full isometry supergroup of the $\text{AdS}_5 \times S^5$ background is $SU(2,2|4)$, which is identical to the $\mathcal{N}=4$ superconformal symmetry. We will see that, in theories with reduced supersymmetry, the S^5 factor is replaced by other compact Einstein spaces Y_5 , but AdS_5 is the “universal” factor present in the dual description of any large- N CFT and makes the $SO(2,4)$ conformal symmetry a geometric one.

The correspondence extends beyond the supergravity limit, and we must think of $\text{AdS}_5 \times Y_5$ as a background of string theory. Indeed, type IIB strings are dual to the electric flux lines in the gauge theory, providing a string-theoretic setup for calculating correlation functions of Wilson loops. Furthermore, if $N \rightarrow \infty$ while $g_{\text{YM}}^2 N$ is held fixed and finite, then there are string scale corrections to the supergravity limit (Maldacena 1998, Gubser *et al.* 1998, Witten 1998) which proceed in powers of $\alpha'/L^2 = (g_{\text{YM}}^2 N)^{-1/2}$. For finite N , there are also

string loop corrections in powers of $\kappa^2/L^8 \sim N^{-2}$. As expected, with $N \rightarrow \infty$ we can take the classical limit of the string theory on $\text{AdS}_5 \times Y_5$. However, in order to understand the large- N gauge theory with finite 't Hooft coupling, we should think of $\text{AdS}_5 \times Y_5$ as the target space of a two-dimensional sigma model describing the classical string physics.

Correlation Functions and the Bulk/Boundary Correspondence

A basic premise of the AdS/CFT correspondence is the existence of a one-to-one map between gauge-invariant operators in the CFT and fields (or extended objects) in AdS. Gubser *et al.* (1998) and Witten (1998) formulated precise methods for calculating correlation functions of various operators in a CFT using its dual formulation. A physical motivation for these methods comes from earlier calculations of absorption by 3-branes. When a wave is absorbed, it tunnels from asymptotic infinity into the throat region, and then continues to propagate toward smaller r . Let us separate the 3-brane geometry into two regions: $r \gtrsim L$ and $r \lesssim L$. For $r \lesssim L$ the metric is approximately that of $\text{AdS}_5 \times S^5$, while for $r \gtrsim L$ it becomes very different and eventually approaches the flat metric. Signals coming in from large r (small $z = L^2/r$) may be considered as disturbing the “boundary” of AdS_5 at $r \sim L$, and then propagating into the bulk of AdS_5 . Discarding the $r \gtrsim L$ part of the 3-brane metric, the gauge theory correlation functions are related to the response of the string theory to boundary conditions at $r \sim L$. It is therefore natural to identify the generating functional of correlation functions in the gauge theory with the string theory path integral subject to the boundary conditions that $\phi(\mathbf{x}, z) = \phi_0(\mathbf{x})$ at $z = L$ (at $z = \infty$ all fluctuations are required to vanish). In calculating correlation functions in a CFT, we will carry out the standard Euclidean continuation; then on the string theory side, we will work with L_5 , which is the Euclidean version of AdS_5 .

More explicitly, we identify a gauge theory quantity W with a string-theory quantity Z_{string} :

$$W[\phi_0(\mathbf{x})] = Z_{\text{string}}[\phi_0(\mathbf{x})] \quad [18]$$

W generates the connected Euclidean Green's functions of a gauge-theory operator \mathcal{O} ,

$$W[\phi_0(\mathbf{x})] = \left\langle \exp \int d^4x \phi_0 \mathcal{O} \right\rangle \quad [19]$$

Z_{string} is the string theory path integral calculated as a functional of ϕ_0 , the boundary condition on the field ϕ related to \mathcal{O} by the AdS/CFT duality. In the

large- N limit, the string theory becomes classical which implies

$$Z_{\text{string}} \sim e^{-I[\phi_0(\mathbf{x})]} \quad [20]$$

where $I[\phi_0(\mathbf{x})]$ is the extremum of the classical string action calculated as a functional of ϕ_0 . If we are further interested in correlation functions at very large 't Hooft coupling, then the problem of extremizing the classical string action reduces to solving the equations of motion in type IIB supergravity whose form is known explicitly. A simple example of such a calculation is presented in the next subsection.

Our reasoning suggests that from the point of view of the metric [5], the boundary conditions are imposed not quite at $z=0$, which is the true boundary of L_5 , but at some finite value $z=\epsilon$. It does not matter which value it is since the metric [5] is unchanged by an overall rescaling of the coordinates (z, \mathbf{x}) ; thus, such a rescaling can take $z=L$ into $z=\epsilon$ for any ϵ . The physical meaning of this cutoff is that it acts as a UV regulator in the gauge theory. Indeed, the radial coordinate z is to be considered as the effective energy scale of the gauge theory, and decreasing z corresponds to increasing the energy. A safe method for performing calculations of correlation functions, therefore, is to keep the cutoff on the z -coordinate at intermediate stages and remove it only at the end.

Two-Point Functions and Operator Dimensions

In the following, we present a brief discussion of two-point functions of scalar operators in CFT_d . The corresponding field in L_{d+1} is a scalar field of mass m whose Euclidean action is proportional to

$$\frac{1}{2} \int d^d x dz z^{-d+1} \left[(\partial_z \phi)^2 + \sum_{a=1}^d (\partial_a \phi)^2 + \frac{m^2 L^2}{z^2} \phi^2 \right] \quad [21]$$

In calculating correlation functions of vertex operators from the AdS/CFT correspondence, the first problem is to reconstruct an on-shell field in L_{d+1} from its boundary behavior. The near-boundary, that is, small z , behavior of the classical solution is

$$\begin{aligned} \phi(z, \mathbf{x}) \rightarrow & z^{d-\Delta} [\phi_0(\mathbf{x}) + O(z^2)] \\ & + z^\Delta [A(\mathbf{x}) + O(z^2)] \end{aligned} \quad [22]$$

where Δ is one of the roots of

$$\Delta(\Delta - d) = m^2 L^2 \quad [23]$$

$\phi_0(\mathbf{x})$ is regarded as a “source” in [19] that couples to the dual gauge-invariant operator \mathcal{O} of dimension Δ , while $A(\mathbf{x})$ is related to the expectation value,

$$A(\mathbf{x}) = \frac{1}{2\Delta - d} \langle \mathcal{O}(\mathbf{x}) \rangle \quad [24]$$

It is possible to regularize the Euclidean action to obtain the following value as a functional of the source:

$$I[\phi_0(\mathbf{x})] = -(\Delta - (d/2))\pi^{-d/2} \frac{\Gamma(\Delta)}{\Gamma(\Delta - (d/2))} \times \int d^d \mathbf{x} \int d^d \mathbf{x}' \frac{\phi_0(\mathbf{x})\phi_0(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^{2\Delta}} \quad [25]$$

Varying twice with respect to ϕ_0 , we find that the two-point function of the corresponding operator is

$$\langle \mathcal{O}(\mathbf{x})\mathcal{O}(\mathbf{x}') \rangle = \frac{(2\Delta - d)\Gamma(\Delta)}{\pi^{d/2}\Gamma(\Delta - (d/2))} \frac{1}{|\mathbf{x} - \mathbf{x}'|^{2\Delta}} \quad [26]$$

Which of the two roots, Δ_+ or Δ_- , of [23]

$$\Delta_{\pm} = \frac{d}{2} \pm \sqrt{\frac{d^2}{4} + m^2 L^2} \quad [27]$$

should we choose for the operator dimension? For positive m^2 , Δ_+ is certainly the right choice: here the other root, Δ_- , is negative. However, it turns out that for

$$-\frac{d^2}{4} < m^2 L^2 < -\frac{d^2}{4} + 1 \quad [28]$$

both roots of [23] may be chosen. Thus, there are two possible CFTs corresponding to the same classical AdS action: in one of them the corresponding operator has dimension Δ_+ , while in the other the dimension is Δ_- . We note that Δ_- is bounded from below by $(d-2)/2$, which is precisely the unitarity bound on dimensions of scalar operators in d -dimensional field theory! Thus, the ability to choose dimension Δ_- is crucial for consistency of the AdS/CFT duality.

Whether string theory on $\text{AdS}_5 \times Y_5$ contains fields with m^2 in the range [28] depends on Y_5 . The example discussed in the next section, $Y_5 = T^{1,1}$, turns out to contain such fields, and the possibility of having dimension Δ_- , [27], is crucial for consistency of the AdS/CFT duality in that case. However, for $Y_5 = S^5$, which is dual to the $\mathcal{N}=4$ large- N SYM theory, there are no such fields and all scalar dimensions are given by [27].

The operators in the $\mathcal{N}=4$ large- N SYM theory naturally break up into two classes: those that correspond to the Kaluza–Klein states of supergravity and those that correspond to massive string

states. Since the radius of the S^5 is L , the masses of the Kaluza–Klein states are proportional to $1/L$. Thus, the dimensions of the corresponding operators are independent of L and therefore also of λ . On the gauge-theory side, this independence is explained by the fact that the supersymmetry protects the dimensions of certain operators from being renormalized: they are completely determined by the representation under the superconformal symmetry. All families of the Kaluza–Klein states, which correspond to such protected operators, were classified long ago. Correlation functions of such operators in the strong 't Hooft coupling limit may be obtained from the dependence of the supergravity action on the boundary values of corresponding Kaluza–Klein fields, as in [19]. A variety of explicit calculations have been performed for two-, three-, and even four-point functions. The four-point functions are particularly interesting because their dependence on operator positions is not determined by the conformal invariance.

On the other hand, the masses of string excitations are $m^2 = 4n/\alpha'$, where n is an integer. For the corresponding operators the formula [27] predicts that the dimensions do depend on the 't Hooft coupling and, in fact, blow up for large $\lambda = g_{\text{YM}}^2 N$ as $2\lambda^{1/4} \sqrt{n}$.

Calculation of Wilson Loops

The Wilson loop operator of a nonabelian gauge theory

$$W(\mathcal{C}) = \text{tr} \left[P \exp \left(i \oint_{\mathcal{C}} A \right) \right] \quad [29]$$

involves the path-ordered integral of the gauge connection A along a contour \mathcal{C} . For $\mathcal{N}=4$ SYM, one typically uses a generalization of this loop operator which incorporates other fields in the $\mathcal{N}=4$ multiplet, the adjoint scalars and fermions. Using a rectangular contour, we can calculate the quark–antiquark potential from the expectation value $\langle W(\mathcal{C}) \rangle$. One thinks of the quarks located a distance L apart for a time T , yielding

$$\langle W \rangle \sim e^{-TV(L)} \quad [30]$$

where $V(L)$ is the potential.

According to Maldacena, and Rey and Yee, the AdS/CFT correspondence relates the Wilson loop expectation value to a sum over string world sheets ending on the boundary of $L_5(z=0)$ along the contour \mathcal{C} :

$$\langle W \rangle \sim \int e^{-S} \quad [31]$$

where S is the action functional of the string world sheet. In the large 't Hooft coupling limit $\lambda \rightarrow \infty$, this path integral may be evaluated using a saddle-point approximation. The leading answer is $\sim e^{-S_0}$, where S_0 is the action for the classical solution, which is proportional to the minimal area of the string world sheet in L_5 subject to the boundary conditions. The area as currently defined is actually divergent, and to regularize it one must position the contour at $z = \epsilon$ (this is the same type of regulator as used in the definition of correlation functions).

Consider a circular Wilson loop of radius a . The action of the corresponding classical string world sheet is

$$S_0 = \sqrt{\lambda} \left(\frac{a}{\epsilon} - 1 \right) \quad [32]$$

Subtracting the linearly divergent term, which is proportional to the length of the contour, one finds

$$\ln \langle W \rangle = \sqrt{\lambda} + O(\ln \lambda) \quad [33]$$

a result which has been duplicated in field theory by summing certain classes of rainbow Feynman diagrams in $\mathcal{N} = 4$ SYM. From these sums, one finds

$$\langle W \rangle_{\text{rainbow}} = \frac{2}{\sqrt{\lambda}} I_1(\sqrt{\lambda}) \quad [34]$$

where I_1 is a Bessel function. This formula is one of the few available proposals for extrapolation of an observable from small to large coupling. At large λ ,

$$\langle W \rangle_{\text{rainbow}} \sim \sqrt{\frac{2}{\pi}} \frac{e^{\sqrt{\lambda}}}{\lambda^{3/4}} \quad [35]$$

in agreement with the geometric prediction.

The quark–antiquark potential is extracted from a rectangular Wilson loop of width L and length T . After regularizing the divergent contribution to the energy, one finds the attractive potential

$$V(L) = -\frac{4\pi^2 \sqrt{\lambda}}{\Gamma(1/4)^4 L} \quad [36]$$

The Coulombic $1/L$ dependence is required by the conformal invariance of the theory. The fact that the potential scales as the square root of the 't Hooft coupling indicates some screening of the charges at large coupling.

Conformal Field Theories and Einstein Manifolds

Interesting generalizations of the duality between $\text{AdS}_5 \times S^5$ and $\mathcal{N} = 4$ SYM with less supersymmetry and more complicated gauge groups can be

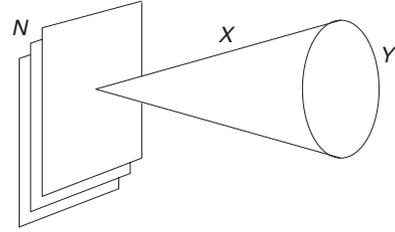


Figure 1 D3 branes placed at the tip of a Ricci-flat cone X .

produced by placing D3 branes at the tip of a Ricci-flat six-dimensional cone X (see Figure 1). The cone metric may be cast in the form

$$ds_X^2 = dr^2 + r^2 ds_Y^2 \quad [37]$$

where Y is the level surface of X . In particular, Y is a positively curved Einstein manifold, that is, one for which $R_{ij} = 4g_{ij}$. In order to preserve the $\mathcal{N} = 1$ supersymmetry, X must be a Calabi–Yau space; then Y is defined to be Sasaki–Einstein.

The D3 branes appear as a point in X and span the transverse Minkowski space $\mathbb{R}^{3,1}$. The ten-dimensional metric they produce assumes the form [9], but with the sphere metric $d\Omega_5^2$ replaced by the metric on Y , ds_Y^2 . The equality of tensions [10] now requires that

$$L^4 = \frac{\sqrt{\pi} \kappa N}{2 \text{vol}(Y)} = 4\pi g_s N \alpha'^2 \frac{\pi^3}{\text{vol}(Y)} \quad [38]$$

In the near-horizon limit, $r \rightarrow 0$, the geometry factors into $\text{AdS}_5 \times Y$. Because the D3 branes are located at a singularity, the gauge theory becomes much more complicated, typically involving a product of several $\text{SU}(N)$ factors coupled to matter in bifundamental representations, often described using a quiver diagram (see Figure 2 for an example).

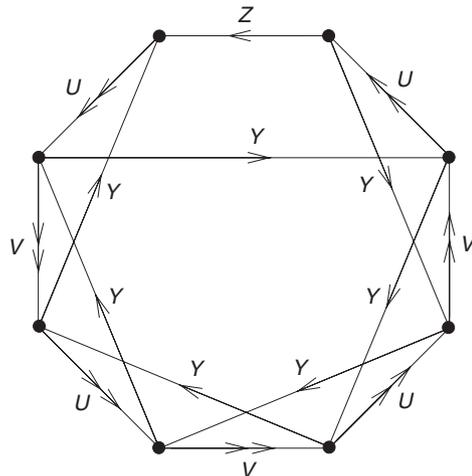


Figure 2 The quiver for $Y^{4,3}$. Each node corresponds to an $\text{SU}(N)$ gauge group and each arrow to a bifundamental chiral superfield.

The simplest examples of X are orbifolds \mathbb{C}^3/Γ , where Γ is a discrete subgroup of $\text{SO}(6)$. Indeed, if $\Gamma \subset \text{SU}(3)$, then $\mathcal{N} = 1$ supersymmetry is preserved. The level surface of such an X is $Y = S^5/\Gamma$. In this case, the product structure of the gauge theory can be motivated by thinking about image stacks of D3 branes from the action of Γ .

The next simplest example of a Calabi–Yau cone X is the conifold which may be described by the following equation in four complex variables:

$$\sum_{a=1}^4 z_a^2 = 0 \quad [39]$$

Since this equation is symmetric under an overall rescaling of the coordinates, this space is a cone. The level surface Y of the conifold is a coset manifold $T^{1,1} = (\text{SU}(2) \times \text{SU}(2))/\text{U}(1)$. This space has the $\text{SO}(4) \sim \text{SU}(2) \times \text{SU}(2)$ symmetry which rotates the z 's, and also the $\text{U}(1)$ R-symmetry under $z_a \rightarrow e^{i\theta} z_a$. The metric on $T^{1,1}$ is known explicitly; it assumes the form of an S^1 bundle over $S^2 \times S^2$.

The supersymmetric field theory on the D3 branes probing the conifold singularity is $\text{SU}(N) \times \text{SU}(N)$ gauge theory coupled to two chiral superfields, A_i , in the (N, \bar{N}) representation and two chiral superfields, B_j , in the (\bar{N}, N) representation. The A 's transform as a doublet under one of the global $\text{SU}(2)$'s, while the B 's transform as a doublet under the other $\text{SU}(2)$. Cancellation of the anomaly in the $\text{U}(1)$ R-symmetry requires that the A 's and the B 's each have R-charge $1/2$. For consistency of the duality, it is necessary that we add an exactly marginal superpotential which preserves the $\text{SU}(2) \times \text{SU}(2) \times \text{U}(1)_R$ symmetry of the theory. Since a marginal superpotential has R-charge equal to 2 it must be quartic, and the symmetries fix it uniquely up to overall normalization:

$$W = \epsilon^{ij} \epsilon^{kl} \text{tr} A_i B_k A_j B_l \quad [40]$$

There are in fact infinite families of Calabi–Yau cones X , but there are two problems one faces in studying these generalized AdS/CFT correspondences. The first is geometric: the cones X are not all well understood and only for relatively few do we have explicit metrics. However, it is often possible to calculate important quantities such as the $\text{vol}(Y)$ without knowing the metric. The second problem is gauge theoretic: although many techniques exist, there is no completely general procedure for constructing the gauge theory on a stack of D-branes at an arbitrary singularity.

Let us mention two important classes of Calabi–Yau cones X . The first class consists of cones over the so-called $Y^{p,q}$ Sasaki–Einstein spaces. Here, p

and q are integers with $p \geq q$. Gauntlett *et al.* (2004) discovered metrics on all the $Y^{p,q}$, and the quiver gauge theories that live on the D-branes probing the singularity are now known. Making contact with the simpler examples discussed above, the $Y^{p,0}$ are orbifolds of $T^{1,1}$ while the $Y^{p,p}$ are orbifolds of S^5 .

In the second class of cones X , a del Pezzo surface shrinks to zero size at the tip of the cone. A del Pezzo surface is an algebraic surface of complex dimension 2 with positive first Chern class. One simple del Pezzo surface is a complex projective space of dimension 2, \mathbb{P}^2 , which gives rise to the $\mathcal{N} = 1$ preserving S^5/\mathbb{Z}_3 orbifold. Another simple case is $\mathbb{P}^1 \times \mathbb{P}^1$, which leads to $T^{1,1}/\mathbb{Z}_2$. The remaining del Pezzo surfaces B_k are \mathbb{P}^2 blown up at k points, $1 \leq k \leq 8$. The cone where B_1 shrinks to zero size has level surface $Y^{2,1}$. Gauge theories for all the del Pezzos have been constructed. Except for the three del Pezzos just discussed, and possibly also for B_6 , metrics on the cones over these del Pezzos are not known. Nevertheless, it is known that for $3 \leq k \leq 8$, the volume of the Sasaki–Einstein manifold Y associated with B_k is $\pi^3(9 - k)/27$.

The Central Charge

The central charge provides one of the most amazing ways to check the generalized AdS/CFT correspondences. The central charge c and conformal anomaly a can be defined as coefficients of certain curvature invariants in the trace of the stress energy tensor of the conformal gauge theory:

$$\langle T_\alpha^\alpha \rangle = -aE_4 - cI_4 \quad [41]$$

(The curvature invariants E_4 and I_4 are quadratic in the Riemann tensor and vanish for Minkowski space.) As discussed above, correlators such as $\langle T_{\mu\nu} \rangle$ can be calculated from supergravity, and one finds

$$a = c = \frac{\pi^3 N^2}{4 \text{vol}(Y)} \quad [42]$$

On the gauge-theory side of the correspondence, anomalies completely determine a and c :

$$\begin{aligned} a &= \frac{3}{32} (3 \text{tr} R^3 - \text{tr} R) \\ c &= \frac{1}{32} (9 \text{tr} R^3 - 5 \text{tr} R) \end{aligned} \quad [43]$$

The trace notation implies a sum over the R-charges of all of the fermions in the gauge theory. (From the geometric knowledge that $a = c$, we can conclude that $\text{tr} R = 0$.)

The R-charges can be determined using the principle of a -maximization. For a superconformal gauge theory, the R-charges of the fermions maximize a subject to the constraints that the

Novikov–Shifman–Vainshtein–Zakharov (NSVZ) beta function of each gauge group vanishes and the R-charge of each superpotential term is 2.

For the $Y^{p,q}$ spaces mentioned above, one finds that

$$\text{vol}(Y^{p,q}) = \frac{q^2(2p + \sqrt{4p^2 - 3q^2})}{3p^2(3q^2 - 2p^2 + p\sqrt{4p^2 - 3q^2})} \pi^3 \quad [44]$$

The gauge theory consists of $p - q$ fields Z , $p + q$ fields Y , $2p$ fields U , and $2q$ fields V . These fields all transform in the bifundamental representation of a pair of $SU(N)$ gauge groups (the quiver diagram for $Y^{4,3}$ is given in [Figure 2](#)). The NSVZ beta function and superpotential constraints determine the R-charges up to two free parameters x and y . Let x be the R-charge of Z and y the R-charge of Y . Then the U have R-charge $1 - (1/2)(x + y)$ and the V have R-charge $1 + (1/2)(x - y)$.

The technique of a maximization leads to the result

$$x = \frac{1}{3q^2} \left(-4p^2 + 2pq + 3q^2 + (2p - q)\sqrt{4p^2 - 3q^2} \right)$$

$$y = \frac{1}{3q^2} \left(-4p^2 - 2pq + 3q^2 + (2p + q)\sqrt{4p^2 - 3q^2} \right)$$

Thus, as calculated by [Benvenuti et al. \(2004\)](#) and [Bertolini et al. \(2004\)](#)

$$a(Y^{p,q}) = \frac{\pi^3 N^2}{4 \text{vol}(Y^{p,q})} \quad [45]$$

in remarkable agreement with the prediction [\[42\]](#) of the AdS/CFT duality.

A Path to a Confining Theory

There exists an interesting way of breaking the conformal invariance for spaces Y whose topology includes an S^2 factor (examples of such spaces include $T^{1,1}$ and $Y^{p,q}$, which are topologically $S^2 \times S^3$). At the tip of the cone over Y , one may add M wrapped D5 branes to the N D3 branes. The gauge theory on such a combined stack is no longer conformal; it exhibits a novel pattern of quasiperiodic renormalization group flow, called a duality cascade.

To date, the most extensive study of a theory of this type has been carried out for the conifold, where one finds an $\mathcal{N} = 1$ supersymmetric $SU(N) \times SU(N + M)$ theory coupled to chiral superfields A_1, A_2 in the $(N, \overline{N + M})$ representation, and B_1, B_2 in the $(\overline{N}, N + M)$ representation. D5 branes source RR 3-form flux; hence, the supergravity dual of this theory has to include M units of this flux. [Klebanov and Strassler \(2000\)](#) found an exact nonsingular supergravity solution incorporating the 3-form and

the 5-form RR field strengths, and their back-reaction on the geometry. This back-reaction creates a “geometric transition” to the deformed conifold

$$\sum_{a=1}^4 z_a^2 = \epsilon^2 \quad [46]$$

and introduces a “warp factor” so that the full ten-dimensional geometry has the form

$$ds_{10}^2 = b^{-1/2}(\tau) (-(dx^0)^2 + (dx^i)^2) + b^{1/2}(\tau) d\tilde{s}_6^2 \quad [47]$$

where $d\tilde{s}_6^2$ is the Calabi–Yau metric of the deformed conifold, which is known explicitly.

The field-theoretic interpretation of this solution is unconventional. After a finite amount of RG flow, the $SU(N + M)$ group undergoes a Seiberg duality transformation. After this transformation, and an interchange of the two gauge groups, the new gauge theory is $SU(\tilde{N}) \times SU(\tilde{N} + M)$ with the same matter and superpotential, and with $\tilde{N} = N - M$. The self-similar structure of the gauge theory under the Seiberg duality is the crucial fact that allows this pattern to repeat many times. If $N = (k + 1)M$, where k is an integer, then the duality cascade stops after k steps, and we find $SU(M) \times SU(2M)$ gauge theory. This IR gauge theory exhibits a multitude of interesting effects visible in the dual supergravity background. One of them is confinement, which follows from the fact that the warp factor b is finite and nonvanishing at the smallest radial coordinate, $\tau = 0$. The methods presented in the section “Calculation of Wilson loops,” then imply that the quark–antiquark potential grows linearly at large distances. Other notable IR effects are chiral symmetry breaking and the Goldstone mechanism. Particularly interesting is the appearance of an entire “baryonic branch” of the moduli space in the gauge theory, whose existence has been demonstrated also in the dual supergravity language.

Conclusions

This article tries to present a logical path from studying gravitational properties of D-branes to the formulation of an exact duality between conformal field theories and string theory in anti-de Sitter backgrounds, and also sketches some methods for breaking the conformal symmetry. Due to space limitations, many aspects and applications of the AdS/CFT correspondence have been omitted. At the moment, practical applications of this duality are limited mainly to very strongly coupled, large- N gauge theories, where the dual string description is well approximated by classical supergravity. To understand the implications of the duality for more general parameters, it is necessary to find better

methods for attacking the world sheet approach to string theories in anti-de Sitter backgrounds with RR background fields turned on. When such methods are found, it is likely that the material presented here will have turned out to be just a tiny tip of a monumental iceberg of dualities between fields and strings.

Acknowledgments

The authors are very grateful to all their collaborators on gauge/string duality for their valuable input over many years. The research of I R Klebanov is supported in part by the National Science Foundation (NSF) grant no. PHY-0243680. The research of C P Herzog is supported in part by the NSF under grant no. PHY99-07949. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

See also: Brane Construction of Gauge Theories; Branes and Black Hole Statistical Mechanics; Einstein Equations: Exact Solutions; Gauge Theories from Strings; Large- N and Topological Strings; Large- N Dualities; Mirror Symmetry: A Geometric Survey; Quantum Chromodynamics; Quantum Field Theory in Curved Spacetime; Superstring Theories.

Further Reading

Aharony O, Gubser SS, Maldacena JM, Ooguri H, and Oz Y (2000) Large N field theories, string theory and gravity. *Physics Reports* 323: 183 (arXiv:hep-th/9905111).
 Benvenuti S, Franco S, Hanany A, Martelli D, and Sparks J (2005) An infinite family of superconformal quiver gauge theories with Sasaki–Einstein duals. *JHEP* 0506: 064 (arXiv:hep-th/0411264).

Bertolini M, Bigazzi F, and Cotrone AL (2004) New checks and subtleties for AdS/CFT and a-maximization. *JHEP* 0412: 024 (arXiv:hep-th/0411249).
 Bigazzi F, Cotrone AL, Petrini M, and Zaffaroni A (2002) Supergravity duals of supersymmetric four dimensional gauge theories. *Rivista del Nuovo Cimento* 25N12: 1 (arXiv:hep-th/0303191).
 D’Hoker E and Freedman DZ (2002) Supersymmetric gauge theories and the AdS/CFT correspondence, arXiv:hep-th/0201253.
 Gauntlett J, Martelli D, Sparks J, and Waldram D (2004) Sasaki–Einstein metrics on $S^2 \times S^3$. *Advances in Theoretical Mathematics in Physics* 8: 711 (arXiv:hep-th/0403002).
 Gubser SS, Klebanov IR, and Polyakov AM (1998) Gauge theory correlators from noncritical string theory. *Physics Letters B* 428: 105 (hep-th/9802109).
 Herzog CP, Klebanov IR, and Ouyang P (2002) D-branes on the conifold and $N=1$ gauge/gravity dualities, arXiv:hep-th/0205100.
 Klebanov IR (2000) TASI lectures: introduction to the AdS/CFT correspondence, arXiv:hep-th/0009139.
 Klebanov IR and Strassler MJ (2000) Supergravity and a confining gauge theory: Duality cascades and χ -resolution of naked singularities. *JHEP* 0008: 052 (arXiv:hep-th/0007191).
 Maldacena J (1998) The large N limit of superconformal field theories and supergravity. *Advances in Theoretical and Mathematical Physics* 2: 231 (hep-th/9711200).
 Maldacena JM (1998) Wilson loops in large N field theories. *Physics Review Letters* 80: 4859 (arXiv:hep-th/9803002).
 Polchinski J (1998) *String Theory*. Cambridge: Cambridge University Press.
 Polyakov AM (1999) The wall of the cave. *International Journal of Modern Physics A* 14: 645.
 Rey SJ and Yee JT (2001) Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity. *European Physics Journal C* 22: 379 (arXiv:hep-th/9803001).
 Semenoff GW and Zarembo K (2002) Wilson loops in SYM theory: from weak to strong coupling. *Nuclear Physics Proceeding Supplements* 108: 106 (arXiv:hep-th/0202156).
 Strassler MJ The duality cascade, TASI 2003 lectures, arXiv:hep-th/0505153.
 Witten E (1998) Anti-de Sitter space and holography. *Advances in Theoretical and Mathematical Physics* 2: 253 (hep-th/9802150).

Affine Quantum Groups

G W Delius and N MacKay, University of York, York, UK

© 2006 G W Delius. Published by Elsevier Ltd.
 All rights reserved.

Affine quantum groups are certain pseudoquasitriangular Hopf algebras that arise in mathematical physics in the context of integrable quantum field theory, integrable quantum spin chains, and solvable lattice models. They provide the algebraic framework behind the spectral parameter dependent Yang–Baxter equation

$$\begin{aligned} R_{12}(u)R_{13}(u+v)R_{23}(v) \\ = R_{23}(v)R_{13}(u+v)R_{12}(u) \end{aligned} \quad [1]$$

One can distinguish three classes of affine quantum groups, each leading to a different dependence of the R -matrices on the spectral parameter u : Yangians lead to rational R -matrices, quantum affine algebras lead to trigonometric R -matrices, and elliptic quantum groups lead to elliptic R -matrices. We will mostly concentrate on the quantum affine algebras but many results hold similarly for the other classes.

After giving mathematical details about quantum affine algebras and Yangians in the first two sections, we describe how these algebras arise in different areas of mathematical physics in the three following sections. We end with a description of boundary quantum groups which extend the formalism to the boundary Yang–Baxter (reflection) equation.

Quantum Affine Algebras

Definition

A quantum affine algebra $U_q(\hat{\mathfrak{g}})$ is a quantization of the enveloping algebra $U(\hat{\mathfrak{g}})$ of an affine Lie algebra (Kac–Moody algebra) $\hat{\mathfrak{g}}$. So we start by introducing affine Lie algebras and their enveloping algebras before proceeding to give their quantizations.

Let \mathfrak{g} be a semisimple finite-dimensional Lie algebra over \mathbb{C} of rank r with Cartan matrix $(a_{ij})_{i,j=1,\dots,r}$, symmetrizable via positive integers d_i , so that $d_i a_{ij}$ is symmetric. In terms of the simple roots α_i , we have

$$a_{ij} = 2 \frac{\alpha_i \cdot \alpha_j}{|\alpha_i|^2} \quad \text{and} \quad d_i = \frac{|\alpha_i|^2}{2}.$$

We can introduce an $\alpha_0 = \sum_{i=1}^r n_i \alpha_i$ in such a way that the extended Cartan matrix $(a_{ij})_{i,j=0,\dots,r}$ is of affine type – that is, it is positive semidefinite of rank r . The integers n_i are referred to as Kac indices. Choosing α_0 to be the highest root of \mathfrak{g} leads to an untwisted affine Kac–Moody algebra while choosing α_0 to be the highest short root of \mathfrak{g} leads to a twisted affine Kac–Moody algebra.

One defines the affine Lie algebra $\hat{\mathfrak{g}}$ corresponding to this affine Cartan matrix as the Lie algebra (over \mathbb{C}) with generators H_i, E_i^\pm for $i=0, 1, \dots, r$ and D with relations

$$\begin{aligned} [H_i, E_j^\pm] &= \pm a_{ij} E_j^\pm; & [H_i, H_j] &= 0 \\ [E_i^+, E_j^-] &= \delta_{ij} H_i & & \\ [D, H_i] &= 0, & [D, E_i^\pm] &= \pm \delta_{i,0} E_i^\pm \end{aligned} \tag{2}$$

$$\sum_{k=0}^{1-a_{ij}} (-1)^k \binom{1-a_{ij}}{k} (E_i^\pm)^k E_j^\pm (E_i^\pm)^{1-a_{ij}-k} = 0, \quad i \neq j$$

The E_i^\pm are referred to as Chevalley generators and the last set of relations are known as Serre relations. The generator D is known as the canonical derivation. We will denote the algebra obtained by dropping the generator D by $\hat{\mathfrak{g}}'$.

In applications to physics, the affine Lie algebra $\hat{\mathfrak{g}}$ often occurs in an isomorphic form as the loop Lie algebra $\mathfrak{g}[z, z^{-1}] \oplus \mathbb{C} \cdot c$ with Lie product (for untwisted $\hat{\mathfrak{g}}$)

$$\begin{aligned} [Xz^k, Yz^l] &= [X, Y]z^{k+l} + \delta_{k,-l}(X, Y)c, \\ &\text{for } X, Y \in \mathfrak{g}, \quad k, l \in \mathbb{Z} \end{aligned} \tag{3}$$

and c being the central element.

The universal enveloping algebra $U(\hat{\mathfrak{g}})$ of $\hat{\mathfrak{g}}$ is the unital algebra over \mathbb{C} with generators H_i, E_i^\pm for $i=0, 1, \dots, r$ and D and with relations given by [2] where now $[,]$ stands for the commutator instead of the Lie product.

To define the quantization of $U(\hat{\mathfrak{g}})$, one can either define $U_b(\hat{\mathfrak{g}})$ (Drinfeld 1985) as an algebra over the ring $\mathbb{C}[[\hbar]]$ of formal power series over an indeterminate \hbar or one can define $U_q(\hat{\mathfrak{g}})$ (Jimbo 1985) as an algebra over the field $\mathbb{Q}(q)$ of rational functions of q with coefficients in \mathbb{Q} . We will present $U_b(\hat{\mathfrak{g}})$ first.

The quantum affine algebra $U_b(\hat{\mathfrak{g}})$ is the unital algebra over $\mathbb{C}[[\hbar]]$ topologically generated by H_i, E_i^\pm for $i=0, 1, \dots, r$ and D with relations

$$\begin{aligned} [H_i, E_j^\pm] &= \pm a_{ij} E_j^\pm; & [H_i, H_j] &= 0 \\ [E_i^+, E_j^-] &= \delta_{ij} \frac{q_i^{H_i} - q_i^{-H_i}}{q_i - q_i^{-1}} & & \\ [D, H_i] &= 0, & [D, E_i^\pm] &= \pm \delta_{i,0} E_i^\pm \end{aligned} \tag{4}$$

$$\sum_{k=0}^{1-a_{ij}} (-1)^k \binom{1-a_{ij}}{k}_{q_i} (E_i^\pm)^k E_j^\pm (E_i^\pm)^{1-a_{ij}-k} = 0, \quad i \neq j$$

where $q_i = q^{d_i}$ and $q = e^{\hbar}$. The q -binomial coefficients are defined by

$$[n]_q = \frac{q^n - q^{-n}}{q - q^{-1}} \tag{5}$$

$$[n]_q! = [n]_q \cdot [n-1]_q \cdots [2]_q [1]_q \tag{6}$$

$$\begin{bmatrix} m \\ n \end{bmatrix}_q = \frac{[m]_q!}{[n]_q! [m-n]_q!} \tag{7}$$

The quantum affine algebra $U_b(\hat{\mathfrak{g}})$ is a Hopf algebra with coproduct

$$\begin{aligned} \Delta(D) &= D \otimes 1 + 1 \otimes D \\ \Delta(H_i) &= H_i \otimes 1 + 1 \otimes H_i \end{aligned} \tag{8}$$

$$\Delta(E_i^\pm) = E_i^\pm \otimes q_i^{-H_i/2} + q_i^{H_i/2} \otimes E_i^\pm$$

antipode

$$\begin{aligned} S(D) &= -D, & S(H_i) &= -H_i \\ S(E_i^\pm) &= -q_i^{\mp 1} E_i^\pm \end{aligned} \tag{9}$$

and co-unit

$$\epsilon(D) = \epsilon(H_i) = \epsilon(E_i^\pm) = 0 \tag{10}$$

It is easy to see that the classical enveloping algebra $U(\hat{\mathfrak{g}})$ can be obtained from the above by setting $\hbar = 0$, or more formally,

$$U_b(\hat{\mathfrak{g}})/\hbar U_b(\hat{\mathfrak{g}}) = U(\hat{\mathfrak{g}})$$

We can also define the quantum affine algebra $U_q(\hat{\mathfrak{g}})$ as the algebra over $\mathbb{Q}(q)$ with generators K_i, E_i^\pm, D for $i=0, 1, \dots, r$ and relations that are

obtained from the ones given above for $U_b(\hat{\mathfrak{g}})$ by setting

$$q_i^{H_i/2} = K_i, \quad i = 0, \dots, r \tag{11}$$

One can go further to an algebraic formulation over \mathbb{C} in which q is a complex number (with some points including $q=0$ not allowed). This has the advantage that it becomes possible to specialize, for example, to q a root of unity, where special phenomena occur.

Representations

For applications in physics, the finite-dimensional representations of $U_b(\hat{\mathfrak{g}}')$ are the most interesting. As will be explained in later sections, these occur, for example, as particle multiplets in 2D quantum field theory or as spin Hilbert spaces in quantum spin chains. In the next subsection, we will use them to derive matrix solutions to the Yang–Baxter equation.

While for a nonaffine quantum algebra $U_b(\mathfrak{g})$ the ring of representations is isomorphic to that of the classical enveloping algebra $U(\mathfrak{g})$ (because in fact the algebras are isomorphic, as Drinfeld has pointed out), the corresponding fact is no longer true for affine quantum groups, except in the case $\hat{\mathfrak{g}} = \mathfrak{a}_n^{(1)} = \widehat{\mathfrak{sl}}_{n+1}$.

For the classical enveloping algebras $U(\hat{\mathfrak{g}}')$, any finite-dimensional representation of $U(\mathfrak{g})$ also carries a finite-dimensional representation of $U(\hat{\mathfrak{g}}')$. In the quantum case, however, in general, an irreducible representation of $U_b(\hat{\mathfrak{g}}')$ reduces to a sum of representations of $U_b(\mathfrak{g})$.

To classify the finite-dimensional representations of $U_b(\hat{\mathfrak{g}}')$, it is necessary to use a different realization of $U_b(\hat{\mathfrak{g}}')$ that looks more like a quantization of the loop algebra realization [3] than the realization in terms of Chevalley generators. In terms of the generators in this alternative realization, which we do not give here because of its complexity, the finite-dimensional representations can be viewed as pseudo-highest-weight representations. There is a set of r “fundamental” representations V^a , $a = 1, \dots, r$, each containing the corresponding $U_b(\mathfrak{g})$ fundamental representation as a component, from the tensor products of which all the other finite-dimensional representations may be constructed. The details can be found in Chari and Pressley (1994).

Given some representation $\rho: U_b(\hat{\mathfrak{g}}') \rightarrow \text{End}(V)$, we can introduce a parameter λ with the help of the automorphism τ_λ of $U_b(\hat{\mathfrak{g}}')$ generated by D and given by

$$\begin{aligned} \tau_\lambda(E_i^\pm) &= \lambda^{\pm s_i} E_i^\pm \\ \tau_\lambda(H_i) &= H_i \end{aligned} \quad i = 0, \dots, r \tag{12}$$

Different choices for the s_i correspond to different gradations. Commonly used are the “homogeneous

gradation,” $s_0 = 1, s_1 = \dots = s_r = 0$, and the “principal gradation,” $s_0 = s_1 = \dots = s_r = 1$. We shall also need the “spin gradation” $s_i = d_i^{-1}$. The representations

$$\rho_\lambda = \rho \circ \tau_\lambda$$

play an important role in applications to integrable models where λ is referred to as the (multiplicative) spectral parameter. In applications to particle scattering introduced in a later section, it is related to the rapidity of the particle. The generator D can be realized as an infinitesimal scaling operator on λ and thus plays the role of the Lorentz boost generator.

The tensor product representations $\rho_\lambda^a \otimes \rho_\mu^b$ are irreducible generically but become reducible for certain values of λ/μ , a fact which again is important in applications (fusion procedure, particle-bound states).

R-Matrices

A Hopf algebra A is said to be “almost cocommutative” if there exists an invertible element $\mathcal{R} \in A \otimes A$ such that

$$\mathcal{R}\Delta(x) = (\sigma \circ \Delta(x))\mathcal{R}, \quad \text{for all } x \in A \tag{13}$$

where $\sigma: x \otimes y \mapsto y \otimes x$ exchanges the two factors in the coproduct. In a quasitriangular Hopf algebra, this element \mathcal{R} satisfies

$$\begin{aligned} (\Delta \otimes \text{id})(\mathcal{R}) &= \mathcal{R}_{13}\mathcal{R}_{23} \\ (\text{id} \otimes \Delta)(\mathcal{R}) &= \mathcal{R}_{13}\mathcal{R}_{12} \end{aligned} \tag{14}$$

and is known as the “universal R -matrix” (see Hopf Algebras and q -Deformation Quantum Groups). As a consequence of [13] and [14], it automatically satisfies the Yang–Baxter equation

$$\mathcal{R}_{12}\mathcal{R}_{13}\mathcal{R}_{23} = \mathcal{R}_{23}\mathcal{R}_{13}\mathcal{R}_{12} \tag{15}$$

For technical reasons, to do with the infinite number of root vectors of $\hat{\mathfrak{g}}$, the quantum affine algebra $U_b(\hat{\mathfrak{g}})$ does not possess a universal R -matrix that is an element of $U_b(\hat{\mathfrak{g}}) \otimes U_b(\hat{\mathfrak{g}})$. However, as pointed out by Drinfeld (1985), it possesses a pseudouniversal R -matrix $\mathcal{R}(\lambda) \in (U_b(\hat{\mathfrak{g}}') \otimes U_b(\hat{\mathfrak{g}}'))((\lambda))$. The λ is related to the automorphism τ_λ defined in [12]. When using the homogeneous gradation, $\mathcal{R}(\lambda)$ is a formal power series in λ .

When the pseudouniversal R -matrix is evaluated in the tensor product of any two indecomposable finite-dimensional representations ρ_1 and ρ_2 , one obtains a numerical R -matrix

$$R^{12}(\lambda) = (\rho^1 \otimes \rho^2)\mathcal{R}(\lambda) \tag{16}$$

The entries of these numerical R -matrices are rational functions of the multiplicative spectral parameter λ but when written in terms of the additive spectral parameter $u = \log(\lambda)$ they are trigonometric functions of u and satisfy the Yang–Baxter equation in the form given in [1]. The matrix

$$\check{R}^{12}(\lambda) = \sigma \circ R^{12}(\lambda)$$

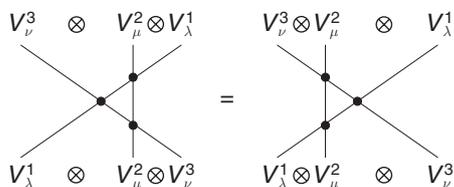
satisfies the intertwining relation

$$\begin{aligned} \check{R}^{12}(\lambda/\mu) \cdot (\rho_\lambda^1 \otimes \rho_\mu^2)(\Delta(x)) \\ = (\rho_\mu^2 \otimes \rho_\lambda^1)(\Delta(x)) \cdot \check{R}^{12}(\lambda/\mu) \end{aligned} \quad [17]$$

for any $x \in U_b(\hat{\mathfrak{g}}')$. It follows from the irreducibility of the tensor product representations that these R -matrices satisfy the Yang–Baxter equations

$$\begin{aligned} (\text{id} \otimes \check{R}^{23}(\mu/\nu))(\check{R}^{13}(\lambda/\nu) \otimes \text{id})(\text{id} \otimes \check{R}^{12}(\lambda/\mu)) \\ = (\check{R}^{12}(\lambda/\mu) \otimes \text{id})(\text{id} \otimes \check{R}^{13}(\lambda/\nu)) \\ \times (\check{R}^{23}(\mu/\nu) \otimes \text{id}) \end{aligned} \quad [18]$$

or, graphically,



Explicit formulas for the pseudouniversal R -matrices were found by Khoroshkin and Tolstoy. However, these are difficult to evaluate explicitly in specific representations so that in practice it is easiest to find the numerical R -matrices $\check{R}^{ab}(\lambda)$ by solving the intertwining relation [17]. It should be stressed that solving the intertwining relation, which is a linear equation for the R -matrix, is much easier than directly solving the Yang–Baxter equation, a cubic equation.

Yangians

As remarked by Drinfeld (1986), for untwisted $\hat{\mathfrak{g}}$ the quantum affine algebra $U_b(\hat{\mathfrak{g}}')$ degenerates as $b \rightarrow 0$ into another quasipseudotriangular Hopf algebra, the “Yangian” $Y(\mathfrak{g})$ (Drinfeld 1985). It is associated with R -matrices which are rational functions of the additive spectral parameter u . Its representation ring coincides with that of $U_b(\hat{\mathfrak{g}}')$.

Consider a general presentation of a Lie algebra \mathfrak{g} , with generators I_a and structure constants f_{abc} , so that

$$[I_a, I_b] = f_{abc}I_c, \quad \Delta(I_a) = I_a \otimes 1 + 1 \otimes I_a$$

(with summation over repeated indices). The Yangian $Y(\mathfrak{g})$ is the algebra generated by these and a second set of generators J_a satisfying

$$\begin{aligned} [I_a, J_b] &= f_{abc}J_c \\ \Delta(J_a) &= J_a \otimes 1 + 1 \otimes J_a + \frac{1}{2}f_{abc}I_c \otimes I_b \end{aligned}$$

The requirement that Δ be a homomorphism imposes further relations:

$$[J_a, [J_b, I_c]] - [I_a, [J_b, J_c]] = \alpha_{abcdeg}\{I_d, I_e, I_g\}$$

and

$$\begin{aligned} [[J_a, J_b], [I_l, J_m]] + [[J_l, J_m], [I_a, J_b]] \\ = (\alpha_{abcdeg}f_{lmc} + \alpha_{lmcdg}f_{abc})\{I_d, I_e, J_g\} \end{aligned}$$

where

$$\alpha_{abcdeg} = \frac{1}{24}f_{adif}f_{bejf}f_{cgkf}f_{ijk}, \quad \{x_1, x_2, x_3\} = \sum_{i \neq j \neq k} x_i x_j x_k$$

When $\mathfrak{g} = \mathfrak{sl}_2$ the first of these is trivial, while for $\mathfrak{g} \neq \mathfrak{sl}_2$ the first implies the second. The co-unit is $\epsilon(I_a) = \epsilon(J_a) = 0$; the antipode is $s(I_a) = -I_a$, $s(J_a) = -J_a + (1/2)f_{abc}I_cI_b$. The Yangian may be obtained from $U_b(\hat{\mathfrak{g}}')$ by expanding in powers of \hbar . For the precise relationship, see Drinfeld (1985) and MacKay (2005). In the spin gradation, the automorphism [12] generated by D descends to $Y(\mathfrak{g})$ as $I_a \mapsto I_a$, $J_a \mapsto J_a + uI_a$.

There are two other realizations of $Y(\mathfrak{g})$. The first (see, for example, Molev 2003) defines $Y(\mathfrak{gl}_n)$ directly from

$$R(u-v)T_1(u)T_2(v) = T_2(v)T_1(u)R(u-v)$$

where $T_1(u) = T(u) \otimes \text{id}$, $T_2(v) = \text{id} \otimes T(v)$, and

$$\begin{aligned} T(u) &= \sum_{i,j=1}^n t_{ij}(u) \otimes e_{ij} \\ t_{ij}(u) &= \delta_{ij} + I_{ij}u^{-1} + J_{ij}u^{-2} + \dots \end{aligned}$$

where e_{ij} are the standard matrix units for \mathfrak{gl}_n . The rational R -matrix for the n -dimensional representation of \mathfrak{gl}_n is

$$R(u-v) = 1 - \frac{P}{u-v}, \quad \text{where } P = \sum_{i,j=1}^n e_{ij} \otimes e_{ji}$$

is the transposition operator. $Y(\mathfrak{gl}_n)$ is then defined to be the algebra generated by I_{ij}, J_{ij} , and must be quotiented by the “quantum determinant” at its center to define $Y(\mathfrak{sl}_n)$. The coproduct takes a particularly simple form,

$$\Delta(t_{ij}(u)) = \sum_{k=1}^n t_{ik}(u) \otimes t_{kj}(u)$$

Here we do not give explicitly the third realization, namely Drinfeld's "new" realization of $Y(\mathfrak{g})$ (Drinfeld 1988), but we remark that it was in this presentation that Drinfeld found a correspondence between certain sets of polynomials and finite-dimensional irreducible representations of $Y(\mathfrak{g})$, thus classifying these (although not thereby deducing their dimension or constructing the action of $Y(\mathfrak{g})$). As remarked earlier, the structure is as in the earlier section: $Y(\mathfrak{g})$ representations are in general \mathfrak{g} -reducible, and there is a set of r fundamental $Y(\mathfrak{g})$ -representations, containing the fundamental \mathfrak{g} -representations as components, from which all other representations can be constructed.

Origins in the Quantum Inverse-Scattering Method

Quantum affine algebras for general $\hat{\mathfrak{g}}$ first appear in Drinfeld (1985, 1986) and Jimbo (1985, 1986), but they have their origin in the "quantum inverse-scattering method" (QISM) of the St. Petersburg school, and the essential features of $U_b(\widehat{\mathfrak{sl}}_2)$ first appear in Kulish and Reshetikhin (1983). In this section, we explain how the quantization of the Lax-pair description of affine Toda theory led to the discovery of the $U_b(\hat{\mathfrak{g}})$ coproduct, commutation relations, and R -matrix. We use the normalizations of Jimbo (1986), in which the H_i are rescaled so that the Cartan matrix $a_{ij} = \alpha_i \cdot \alpha_j$ is symmetric.

We begin with the affine Toda field equations

$$\partial^\mu \partial_\mu \phi_i = -\frac{m^2}{\beta} \sum_{j=1}^r (e^{\beta a_{ij} \phi_j} - n_j e^{\beta \alpha_0 \cdot \alpha_j \phi_j})$$

an integrable model in \mathbb{R}^{1+1} of r real scalar fields $\phi_i(x, t)$ with a mass parameter m and coupling constant β . Equivalently, we may write $[\partial_x + L_x, \partial_t + L_t] = 0$ for the Lax pair

$$\begin{aligned} L_x(x, t) &= \frac{\beta}{2} \sum_{i=1}^r H_i \partial_t \phi_i + \frac{m}{2} \sum_{i,j=1}^r e^{(\beta/2) a_{ij} \phi_j} (E_i^+ + E_i^-) \\ &\quad + \frac{m}{2} \sum_{j=1}^r e^{(\beta/2) a_{0j} \phi_j} \left(\lambda E_0^+ + \frac{1}{\lambda} E_0^- \right) \\ L_t(x, t) &= \frac{\beta}{2} \sum_{i=1}^r H_i \partial_x \phi_i + \frac{m}{2} \sum_{i,j=1}^r e^{(\beta/2) a_{ij} \phi_j} (E_i^+ - E_i^-) \\ &\quad + \frac{m}{2} \sum_{j=1}^r e^{(\beta/2) a_{0j} \phi_j} \left(\lambda E_0^+ - \frac{1}{\lambda} E_0^- \right) \end{aligned}$$

with arbitrary $\lambda \in \mathbb{C}$. The classical integrability of the system is seen in the existence of $r(\lambda, \lambda')$ such that

$$\{T(\lambda) \otimes T(\lambda')\} = [r(\lambda, \lambda'), T(\lambda) \otimes T(\lambda')]$$

where $T(\lambda) = T(-\infty, \infty; \lambda)$ and $T(x, y; \lambda) = P \exp(\int_x^y L(\xi; \lambda) d\xi)$. Taking the trace of this relation gives an infinity of charges in involution.

Quantization is problematic, owing to divergences in T . The QISM regularizes these by putting the model on a lattice of spacing Δ , defining the lattice Lax operator to be

$$\begin{aligned} L_n(\lambda) &= T((n-1/2)\Delta, (n+1/2)\Delta; \lambda) \\ &= P \exp \left(\int_{(n-(1/2)\Delta}^{(n+(1/2)\Delta)} L(\xi; \lambda) d\xi \right) \end{aligned}$$

The lattice monodromy matrix is then $T(\lambda) = \lim_{l \rightarrow -\infty, m \rightarrow \infty} T_l^m$ where $T_l^m = L_m L_{m-1} \cdots L_{l+1}$, and its trace again yields an infinity of commuting charges, provided that there exists a quantum R -matrix $R(\lambda_1, \lambda_2)$ such that

$$\begin{aligned} R(\lambda_1, \lambda_2) L_n^1(\lambda_1) L_n^2(\lambda_2) \\ = L_n^2(\lambda_2) L_n^1(\lambda_1) R(\lambda_1, \lambda_2) \end{aligned} \quad [19]$$

where $L_n^1(\lambda_1) = L_n(\lambda_1) \otimes \text{id}$, $L_n^2(\lambda_2) = \text{id} \otimes L_n(\lambda_2)$. That R solves the Yang-Baxter equation follows from the equivalence of the two ways of intertwining $L_n(\lambda_1) \otimes L_n(\lambda_2) \otimes L_n(\lambda_3)$ with $L_n(\lambda_3) \otimes L_n(\lambda_2) \otimes L_n(\lambda_1)$.

To compute $L_n(\lambda)$, one uses the canonical, equal-time commutation relations for the ϕ_i and $\dot{\phi}_i$. In terms of the lattice fields

$$\begin{aligned} p_{i,n} &= \int_{(n-(1/2)\Delta}^{(n+(1/2)\Delta)} \dot{\phi}_i(x) dx \\ q_{i,n} &= \int_{(n-(1/2)\Delta}^{(n+(1/2)\Delta)} \sum_j e^{(\beta/2) a_{ij} \phi_j(x)} dx \end{aligned}$$

the only nontrivial relation is $[p_{i,n}, q_{j,n}] = (i\hbar\beta/2)\delta_{ij}q_{j,n}$, and one finds

$$\begin{aligned} L_n(\lambda) &= \exp \left(\frac{\beta}{2} \sum_i H_i p_{i,n} \right) + \exp \left(\frac{\beta}{4} \sum_j H_j p_{j,n} \right) \\ &\quad \times \frac{m}{2} \left[\sum_i q_{i,n} (E_i^+ + E_i^-) \right. \\ &\quad \left. + \prod_i q_{i,n}^{-n_i} \left(\lambda E_0^+ + \frac{1}{\lambda} E_0^- \right) \right] \\ &\quad \times \exp \left(\frac{\beta}{4} \sum_j H_j p_{j,n} \right) + O(\Delta^2) \end{aligned}$$

the expression used by the St Petersburg school and by Jimbo. We now make the replacement $E_i^\pm \mapsto q^{-H_i/4} E_i^\pm q^{H_i/4}$, where $q = \exp(i\hbar\beta^2/2)$, and compute the $O(\Delta)$ terms in [19], which reduce to

$$\begin{aligned}
 &R(z)(H_i \otimes 1 + 1 \otimes H_i) \\
 &= (H_i \otimes 1 + 1 \otimes H_i)R(z) \\
 &R(z)\left(E_i^\pm \otimes q^{-H_i/2} + q^{H_i/2} \otimes E_i^\pm\right) \\
 &= \left(q^{-H_i/2} \otimes E_i^\pm + E_i^\pm \otimes q^{H_i/2}\right)R(z) \\
 &R(z)\left(z^{\pm 1}E_0^\pm \otimes q^{-H_0/2} + q^{H_0/2} \otimes E_0^\pm\right) \\
 &= \left(q^{-H_0/2} \otimes E_0^\pm + z^{\pm 1}E_0^\pm \otimes q^{H_0/2}\right)R(z)
 \end{aligned}$$

where $z = \lambda_1/\lambda_2$. We recognize in these the $U_b(\hat{\mathfrak{g}})$ coproduct and thus the intertwining relations, in the homogeneous gradation. These equations were solved for R in defining representations of nonexceptional \mathfrak{g} by Jimbo (1986).

For $\hat{\mathfrak{g}} = \widehat{\mathfrak{sl}}_2$, it was Kulish and Reshetikhin (1983) who first discovered that the requirement that the coproduct must be an algebra homomorphism forces the replacement of the commutation relations of $U(\widehat{\mathfrak{sl}}_2)$ by those of $U_b(\widehat{\mathfrak{sl}}_2)$; more generally it requires the replacement of $U(\hat{\mathfrak{g}})$ by $U_b(\hat{\mathfrak{g}})$.

Affine Quantum Group Symmetry and the Exact S-Matrix

In the last section, we saw the origins of $U_b(\hat{\mathfrak{g}})$ in the ‘‘auxiliary’’ algebra introduced in the Lax pair. However, the quantum affine algebras also play a second role, as a symmetry algebra. An imaginary-coupled affine Toda field theory based on the affine algebra $\hat{\mathfrak{g}}^\vee$ possesses the quantum affine algebra $U_b(\hat{\mathfrak{g}})$ as a symmetry algebra, where $\hat{\mathfrak{g}}^\vee$ is the Langland dual to $\hat{\mathfrak{g}}$ (the algebra obtained by replacing roots by coroots).

The solitonic particle states in affine Toda theories form multiplets which transform in the fundamental representations of the quantum affine algebra. Multi-particle states transform in tensor product representations $V^a \otimes V^b$. The scattering of two solitons of type a and b with relative rapidity θ is described by the S -matrix $S^{ab}(\theta): V^a \otimes V^b \rightarrow V^b \otimes V^a$, graphically represented in Figure 1a. It then follows from the symmetry that the two-particle scattering matrix

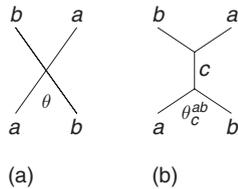


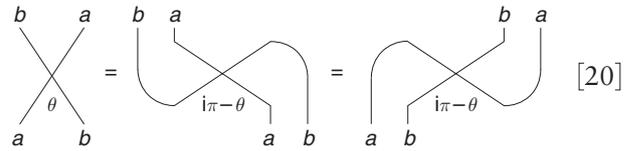
Figure 1 (a) Graphical representation of a two-particle scattering process described by the S -matrix $S_{ab}(\theta)$. (b) At special values θ_c^{ab} of the relative spectral parameter, the two particles of types a and b form a bound state of type c .

(S -matrix) for solitons must be proportional to the intertwiner for these tensor product representations, the R matrix:

$$S^{ab}(\theta) = f^{ab}(\theta)\check{R}^{ab}(\theta)$$

with θ proportional to u , the additive spectral parameter. The scalar prefactor $f^{ab}(\theta)$ is not determined by the symmetry but is fixed by other requirements like unitarity, crossing symmetry, and the bootstrap principle.

It turns out that the axiomatic properties of the R -matrices are in perfect agreement with the axiomatic properties of the analytic S -matrix. For example, crossing symmetry of the S -matrix, graphically represented by



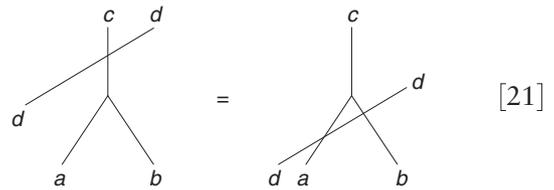
is a consequence of the property of the universal R -matrix with respect to the action of the antipode S ,

$$(S \otimes 1)\mathcal{R} = \mathcal{R}^{-1}$$

An S -matrix will have poles at certain imaginary rapidities θ_c^{ab} corresponding to the formation of virtual bound states. This is graphically represented in Figure 1b. The location of the pole is determined by the masses of the three particles involved,

$$m_c^2 = m_a^2 + m_b^2 + 2m_a m_b \cos(i\theta_c^{ab})$$

At the bound state pole, the S -matrix will project onto the multiplet V^c . Thus, the \check{R} matrix has to have this projection property as well and indeed, this turns out to be the case. The bootstrap principle, whereby the S -matrix for a bound state is obtained from the S -matrices of the constituent particles,



is a consequence of the property [14] of the universal R -matrix with respect to the coproduct.

There is a famous no-go theorem due to Coleman and Mandula which states the ‘‘impossibility of combining space-time and internal symmetries in any but a trivial way.’’ Affine quantum group symmetry circumvents this no-go theorem. In fact, the derivation D is the infinitesimal two-dimensional Lorentz boost generator and the other symmetry

charges transform nontrivially under these Lorentz transformations, see [2].

The noncocommutative coproduct [8] means that a $U_b(\hat{\mathfrak{g}})$ symmetry generator, when acting on a 2-soliton state, acts differently on the left soliton than on the right soliton. This is only possible because the generator is a nonlocal symmetry charge – that is, a charge which is obtained as the space integral of the time component of a current which itself is a nonlocal expression in terms of the fields of the theory.

Similarly, many nonlinear sigma models possess nonlocal charges which form $Y(\mathfrak{g})$, and the construction proceeds similarly, now utilizing rational R -matrices, and with particle multiplets forming fundamental representations of $Y(\mathfrak{g})$. In each case, the three-point couplings corresponding to the formation of bound states, and thus the analogs for $U_b(\hat{\mathfrak{g}})$ and $Y(\mathfrak{g})$ of the Clebsch–Gordan couplings, obey a rather beautiful geometric rule originally deduced in simpler, purely elastic scattering models (Chari and Pressley 1996).

More details about this topic can be found in Delius (1995) and MacKay (2005).

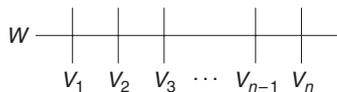
Integrable Quantum Spin Chains

Affine quantum groups provide an unlimited supply of integrable quantum spin chains. From any R -matrix $R(\theta)$ for any tensor product of finite-dimensional representations $W \otimes V$, one can produce an integrable quantum system on the Hilbert space $V^{\otimes n}$. This Hilbert space can then be interpreted as the space of n interacting spins. The space W is an auxiliary space required in the construction but not playing a role in the physics.

Given an arbitrary R -matrix $R(\theta)$, one defines the monodromy matrix $T(\theta) \in \text{End}(W \otimes V^{\otimes n})$ by

$$T(\theta) = R_{01}(\theta - \theta_1)R_{02}(\theta - \theta_2) \cdots R_{0n}(\theta - \theta_n)$$

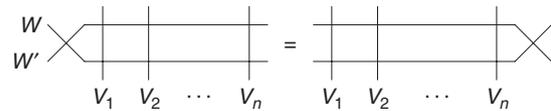
where, as usual, R_{ij} is the R -matrix acting on the i th and j th component of the tensor product space. The θ_i can be chosen arbitrarily for convenience. Graphically the monodromy matrix can be represented as



As a consequence of the Yang–Baxter equation satisfied by the R -matrices the monodromy matrix satisfies

$$RTT = TTR \tag{22}$$

or, graphically,



One defines the transfer matrix

$$\tau(\theta) = \text{tr}_W T(\theta)$$

which is now an operator on $V^{\otimes n}$, the Hilbert space of the quantum spin chain. Due to [22], two transfer matrices commute,

$$[\tau(\theta), \tau(\theta')] = 0$$

and thus the $\tau(\theta)$ can be seen as a generating function of an infinite number of commuting charges, one of which will be chosen as the Hamiltonian. This Hamiltonian can then be diagonalized using the algebraic Bethe ansatz.

One is usually interested in the thermodynamic limit where the number of spins goes to infinity. In this limit, it has been conjectured, the Hilbert space of the spin chain carries a certain infinite-dimensional representation of the quantum affine algebra and this has been used to solve the model algebraically, using vertex operators (Jimbo and Miwa 1995).

Boundary Quantum Groups

In applications to physical systems that have a boundary, the Yang–Baxter equation [1] appears in conjunction with the boundary Yang–Baxter equation, also known as the reflection equation,

$$R_{12}(u - v)K_1(u)R_{21}(u + v)K_2(v) = K_2(v)R_{12}(u + v)K_1(u)R_{21}(u - v) \tag{23}$$

The matrices K are known as reflection matrices. This equation was originally introduced by Cherednik to describe the reflection of particles from a boundary in an integrable scattering theory and was used by Sklyanin to construct integrable spin chains and quantum field theories with boundaries.

Boundary quantum groups are certain co-ideal subalgebras of affine quantum groups. They provide the algebraic structures underlying the solutions of the boundary Yang–Baxter equation in the same way in which affine quantum groups underlie the solutions of the ordinary Yang–Baxter equation. Both allow one to find solutions of the respective Yang–Baxter equation by solving a linear intertwining relation. In the case without spectral parameters these algebras appear in the theory of braided groups (see Hopf Algebras and q -Deformation Quantum Groups and Braided and Modular Tensor Categories).

For example, the subalgebra $B_\epsilon(\hat{\mathfrak{g}})$ of $U_b(\hat{\mathfrak{g}}')$ generated by

$$Q_i = q_i^{H_i/2}(E_i^+ + E_i^-) + \epsilon_i(q_i^{H_i} - 1),$$

$$i = 0, \dots, r \tag{24}$$

is a boundary quantum group for certain choices of the parameters $\epsilon_i \in \mathbb{C}[[\hbar]]$. It is a left co-ideal subalgebra of $U_b(\hat{\mathfrak{g}}')$ because

$$\Delta(Q_i) = Q_i \otimes 1 + q_i^{H_i} \otimes Q_i \in U_b(\hat{\mathfrak{g}}') \otimes B_\epsilon(\hat{\mathfrak{g}}) \tag{25}$$

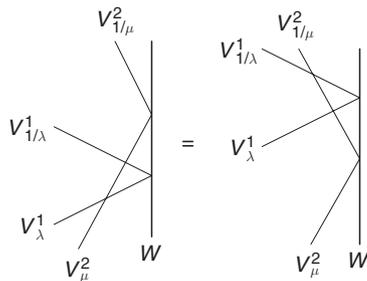
Intertwiners $K(\lambda) : V_{\eta\lambda} \rightarrow V_{\eta/\lambda}$ for some constant η satisfying

$$K(\lambda)\rho_{\eta\lambda}(Q) = \rho_{\eta/\lambda}(Q)K(\lambda), \text{ for all } Q \in B_\epsilon(\hat{\mathfrak{g}}) \tag{26}$$

provide solutions of the reflection equation in the form

$$\begin{aligned} &(\text{id} \otimes K^2(\mu))\check{R}^{12}(\lambda\mu)(\text{id} \otimes K^1(\lambda))\check{R}^{21}(\lambda/\mu) \\ &= \check{R}^{12}(\lambda/\mu)(\text{id} \otimes K^1(\lambda)) \\ &\times \check{R}^{21}(\lambda\mu)(\text{id} \otimes K^2(\mu)) \end{aligned} \tag{27}$$

This can be extended to the case where the boundary itself carries a representation W of $B_\epsilon(\hat{\mathfrak{g}})$. The boundary Yang–Baxter equation can be represented graphically as



Another example is provided by twisted Yangians where, when the I_a and J_a are constructed as nonlocal charges in sigma models, it is found that a boundary condition which preserves integrability leaves only the subset

$$I_i \quad \text{and} \quad \tilde{J}_p = J_p + \frac{1}{4}f_{piq}(I_i I_q + I_q I_i)$$

conserved, where i labels the \mathfrak{h} -indices and p, q the \mathfrak{k} -indices of a symmetric splitting $\mathfrak{g} = \mathfrak{h} + \mathfrak{k}$. The

algebra $Y(\mathfrak{g}, \mathfrak{h})$ generated by the I_i, \tilde{J}_p is, like $B_\epsilon(\hat{\mathfrak{g}})$, a co-ideal subalgebra, $\Delta(Y(\mathfrak{g}, \mathfrak{h})) \subset Y(\mathfrak{g}) \otimes Y(\mathfrak{g}, \mathfrak{h})$, and again yields an intertwining relation for K -matrices. For $\mathfrak{g} = \mathfrak{sl}_n$ and $\mathfrak{h} = \mathfrak{so}_n$ or \mathfrak{sp}_{2n} , $Y(\mathfrak{g}, \mathfrak{h})$ is the “twisted Yangian” described in Molev (2003).

All the constructions in earlier sections of this review have analogs in the boundary setting. For more details see Delius and MacKay (2003) and MacKay (2005).

See also: Bethe Ansatz; Boundary Conformal Field Theory; Classical r -Matrices, Lie Bialgebras, and Poisson Lie Groups; Hopf Algebras and q -Deformation Quantum Groups; Riemann–Hilbert Problem; Solitons and Kac–Moody Lie Algebras; Yang–Baxter Equations.

Further Reading

Chari V and Pressley AN (1994) *Quantum Groups*. Cambridge: Cambridge University Press.

Chari V and Pressley AN (1996) Yangians, integrable quantum systems and Dorey’s rule. *Communications in Mathematical Physics* 181: 265–302.

Delius GW (1995) Exact S-matrices with affine quantum group symmetry. *Nuclear Physics B* 451: 445–465.

Delius GW and MacKay NJ (2003) Quantum group symmetry in sine-Gordon and affine Toda field Theories on the Half-Line. *Communications in Mathematical Physics* 233: 173–190.

Drinfeld V (1985) Hopf algebras and the quantum Yang–Baxter equation. *Soviet Mathematics Doklady* 32: 254–258.

Drinfeld V (1986) *Quantum Groups*, Proc. Int. Cong. Math. (Berkeley), pp. 798–820.

Drinfeld V (1988) A new realization of Yangians and quantized affine algebras. *Soviet Mathematics Doklady* 36: 212–216.

Jimbo M (1985) A q -difference analogue of $U(\mathfrak{g})$ and the Yang–Baxter equation. *Letters in Mathematical Physics* 10: 63–69.

Jimbo M (1986) Quantum R-matrix for the generalized Toda system. *Communications in Mathematical Physics* 102: 537–547.

Jimbo M and Miwa T (1995) *Algebraic Analysis of Solvable Lattice Models*. Providence, RI: American Mathematical Society.

Kulish PP and Reshetikhin NY (1983) Quantum linear problem for the sine-Gordon equation and higher representations. *Journal of Soviet Mathematics* 23: 2435.

MacKay NJ (2005) Introduction to Yangian symmetry in integrable field theory. *International Journal of Modern Physics* (to appear).

Molev A (2003) Yangians and their applications. In: Hazewinkel M (ed.) *Handbook of Algebra*, vol. 3, pp. 907–959. Elsevier.

Aharonov–Bohm Effect

M Socolovsky, Universidad Nacional Autónoma de México, México DF, México

© 2006 Elsevier Ltd. All rights reserved.

Introduction

In classical electrodynamics, the interaction of charged particles with the electromagnetic field is local, through the pointlike coupling of the electric charge of the particles with the electric and magnetic fields, \mathbf{E} and \mathbf{B} , respectively. This is mathematically expressed by the Lorentz-force law. The scalar and vector potentials, φ and \mathbf{A} , which are the time and space components of the relativistic 4-potential A_μ , are considered auxiliary quantities in terms of which the field strengths \mathbf{E} and \mathbf{B} , the observables, are expressed in a gauge-invariant manner. The homogeneous or first pair of Maxwell equations are a direct consequence of the definition of the field strengths in terms of A_μ . The inhomogeneous or second pair of Maxwell equations, which involve the charges and currents present in the problem, are also usually written in terms of \mathbf{E} and \mathbf{B} ; however when writing them in terms of A_μ , the number of degrees of freedom of the electromagnetic field is explicitly reduced from six to four; and finally, with two additional gauge transformations, one ends with the two physical degrees of freedom of the electromagnetic field.

In quantum mechanics, however, both the Schrödinger equation and the path-integral approaches for scalar and unpolarized charged particles in the presence of electromagnetic fields, are written in terms of the potential and not of the field strengths. Even in the case of the Schrödinger–Pauli equation for spin 1/2 electrons with magnetic moment $\boldsymbol{\mu}$ interacting with a magnetic field \mathbf{B} , one knows that the coupling $-\boldsymbol{\mu} \cdot \mathbf{B}$ is the nonrelativistic limit of the Dirac equation, which depends on A_μ but not on \mathbf{E} and \mathbf{B} . Since gauge invariance also holds in the quantum domain, it was thought that \mathbf{A} and φ were mere auxiliary quantities, like in the classical case.

Aharonov and Bohm, in 1959, predicted a quantum interference effect due to the motion of charged particles in regions where $\mathbf{B}(\mathbf{E})$ vanishes, but not $\mathbf{A}(\varphi)$, leading to a nonlocal gauge-invariant effect depending on the flux of the magnetic field in the inaccessible region, in the magnetic case, and on the difference of the integrals over time of time-varying potentials, in the electric case. (The magnetic effect was already noticed 10 years before by Ehrenberg and Siday in a paper on the refractive index of electrons.)

In the context of the Schrödinger equation, one can show that due to gauge invariance, if ψ_0 is a solution to the equation in the absence of an electromagnetic potential, then the product of $\psi_0(\mathbf{x})$ times the integral of A_μ over a path joining an arbitrary reference point \mathbf{x}_0 to \mathbf{x} is also a solution, if the integral is path independent. However, it is the path integral of Feynman which in the formulas for propagators of charged particles in the presence of electromagnetic fields clearly shows that *the action of these fields on charged particles is nonlocal, and it is given by the celebrated non-integrable (path-dependent) phase factor of Wu and Yang (1975)*. Moreover, this fact provides an additional proof of the nonlocal character of quantum mechanics: to surround fluxes, or to develop a potential difference, the particle has to travel simultaneously at least through two paths.

Thus, the fact that the Aharonov–Bohm (A–B) effect was verified experimentally, by Chambers and others, demonstrates the necessity of introducing the (gauge-dependent) potential A_μ in describing the electromagnetic interactions of the quantum particle. This is widely regarded as the single most important piece of evidence for electromagnetism being a gauge theory. Moreover, it shows, to paraphrase Yang, that the field underdescribes the physical theory, while the potential overdescribes it, and it is the phase factor which describes it exactly.

The content of this article is essentially twofold. The first four sections are mainly physical, where we describe the magnetic A–B effect using the Schrödinger equation and the Feynman path integral. The fifth section is geometrical and is the longest of the article. We describe the effect in the context of fiber bundles and connections, namely as a result of the coupling of the wave function (section of an associated bundle) to a nontrivial flat connection (non-pure gauge vector potential with zero magnetic field) in a trivial bundle (the A–B bundle) with topologically nontrivial (non-simply-connected) base space. We discuss the moduli space of flat connections and the holonomy groups giving the phase shifts of the interference patterns. Finally, in the last section, we briefly comment on the nonabelian A–B effect.

Electromagnetic Fields in Classical Physics

In classical physics, the motion of charged particles in the presence of electromagnetic fields is governed by the equation

$$\frac{d}{dt}\mathbf{p} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \quad [1]$$

where

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - (\mathbf{v}^2/c^2)}}$$

is the mechanical momentum of the particle with electric charge q , mass m , and velocity $\mathbf{v} = \dot{\mathbf{x}}$ (c is the velocity of light in vacuum, and for $|\mathbf{v}| \ll c$ the left-hand side (LHS) of [1] is approximately $m\mathbf{v}$); the right-hand side (RHS) is the Lorentz force, where \mathbf{E} and \mathbf{B} are, respectively, the electric and magnetic fields at the spacetime point (t, \mathbf{x}) where the particle is located. Equation [1] is easily derived from the Euler–Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \mathbf{v}}\right) - \frac{\partial L}{\partial \mathbf{x}} = 0 \quad [2]$$

with the Lagrangian L given by the sum of the free Lagrangian for the particle,

$$L_0 = -mc^2\sqrt{1 - \frac{\mathbf{v}^2}{c^2}} \quad [3]$$

and the Lagrangian describing the particle–field interaction,

$$L_{\text{int}} = \frac{q}{c}\mathbf{A} \cdot \mathbf{v} - q\varphi \quad [4]$$

In [4], \mathbf{A} and φ are, respectively, the vector potential and the scalar potential, which together form the 4-potential $A_\mu = (A_0, -\mathbf{A}) = (\varphi, -A^i)$, $i = 1, 2, 3$, in terms of which the electric and magnetic field strengths are given by

$$\mathbf{E} = -\frac{1}{c}\frac{\partial}{\partial t}\mathbf{A} - \nabla\varphi \quad [5a]$$

$$\mathbf{B} = \nabla \times \mathbf{A} \quad [5b]$$

The classical action corresponding to a given path of the particle is

$$\begin{aligned} S &= \int_{t_1}^{t_2} dt L = \int_{t_1}^{t_2} dt (L_0 + L_{\text{int}}) \\ &= \int_{t_1}^{t_2} dt L_0 + \int_{t_1}^{t_2} dt L_{\text{int}} \equiv S_0 + S_{\text{int}} \end{aligned} \quad [6]$$

\mathbf{E} , \mathbf{B} , and S are invariant under the gauge transformation

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} - \nabla\Lambda \quad [7a]$$

$$\varphi \rightarrow \varphi' = \varphi + \frac{1}{c}\frac{\partial}{\partial t}\Lambda \quad [7b]$$

where Λ is a real-valued differentiable scalar function (at least of class C^2) on spacetime. That is, if \mathbf{E}' , \mathbf{B}' , and S'_{int} are defined in terms of \mathbf{A}' and φ' as \mathbf{E} , \mathbf{B} , and S_{int} are defined in terms of \mathbf{A} and φ , then $\mathbf{E}' = \mathbf{E}$, $\mathbf{B}' = \mathbf{B}$, and $S'_{\text{int}} = S_{\text{int}}$. This fact leads to the concept that, classically, the observables \mathbf{E} and \mathbf{B} are the physical quantities, while A_μ is only an auxiliary quantity. Also, and most important in the present context, eqn [1] states that the motion of the particles is determined by the values or state of the field strengths in an infinitesimal neighborhood of the particles, that is, classically, \mathbf{E} and \mathbf{B} act locally. If one defines the differential 1-form $A \equiv A_\mu dx^\mu$ (with $dx^0 = c dt$), then the components of the differential 2-form $F = dA = (1/2)(\partial_\mu A_\nu - \partial_\nu A_\mu)dx^\mu \wedge dx^\nu \equiv (1/2)F_{\mu\nu} dx^\mu \wedge dx^\nu$ are precisely the electric and magnetic fields:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E^1 & E^2 & E^3 \\ -E^1 & 0 & -B^3 & B^2 \\ -E^2 & B^3 & 0 & -B^1 \\ -E^3 & -B^2 & B^1 & 0 \end{pmatrix} \quad [8]$$

At the level of A ,

$$dF = d^2A = 0 \quad [9]$$

is an identity, but at the level of \mathbf{E} and \mathbf{B} , [9] amounts to the homogeneous (or first pair of) Maxwell equations obeyed by the field strengths:

$$\nabla \cdot \mathbf{B} = 0 \quad [10a]$$

$$\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{B} = 0 \quad [10b]$$

Therefore, these equations have a geometrical origin. The second pair of Maxwell equations is dynamical, and is obtained from the field action (in the Heaviside system of units)

$$S_{\text{field}} = -\frac{1}{4c}\int d^4x F_{\mu\nu}F^{\mu\nu} \quad [11]$$

which leads to

$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad [12a]$$

$$\nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E} = \frac{4\pi\mathbf{j}}{c} \quad [12b]$$

where $(\rho, -\mathbf{j}) = (j^0, -\mathbf{j})$ is the 4-current satisfying, as a consequence of [12a] and [12b], the conservation law

$$\partial_\mu j^\mu = 0 \quad [13]$$

For a pointlike particle, $\rho(t, \mathbf{x}) = q\delta^3(\mathbf{x} - \mathbf{x}(t))$ and $\mathbf{j} = \rho\mathbf{v}$.

Electromagnetic Fields in Quantum Physics

In quantum physics, the motion of charged particles in external electromagnetic fields is governed by the Schrödinger equation or, equivalently, by the Feynman path integral. In both cases, however, it is the 4-potential A_μ which appears in the equations, and not the field strengths. For simplicity, we consider here scalar (spinless) charged particles or unpolarized electrons (spin-(1/2)particles), both of which, in the nonrelativistic approximation, can be described quantum mechanically by a complex wave function $\psi(t, \mathbf{x})$.

To derive the Schrödinger equation, one starts from the classical Hamiltonian

$$H = \mathbf{P} \cdot \mathbf{v} - L - mc^2 = \frac{1}{2} \left(\mathbf{P} - \frac{q}{c} \mathbf{A} \right)^2 + q\varphi \quad [14]$$

where

$$\mathbf{P} = \frac{\partial}{\partial \mathbf{v}} L = \mathbf{p} + \frac{q}{c} \mathbf{A}$$

is the canonical momentum of the particle, and we have subtracted its rest energy. The replacements $\mathbf{P} \rightarrow -i\hbar\nabla$ and $H \rightarrow i\hbar\partial/\partial t$ lead to

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \psi &= \left(\frac{1}{2m} \left(i\hbar\nabla + \frac{q}{c} \mathbf{A} \right)^2 + q\varphi \right) \psi \\ &= \left(-\frac{\hbar^2}{2m} \nabla^2 + \frac{q^2}{2mc^2} A^2 \right. \\ &\quad \left. + \frac{i\hbar q}{2mc} \nabla \cdot \mathbf{A} + \frac{i\hbar q}{mc} \mathbf{A} \cdot \nabla + q\varphi \right) \psi \end{aligned} \quad [15]$$

The gauge transformation [7a] and [7b] is a symmetry of this equation, if simultaneously to the change of the 4-potential, the wave function transforms as follows:

$$\psi(t, \mathbf{x}) \rightarrow \psi'(t, \mathbf{x}) = e^{-(iq/\hbar c)\Lambda} \psi(t, \mathbf{x}) \quad [7c]$$

So, \mathbf{A}' and ψ' obey [15]. At each (t, \mathbf{x}) , $e^{-(iq/\hbar c)\Lambda}$ belongs to U(1), the unit circle in the complex plane.

In the path-integral approach, the kernel $K(t', \mathbf{x}'; t, \mathbf{x})$, which gives the probability amplitude for the propagation of the particle from the spacetime point (t, \mathbf{x}) to the spacetime point (t', \mathbf{x}') ($t < t'$), is given by

$$\begin{aligned} K(t', \mathbf{x}'; t, \mathbf{x}) &= \int_{\mathbf{x}(t)=\mathbf{x}}^{\mathbf{x}(t')=\mathbf{x}'} D\mathbf{x}(\tau) \exp\left(\frac{i}{\hbar} (S_0 + S_{\text{int}})\right) \\ &= \int_{\mathbf{x}(t)=\mathbf{x}}^{\mathbf{x}(t')=\mathbf{x}'} D\mathbf{x}(\tau) \exp\left(\frac{i}{\hbar} \int_t^{t'} d\tau \left(\frac{1}{2} m \dot{\mathbf{x}}^2 \right. \right. \\ &\quad \left. \left. + \frac{q}{c} \mathbf{A} \cdot \mathbf{v} - q\varphi \right) \right) \end{aligned}$$

$$\begin{aligned} &= \int_{\mathbf{x}(t)=\mathbf{x}}^{\mathbf{x}(t')=\mathbf{x}'} D\mathbf{x}(\tau) \exp\left(\frac{i}{\hbar} \int_t^{t'} d\tau \frac{1}{2} m \dot{\mathbf{x}}^2\right) \\ &\quad \times \exp\left(\frac{iq}{\hbar c} \int_t^{t'} (\mathbf{A} \cdot d\mathbf{x} - \varphi dx^0)\right) \\ &= \int_{\mathbf{x}(t)=\mathbf{x}}^{\mathbf{x}(t')=\mathbf{x}'} D\mathbf{x}(\tau) \exp\left(\frac{i}{\hbar} \int_t^{t'} d\tau \frac{1}{2} m \dot{\mathbf{x}}^2\right) \\ &\quad \times \exp\left(\frac{iq}{\hbar c} \int_t^{t'} dx^\mu A_\mu\right) \end{aligned} \quad [16]$$

where the integral $\int D\mathbf{x}(\tau) \dots$ is over all continuous spacetime paths $(\tau, \mathbf{x}(\tau))$ which join (t, \mathbf{x}) with (t', \mathbf{x}') . If one knows the wave function at (t, \mathbf{x}) , then the wave function at (t', \mathbf{x}') is given by

$$\psi(t', \mathbf{x}') = \int d^3\mathbf{x} K(t', \mathbf{x}'; t, \mathbf{x}) \psi(t, \mathbf{x}) \quad [17]$$

An important point is the natural appearance in the integrand of the functional integral of the factor

$$e^{(iq/\hbar c) \int_\gamma A}$$

for each path γ joining (t, \mathbf{x}) with (t', \mathbf{x}') .

A Solution to the Schrödinger Equation

In what follows, we shall restrict ourselves to static magnetic fields; then in the previous formulas, we set $\varphi = 0$ and $\mathbf{A}(t, \mathbf{x}) = \mathbf{A}(\mathbf{x})$. It is then easy to show that if \mathbf{x}_0 is an arbitrary reference point and the integral $\int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}(\mathbf{x}') \cdot d\mathbf{x}'$ is independent of the integration path from \mathbf{x}_0 to \mathbf{x} , that is, it is a well-defined function f of \mathbf{x} , and if ψ_0 is a solution of the free Schrödinger equation, that is,

$$i\hbar \frac{\partial}{\partial t} \psi_0 = -\frac{\hbar^2}{2m} \nabla^2 \psi_0 \quad [18]$$

then

$$\psi(t, \mathbf{x}) = \exp\left(\frac{iq}{\hbar c} \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}(\mathbf{x}') \cdot d\mathbf{x}'\right) \psi_0(t, \mathbf{x}) \quad [19]$$

is a solution of [15]. In fact, replacing [19] in [15], the LHS gives

$$\exp\left(\frac{iq}{\hbar c} f(\mathbf{x})\right) i\hbar \frac{\partial}{\partial t} \psi_0$$

while for the RHS one has

$$\exp\left(\frac{iq}{\hbar c} f(\mathbf{x})\right) \left(-\frac{\hbar^2}{2m}\right) \nabla^2 \psi_0$$

The cancelation of the exponential factors shows that, under the condition of path independence, there is no effect of the potential on the charged particles. Another way to see this is by making a gauge transformation [7a]–[7c] with $\Lambda(\mathbf{x}) = f(\mathbf{x})$, which changes $\psi \rightarrow \psi_0$ and $\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} - \nabla \int_{x_0}^x \mathbf{A}(\mathbf{x}') \cdot d\mathbf{x}' = \mathbf{A} - \mathbf{A} = 0$.

The condition of path independence amounts, however, to the condition that no magnetic field is present since, if $\int_{\gamma} \mathbf{A}$ depends on γ , then for some pair of paths γ and γ' from (t, \mathbf{x}) to (t', \mathbf{x}') , $0 \neq \int_{\gamma} \mathbf{A} - \int_{\gamma'} \mathbf{A} = \int_{\gamma} \mathbf{A} + \int_{-\gamma'} \mathbf{A} = \oint_{\gamma \cup (-\gamma')} \mathbf{A} = \int_{\Sigma} d\boldsymbol{\sigma} \cdot (\nabla \times \mathbf{A})$, where in the last equality we applied Stokes theorem (Σ is any surface with boundary $\gamma \cup (-\gamma')$), which shows that $\mathbf{B} = \nabla \times \mathbf{A}$ must not vanish everywhere and has a nonzero flux Φ through Σ given by

$$\Phi = \int_{\Sigma} d\boldsymbol{\sigma} \cdot \mathbf{B} \quad [20]$$

The conclusion of this section is that the ansatz [19] for solving [15] can only be applied in simply connected regions with no magnetic field strength present.

Aharonov–Bohm Proposal

In 1959, Aharonov and Bohm proposed an experiment to test, in quantum mechanics, the coupling of electric charges to electromagnetic field strengths through a local interaction with the electromagnetic potential A_{μ} , but not with the field strengths themselves. However, as we saw before, no physical effect exists, that is, A_{μ} can be gauged away, unless magnetic and/or electric fields exist somewhere, although not necessarily overlapping the wave function of the particles.

Consider the usual two-slit experiment as depicted in Figure 1, with the additional presence, behind the slits, of a long and narrow solenoid enclosing a nonvanishing magnetic flux Φ due to a constant and homogeneous magnetic field \mathbf{B} normal to the plane

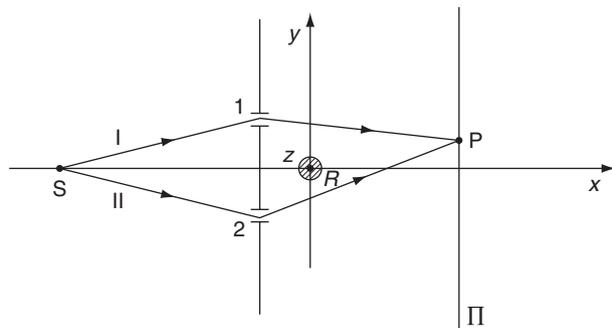


Figure 1 Magnetic Aharonov–Bohm effect.

of the figure (in direction z); outside of the solenoid, the magnetic field is zero. If the radius of the solenoid is R , a vector potential \mathbf{A} that produces such field strength is given by

$$\mathbf{A}(\mathbf{x}) = \begin{cases} (|\mathbf{B}|r/2)\hat{\varphi}, & r \leq R \\ (\Phi/2\pi r)\hat{\varphi}, & r > R \end{cases} \quad [21]$$

where $\Phi = \pi R^2|\mathbf{B}|$ and $\hat{\varphi}$ is a unit vector in the azimuthal direction. In fact,

$$\mathbf{B} = \nabla \times \mathbf{A}(\mathbf{x}) = \begin{cases} |\mathbf{B}|\hat{z}, & r \leq R \\ \mathbf{0}, & r > R \end{cases} \quad [22]$$

Notice that at $r = R$, \mathbf{A} is continuous but not continuously differentiable. Also, the ideal limit of an infinitely long solenoid makes the problem two-dimensional, that is, in the x – y plane.

The probability amplitude for an electron emitted at the source S to arrive at the point P on the screen Π , is given by the sum of two probability amplitudes, namely those corresponding to passing through the slits 1 and 2. The solenoid is assumed to be impenetrable to the electrons; mathematically, this corresponds to a motion in a non-simply-connected region. In the approximation for the path integral [16], in which one considers the contribution of only two classes of paths, that is, the class $\{\gamma\}$ represented by path I, and the class $\{\gamma'\}$ represented by path II, if the wave function at the source is ψ_S , then the wave function at P is given by

$$\begin{aligned} \psi_P &= \left(\int_{\{\gamma\}} e^{(i/\hbar)S_0(\gamma)} e^{-(i|e|/\hbar c) \int_{\gamma} \mathbf{A}} \right. \\ &\quad \left. + \int_{\{\gamma'\}} e^{(i/\hbar)S_0(\gamma')} e^{-(i|e|/\hbar c) \int_{\gamma'} \mathbf{A}} \right) \psi_S \\ &= e^{-(i|e|/\hbar c) \int_I \mathbf{A}} \int_{\{\gamma\}} e^{(i/\hbar)S_0(\gamma)} \psi_S \\ &\quad + e^{-(i|e|/\hbar c) \int_{II} \mathbf{A}} \int_{\{\gamma'\}} e^{(i/\hbar)S_0(\gamma')} \psi_S \\ &= e^{-(i|e|/\hbar c) \int_I \mathbf{A}} \left(\psi_P^0(\text{I}) \right. \\ &\quad \left. + e^{-(i|e|/\hbar c) \left(\int_{II \cup (-I)} \mathbf{A} \right)} \psi_P^0(\text{II}) \right) \\ &= e^{-(i|e|/\hbar c) \int_I \mathbf{A}} \left(\psi_P^0(\text{I}) + e^{-2\pi i(\Phi/\Phi_0)} \psi_P^0(\text{II}) \right) \end{aligned} \quad [23]$$

where, in the second line, we used the path independence of the integral of \mathbf{A} within each class of paths;

$$\psi_P^0(\text{I}) = \int_{\{\gamma\}} e^{(i/\hbar) \int_{\gamma} S_0(\gamma)} \psi_S$$

and

$$\psi_{\text{P}}^0(\text{II}) = \int_{\{\gamma'\}} e^{(i/\hbar)S_0(\gamma')} \psi_{\text{S}}$$

and, in the last equality, we applied the extended version of Stokes theorem (by Craven), to allow for noncontinuously differentiable vector potentials; and the quantum of magnetic flux associated with the charge $|e|$ is defined by

$$\Phi_0 = 2\pi \frac{\hbar c}{|e|} \cong 4.135 \times 10^{-7} \text{ G cm}^2 \quad [24]$$

($= 2\pi/|e| = \sqrt{\pi/\alpha} \cong \sqrt{137\pi}$ in the natural system of units (n.s.u.) $\hbar = c = 1$; α is the fine structure constant). Then the probability of finding the electron at P is proportional to

$$|\psi_{\text{P}}|^2 = |\psi_{\text{P}}^0(\text{I})|^2 + |\psi_{\text{P}}^0(\text{II})|^2 + 2\text{Re}(e^{2\pi i(\Phi/\Phi_0)} \psi_{\text{P}}^0(\text{I}) \psi_{\text{P}}^0(\text{II})^*) \quad [25]$$

which exhibits an interference pattern shifted with respect to that without the magnetic field: as \mathbf{B} and therefore Φ change, dark and bright interference fringes alternate periodically at the screen II, with period Φ_0 . This is the magnetic A–B effect, which has been quantitatively verified in many experiments, the first one in 1960 by [Chambers](#). The effect is:

1. *gauge invariant*, since \mathbf{B} and therefore Φ are gauge invariant;
2. *nonlocal*, since it depends on the magnetic field inside the solenoid, where the electrons never enter;
3. *quantum mechanical*, since classically the charges do not feel any force and therefore no effect would be expected in this limit; and
4. *topological*, since the electrons necessarily move in a non-simply-connected space.

But perhaps the most important implication of the A–B effect is a dramatic additional confirmation of the nonlocal character of quantum mechanics: the electron has to “travel” along the two paths (I and II) simultaneously; on the contrary, no flux would be surrounded and then no shift of the (then nonexistent) interference fringes would be observed at the screen II.

Calculations in the path-integral approach including the whole set of homotopy classes of paths around the solenoid, indexed by an integer m , have been performed by several authors, leading to a formula of the type

$$\psi_{\text{P}} = \sum_{m=-\infty}^{\infty} e^{-im\delta} \psi_{\text{P}}^0(m) \quad [26]$$

with

$$\delta = 2\pi \frac{\Phi}{\Phi_0} \quad [27]$$

([Schulman 1971](#), [Kobe 1979](#)). As in [23],

$$\psi_{\text{P}}(\Phi + k\Phi_0) = \psi_{\text{P}}(\Phi), \quad k \in \mathbb{Z} \quad [28]$$

There is a close relation between the A–B effect and the Dirac quantization condition (DQC) in the presence of electric and magnetic charges: according to [25] (or [26]) the A–B effect disappears when the flux Φ equals $n\Phi_0 = 2\pi n(\hbar c/|e|)$, $n \in \mathbb{Z}$, that is, when the condition

$$|e|\Phi = nhc \quad [29]$$

holds. But this is the DQC ([Dirac 1931](#)) when Φ is the flux associated with a magnetic charge g : $\Phi(g) = (g/4\pi r^2) \times 4\pi r^2 = g$, leading to $|e|g = nhc$ ($2\pi n$ in the n.s.u.). This is precisely the condition for the Dirac string to be unobservable in quantum mechanics: to give no A–B effect.

Geometry of the A–B Effect

In this section we study the space of gauge classes of flat potentials outside the solenoid, which determine the A–B effect; the topological structure of the A–B bundle; and the holonomy groups of the connections, which precisely give the phase shifts of the wave functions. We use the n.s.u. system; in particular, if [L] is the unit of length, then $[A_\mu] = [\text{L}]^{-1}$, $[|e|] = [\text{L}]^0$, and $\Phi_0 = 2\pi/|e| = \sqrt{\pi/\alpha} \cong \sqrt{137\pi}$, where α is the fine structure constant.

To synthesize, one can say that *the abelian A–B effect is a nonlocal gauge-invariant quantum effect due to the coupling of the wave function (section of an associated bundle) to a nontrivial (non-exact) flat (closed) connection in a trivial principal bundle with a non-simply-connected base space*. In the following subsections, we will give a detailed explanation of these statements.

The A–B Bundle

The gauge group of electromagnetism is the abelian Lie group $U(1)$ with Lie algebra (the tangent space at the identity) $\mathfrak{u}(1) = i\mathbb{R}$. In the limit of an infinitely long and infinitesimally thin solenoid carrying the magnetic flux Φ , the space available to the electrons is the plane minus a point, that is, \mathbb{R}^{2*} , which is of the same homotopy type as the circle S^1 . Then the set of isomorphism classes of $U(1)$ bundles over \mathbb{R}^{2*} is in one-to-one correspondence with the set of homotopy classes of maps from S^0 to S^1 ([Steenrod 1951](#)), which consists of only one point: if f, g :

$S^0 \rightarrow S^1$ are given by $f(1) = e^{i\varphi_1}$, $f(-1) = e^{i\varphi_2}$, $g(1) = e^{i\theta_1}$, and $g(-1) = e^{i\theta_2}$, then $H : S^0 \times [0, 1] \rightarrow S^1$ given by $H(1, t) = e^{i((1-t)\varphi_1 + t\theta_1)}$ and $H(-1, t) = e^{i((1-t)\varphi_2 + t\theta_2)}$ is a homotopy between f and g . Then, up to equivalence, the relevant bundle for the A-B effect is the product bundle

$$\xi_{A-B} : U(1) \rightarrow \mathbb{R}^{2*} \times U(1) \rightarrow \mathbb{R}^{2*} \quad [30a]$$

Since \mathbb{R}^{2*} is homeomorphic to an open disk minus a point $(D_0^2)^*$, then the total space of the bundle is homeomorphic to an open solid 2-torus minus a circle, since $(T_0^2)^* = (D_0^2)^* \times S^1$. Then the A-B bundle has the topological structure

$$\xi_{A-B} : S^1 \rightarrow (T_0^2)^* \rightarrow (D_0^2)^* \quad [30b]$$

The Gauge Group and the Moduli Space of Flat Connections

The gauge group of the bundle ξ_{A-B} is the set of smooth functions from the base space to the structure group, that is, $\mathcal{G} = C^\infty(\mathbb{R}^{2*}, U(1))$. Since $\mathcal{G} \subset C^0(\mathbb{R}^{2*}, U(1)) = \{\text{continuous functions } \mathbb{R}^{2*} \rightarrow U(1)\}$ and $[\mathbb{R}^{2*}, U(1)] = \{\text{homotopy classes of continuous functions } \mathbb{R}^{2*} \rightarrow U(1)\} \cong [S^1, S^1] \cong \pi_1(S^1) \cong \mathbb{Z}$, given $f \in \mathcal{G}$ there exists a unique $n \in \mathbb{Z}$ such that f is homotopic to $f_n (f \sim f_n)$, where $f_n : \mathbb{R}^{2*} \rightarrow U(1)$ is given by $f_n(re^{i\varphi}) = e^{in\varphi}$, $\varphi \in [0, 2\pi)$.

\mathcal{G} acts on the space of flat connections on ξ_{A-B} given by the closed $u(1)$ -valued differential 1-forms on \mathbb{R}^{2*} :

$$\mathcal{C}_0 = \{\mathcal{A} \in \Omega^1(\mathbb{R}^{2*}; u(1)), d\mathcal{A} = 0\} \quad [31]$$

through

$$\mathcal{C}_0 \times \mathcal{G} \rightarrow \mathcal{C}_0, \quad (\mathcal{A}, f) \rightarrow \mathcal{A} + f^{-1}df \quad [32]$$

where $f^{-1}(x, y) = (f(x, y))^{-1}$. The moduli space

$$\begin{aligned} \mathcal{M}_0 &= \frac{\mathcal{C}_0}{\mathcal{G}} = \{\text{gauge equivalence classes} \\ &\text{of flat connections on } \xi_{A-B}\} \\ &= \{[\mathcal{A}] = \{\mathcal{A} + f^{-1}df, f \in \mathcal{G}\}, \mathcal{A} \in \mathcal{C}_0\} \quad [33] \end{aligned}$$

is isomorphic to the circle S^1 with length 1. This can be seen as follows: the de Rham cohomology of \mathbb{R}^{2*} with coefficients in $i\mathbb{R}$ in dimension 1 is

$$\begin{aligned} H_{\text{DR}}^1(\mathbb{R}^{2*}; i\mathbb{R}) &= \{\lambda[\mathcal{A}_0]_{\text{DR}}, \lambda \in \mathbb{R}\} \\ &\cong H_{\text{DR}}^1(S^1; i\mathbb{R}) \cong \mathbb{R} \quad [34] \end{aligned}$$

where

$$\mathcal{A}_0 = i \frac{x dy - y dx}{x^2 + y^2} \in \mathcal{C}_0 \quad [35]$$

is the connection that, once multiplied by $-|e|^{-1}$ (see below) generates the flux $-\Phi_0$ and therefore no A-B effect: \mathcal{A}_0 is closed ($d\mathcal{A}_0 = 0$) but not exact ($(x dy - y dx)/(x^2 + y^2) = d\varphi$ only for $\varphi \in (0, 2\pi)$, $\varphi = 0$ is excluded); $[\mathcal{A}_0]_{\text{DR}} = \mathcal{A}_0 + d\beta$ with $\beta \in \Omega^0(\mathbb{R}^{2*}; i\mathbb{R})$. β gives an element of \mathcal{G} through the composite $\exp \circ \beta : \mathbb{R}^{2*} \rightarrow U(1)$, $(x, y) \mapsto e^{i\beta(x, y)}$. The A-B effect with flux $\Phi = -\lambda\Phi_0$ is produced by the connection $\mathcal{A} = \lambda\mathcal{A}_0$. To determine \mathcal{M}_0 , one finds the smallest $\sigma \in \mathbb{R}$ such that $(\lambda + \sigma)\mathcal{A}_0 \sim \lambda\mathcal{A}_0$, that is, $(\lambda + \sigma)\mathcal{A}_0 \in [\lambda\mathcal{A}_0]$, which means, from [33], that $(\lambda + \sigma)\mathcal{A}_0 = \lambda\mathcal{A}_0 + f^{-1}df$ or $\sigma\mathcal{A}_0 = f^{-1}df$. For $\varphi \neq 0$, $\mathcal{A}_0 = \text{id}\varphi$ and $f_1^{-1}df_1 = \text{id}\varphi$, then $\sigma = 1$, and therefore $(\lambda + 1)\mathcal{A}_0 \sim \lambda\mathcal{A}_0$, in particular $\mathcal{A}_0 \sim 0$.

A remark concerning the gauge group \mathcal{G} is the following. In classical electrodynamics, according to [7a] and [7b], the symmetry group could be taken to be the additive group $(\mathbb{R}, +)$ instead of the multiplicative group $U(1)$. Since \mathbb{R} is contractible, then the gauge group would be $\mathcal{G}_{\text{cl}} = C^\infty(\mathbb{R}^{2*}, \mathbb{R})$ with $[\mathbb{R}^{2*}, \mathbb{R}] \cong 0$, so that the homomorphism $\Psi : \mathcal{G}_{\text{cl}} \rightarrow \mathcal{G}$, $\Psi(f)(x) = e^{if(x)}$ would not exhaust \mathcal{G} since $\Psi(f) \in [1]$ for any $f \in \mathcal{G}_{\text{cl}}$: in fact, $H : \mathbb{R}^{2*} \times [0, 1] \rightarrow U(1)$ given by $H(x, t) = e^{i(1-t)f(x)}$ is a homotopy between $\Psi(f)$ and 1. However, the quantization of electric charges implies that in fact the gauge group is $U(1)$ and not \mathbb{R} . This is equivalent mathematically to the possible existence of magnetic monopoles which require nontrivial bundles for their description.

Covariant Derivative, Parallel Transport, and Holonomy

Let G be a matrix Lie group with Lie algebra \mathfrak{g} , B a differentiable manifold, $\xi : G \rightarrow P \xrightarrow{\pi} B$ a principal bundle, V a vector space, $G \times V \rightarrow V$ an action, and $\xi_V : V \rightarrow P \times_G V \xrightarrow{\pi} B$ the corresponding associated vector bundle (ξ_V is trivial if ξ is trivial). Call $\Gamma(\xi_V)$ the sections of ξ_V , $\Gamma(TB)(\Gamma(TP))$ the sections of the tangent bundle of $B(P)$, and $\Gamma_{\text{eq}}(P, V)$ the set of functions $\gamma : P \rightarrow V$ satisfying $\gamma(pg) = g^{-1}\gamma(p)$ (equivariant functions from P to V). $s \in \Gamma(\xi_V)$ induces $\gamma_s \in \Gamma_{\text{eq}}(P, V)$ with $\gamma_s(p) = s$, where $s(\pi(p)) = [p, s]$ and $\gamma \in \Gamma_{\text{eq}}(P, V)$ induces $s_\gamma \in \Gamma(\xi_V)$ with $s_\gamma(b) = [p, \gamma(p)]$, where $p \in \pi^{-1}(\{b\})$. If H is a connection on ξ , that is, a smooth assignment of a (horizontal) vector subspace H_p of T_pP at each p of P , algebraically determined by a smooth \mathfrak{g} -valued 1-form ω on P through $H_p = \ker(\omega_p)$, $s \in \Gamma(\xi_V)$, $X \in \Gamma(TB)$, and $X^\dagger \in \Gamma(TP)$ the horizontal lifting of X by ω , then $X^\dagger(\gamma_s) \in \Gamma_{\text{eq}}(P, V)$, and covariant

derivative of s with respect to ω in the direction of X is defined by

$$\nabla_X^\omega s := s_{X^\uparrow(\gamma_s)} \quad [36a]$$

If $\phi: \pi^{-1}(U) \rightarrow U \times G$ is a local trivialization of ξ , x^μ , $\mu = 1, \dots, \dim B$ are local coordinates on U , and e_i , $i = 1, \dots, \dim V$ is a basis of the local sections in $\pi^{-1}(U)$, then the local expression of [36a] is

$$\nabla_{X^\mu \partial / \partial x^\mu}^{\omega_U} (s^i e_i) = X^\mu \left(\delta_i^j \frac{\partial}{\partial x^\mu} + \mathcal{A}_{\mu i}^j \right) s^j e_j \quad [36b]$$

where

$$\mathcal{A}_{U i}^j = \mathcal{A}_{\mu i}^j dx^\mu = (\sigma^* \omega_U)_i^j \quad [36c]$$

is the geometrical gauge potential in U , given by the pullback of ω_U , the restriction of ω to $\pi^{-1}(U)$, by the local section $\sigma: U \rightarrow \pi^{-1}(U)$, $\sigma(b) = \phi^{-1}(b, 1)$. ($\mathcal{A}_{\mu i}^j$ is defined through $\nabla_{\partial / \partial x^\mu}^{\omega_U} e_i = \mathcal{A}_{\mu i}^j e_j$.) The operator

$$D_{\mu i}^j = \delta_i^j \frac{\partial}{\partial x^\mu} + \mathcal{A}_{\mu i}^j \quad [36d]$$

is the usual local covariant derivative. In an overlapping trivialization, [36b] is replaced by

$$\nabla_{X^\mu \partial / \partial x^\mu}^{\omega_{U'}} (s'^i e'_i) = X^\mu \left(\delta_i^j \frac{\partial}{\partial x^\mu} + \mathcal{A}_{\mu i}^j \right) s'^j e'_j$$

with $e'_i = g_j^k e_k$ and $s'^i = g^{-1j} s^j$ on $U \cap U'$, then the local potential transforms as

$$\mathcal{A}_{\mu l}^j = g_k^j \mathcal{A}_{\mu i}^k g^{-1i} + (\partial_\mu g_k^j) g^{-1k} \quad [36e]$$

which for G abelian has the form [32].

For each smooth path $c: [0, 1] \rightarrow B$ joining the points b and b' , and each $p \in P_b = \pi^{-1}(\{b\})$, there exists a unique path c^\uparrow in P through p with $\dot{c}^\uparrow(t) \in H_{c(t)}$ for all $t \in [0, 1]$. c^\uparrow is the horizontal lifting of c by ω through p . Thus, for each connection and path there exists a diffeomorphism $P_c^\omega: P_b \rightarrow P_{b'}$ called parallel transport. If c is a loop at b , then $P_c^\omega \in \text{Diff}(P_b)$ is called the holonomy of ω at b along c . To the loop space of B at b , $\Omega(B; b)$, corresponds a subgroup Hol_b^ω of $\text{Diff}(P_b)$ called the holonomy of ω at b . If $c \in \Omega(B; b)$ and β is a lifting of c through $q \in P_b$, then there exists a unique path $g: [0, 1] \rightarrow G$ such that $c^\uparrow(t) = \beta(t)g(t)$ with $c^\uparrow(0) = qg(0) = p$; g satisfies the differential equation

$$\frac{d}{dt} g(t) + \omega_{\beta(t)}(\dot{\beta}(t)) = 0 \quad [37]$$

whose solution is the time-ordered exponential

$$\begin{aligned} g(t)g(0)^{-1} &= T \exp \left(\int_0^t d\tau \omega_{\beta(\tau)}(\dot{\beta}(\tau)) \right) \\ &= 1 + \sum_{m=1}^{\infty} (-1)^m \int_0^t d\tau_1 \omega_{\beta(\tau_1)}(\dot{\beta}(\tau_1)) \\ &\quad \times \int_0^{\tau_1} d\tau_2 \omega_{\beta(\tau_2)}(\dot{\beta}(\tau_2)) \cdots \\ &\quad \times \int_0^{\tau_{m-1}} d\tau_m \omega_{\beta(\tau_m)}(\dot{\beta}(\tau_m)) \end{aligned} \quad [38]$$

If $q = p$ then $g(0) = 1$. For each $p \in P$, the set of elements $g' \in G$ such that $c^\uparrow(1) = pg'$ for $c \in \Omega(B; \pi(p))$ is a subgroup of G , Hol_p^ω , called the holonomy of ω at p . (For each p , there exists a group isomorphism $\text{Hol}_p^\omega \rightarrow \text{Hol}_{\pi(p)}^\omega$, and if p and p' are connected by a horizontal curve, then $\text{Hol}_p^\omega = \text{Hol}_{p'}^\omega$; if all p 's in P are horizontally connected, then $\text{Hol}_p^\omega = G$ for all $p \in P$.) If (U, ϕ) is a local trivialization of ξ , $c \subset U$, and $\beta(t) = \sigma(c(t))$, then one has the local formula

$$c^\uparrow(t) = \phi^{-1}(c(t), 1) \left(T \exp \left(- \int_{c(0)}^{c(t)} \mathcal{A}_U \right) \right) g(0) \quad [39]$$

In particular, if ξ is a product bundle, then ϕ is the identity, and choosing $g(0) = 1$ gives

$$c^\uparrow(t) = \left(c(t), T \exp \left(- \int_{c(0)}^{c(t)} \mathcal{A}_U \right) \right) \quad [40]$$

In our case, $V = \mathbb{C}$, ξ is a product bundle, $s = \psi$, the wave function, is a global section of the associated bundle

$$\xi_{\mathbb{C}}: \mathbb{C} \rightarrow \mathbb{R}^{2*} \times \mathbb{C} \xrightarrow{\pi_{\mathbb{C}}} \mathbb{R}^{2*} \quad [41]$$

$G = \text{U}(1)$ with $\mathfrak{g} = i\mathbb{R}$ and an action $\text{U}(1) \times \mathbb{C} \rightarrow \mathbb{C}$, $(e^{i\varphi}, z) \mapsto e^{i\varphi}z$; therefore, $\mathcal{A}_\mu = \mathcal{A}_{0\mu} = ia_\mu$ with a_μ real valued, and the covariant derivative is

$$D_\mu \psi = \left(\frac{\partial}{\partial x^\mu} + ia_\mu \right) \psi \quad [36f]$$

If ψ carries the electric charge q , we define the physical gauge potential A_μ through

$$a_\mu = qA_\mu \quad [42]$$

and, for the covariant derivative, after multiplying by i , we obtain the operator appearing in eqn [15], $iD_\mu \psi = (i(\partial/\partial x^\mu) - qA_\mu)\psi$: in fact, for the spatial part the coupling is $(i\nabla + q\mathbf{A})\psi$, and for the temporal part one has $(i\partial/\partial t - q\varphi)\psi$. For the electron, $q = -|e|$ and $a_\mu = -|e|A_\mu = -(2\pi/\Phi_0)A_\mu$.

For $c \in \Omega(\mathbb{R}^{2*}; (x_0, y_0))$, which turns n times around the solenoid at $(0, 0)$, eqn [40] gives

$$\begin{aligned} c^\dagger &= ((x_0, y_0), e^{-n \oint_c A}) = ((x_0, y_0), e^{-in \oint_c a}) \\ &= ((x_0, y_0), e^{-i|e|n \oint_c A \cdot dx}) = ((x_0, y_0), e^{-2\pi in \Phi / \Phi_0}) \end{aligned}$$

and therefore, for $\Phi / \Phi_0 = \lambda \in [0, 1)$ we have the holonomy groups

$$\begin{aligned} \text{Hol}_{((x_0, y_0), 1)}^{\omega(\Phi)} &= \{e^{-2\pi in(\Phi/\Phi_0)}\}_{n \in \mathbb{Z}} \\ &= \begin{cases} \mathbb{Z}_q, & \lambda = p/q, p, q \in \mathbb{Z}, (p, q) = 1 \\ \mathbb{Z}, & \lambda \notin \mathbb{Q} \end{cases} \end{aligned} \quad [43]$$

In the second case, $\text{Hol}_{((x_0, y_0), 1)}^{\omega(\Phi)}$ is dense in $U(1)$: in fact, suppose that for $n_1, n_2 \in \mathbb{Z}$, $n_1 \neq n_2$, $e^{2\pi in_1 \lambda} = e^{2\pi in_2 \lambda}$, then $e^{2\pi i(n_1 - n_2)\lambda} = 1$ and so $(n_1 - n_2)\lambda = m$ for some $m \in \mathbb{Z}$; therefore, $\lambda \in \mathbb{Q}$, which is a contradiction.

Finally, we should mention that the A–B effect can be understood as a geometric phase *à la* Berry, though not necessarily through an adiabatic change of the parameters on which the Hamiltonian depends. The Berry potential a_B turns out to be proportional to the real magnetic vector potential A : in the n.s.u., and for electrons,

$$a_B = -|e|A \quad [44]$$

Nonabelian and Gravitational A–B Effects

Since the fundamental group $\Pi_1(\mathbb{R}^{2*}, (x_0, y_0)) \cong \mathbb{Z}$, eqn [43] shows that there is a homomorphism $\varphi(\omega): \Pi_1(\mathbb{R}^{2*}, (x_0, y_0)) \rightarrow U(1)$, $\varphi(\omega)(n) = e^{-2\pi in\lambda}$, with $\varphi(\omega) (\Pi_1(\mathbb{R}^{2*})) = \text{Hol}_{((x_0, y_0), 1)}^{\omega(\Phi)}$, which characterizes the A–B effect in that case. In general, an A–B effect in a G -bundle with a connection ω is characterized by a group homomorphism from the fundamental group of the base space B onto the holonomy group of the connection, which is a subgroup of the structure group. The A–B effect is nonabelian if the holonomy group is nonabelian, which requires both G and $\Pi_1(B, x)$ to be

nonabelian. Examples with Yang–Mills and gravitational fields are considered in the literature.

Acknowledgment

The author thanks the University of Valencia, Spain, where part of this work was done.

See also: Deformation Quantization and Representation Theory; Fractional Quantum Hall Effect; Geometric Phases; Moduli Spaces: An Introduction; Quantum Chromodynamics; Variational Techniques for Ginzburg–Landau Energies.

Further Reading

- Aguilar MA and Socolovsky M (2002) Aharonov–Bohm effect, flat connections, and Green’s theorem. *International Journal of Theoretical Physics* 41: 839–860.
- Aharonov Y and Bohm D (1959) Significance of electromagnetic potentials in the quantum theory. *Physical Review* 15: 485–491.
- Berry MV (1984) Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London A* 392: 45–57.
- Chambers RG (1960) Shift of an electron interference pattern by enclosed magnetic flux. *Physical Review Letters* 5: 3–5.
- Corichi A and Pierri M (1995) Gravity and geometric phases. *Physical Review D* 51: 5870–5875.
- Dirac PMA (1931) Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London A* 133: 60–72.
- Kobe DH (1979) Aharonov–Bohm effect revisited. *Annals of Physics* 123: 381–410.
- Peshkin M and Tonomura A (1989) *The Aharonov–Bohm Effect*. Berlin: Springer.
- Schulman LS (1971) Approximate topologies. *Journal of Mathematical Physics* 12: 304–308.
- Steenrod N (1951) *The Topology of Fibre Bundles*. Princeton, NJ: Princeton University Press.
- Sundrum R and Tassie LJ (1986) Non-abelian Aharonov–Bohm effects, Feynman paths, and topology. *Journal of Mathematical Physics* 27: 1566–1570.
- Wu TT and Yang CN (1975) Concept of nonintegrable phase factors and global formulation of gauge fields. *Physical Review D* 12: 3845–3857.

Algebraic Approach to Quantum Field Theory

R Brunetti and K Fredenhagen, Universität Hamburg, Hamburg, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Quantum field theory may be understood as the incorporation of the principle of locality, which is at the basis of classical field theory, into quantum

physics. There are, however, severe obstacles against a straightforward translation of concepts of classical field theory into quantum theory, among them the notorious divergences of quantum field theory and the intrinsic nonlocality of quantum physics. Therefore, the concept of locality is somewhat obscured in the formalism of quantum field theory as it is typically exposed in textbooks. Nonlocal concepts such as the vacuum, the notion of particles or the S -matrix play a fundamental role, and neither the

relation to classical field theory nor the influence of background fields can be properly treated.

Algebraic quantum field theory (AQFT; synonymously, local quantum physics), on the contrary, aims at emphasizing the concept of locality at every instance. As the nonlocal features of quantum physics occur at the level of states (“entanglement”), not at the level of observables, it is better not to base the theory on the Hilbert space of states but on the algebra of observables. Subsystems of a given system then simply correspond to subalgebras of a given algebra. The locality concept is abstractly encoded in a notion of independence of subsystems; two subsystems are independent if the algebra of observables which they generate is isomorphic to the tensor product of the algebras of the subsystems.

Spacetime can then – in the spirit of Leibniz – be considered as an ordering device for systems. So, one associates with regions of spacetime the algebras of observables which can be measured in the pertinent region, with the condition that the algebras of subregions of a given region can be identified with subalgebras of the algebra of the region.

Problems arise if one aims at a generally covariant approach in the spirit of general relativity. Then, in order to avoid pitfalls like in the “hole problem,” systems corresponding to isometric regions must be isomorphic. Since isomorphic regions may be embedded into different spacetimes, this amounts to a simultaneous treatment of all spacetimes of a suitable class. We will see that category theory furnishes such a description, where the objects are the systems and the morphisms the embeddings of a system as a subsystem of other systems.

States arise as secondary objects via Hilbert space representations, or directly as linear functionals on the algebras of observables which can be interpreted as expectation values and are, therefore, positive and normalized. It is crucial that inequivalent representations (“sectors”) can occur, and the analysis of the structure of the sectors is one of the big successes of AQFT. One can also study the particle interpretation of certain states as well as (equilibrium and nonequilibrium) thermodynamical properties.

The mathematical methods in AQFT are mainly taken from the theory of operator algebras, a field of mathematics which developed in close contact to mathematical physics, in particular to AQFT. Unfortunately, the most important field theories, from the point of view of elementary particle physics, as quantum electrodynamics or the standard model could not yet be constructed beyond formal perturbation theory with the annoying consequence that it seemed that the concepts of AQFT could not

be applied to them. However, it has recently been shown that formal perturbation theory can be reshaped in the spirit of AQFT such that the algebras of observables of these models can be constructed as algebras of formal power series of Hilbert space operators. The price to pay is that the deep mathematics of operator algebras cannot be applied, but the crucial features of the algebraic approach can be used.

AQFT was originally proposed by Haag as a concept by which scattering of particles can be understood as a consequence of the principle of locality. It was then put into a mathematically precise form by Araki, Haag, and Kastler. After the analysis of particle scattering by Haag and Ruelle and the clarification of the relation to the Lehmann–Symanzik–Zimmermann (LSZ) formalism by Hepp, the structure of superselection sectors was studied first by Borchers and then in a fundamental series of papers by Doplicher, Haag, and Roberts (DHR) (see, e.g., Doplicher *et al.* (1971, 1974)) (soon after Buchholz and Fredenhagen established the relation to particles), and finally Doplicher and Roberts uncovered the structure of superselection sectors as the dual of a compact group thereby generalizing the Tannaka–Krein theorem of characterization of group duals.

With the advent of two-dimensional conformal field theory, new models were constructed and it was shown that the DHR analysis can be generalized to these models. Directly related to conformal theories is the algebraic approach to holography in anti-de Sitter (AdS) spacetime by Rehren.

The general framework of AQFT may be described as a covariant functor between two categories. The first one contains the information on local relations and is crucial for the interpretation. Its objects are topological spaces with additional structures (typically globally hyperbolic Lorentzian spaces, possibly spin bundles with connections, etc.), its morphisms being the structure-preserving embeddings. In the case of globally hyperbolic Lorentzian spacetimes, one requires that the embeddings are isometric and preserve the causal structure. The second category describes the algebraic structure of observables. In quantum physics the standard assumption is that one deals with the category of C^* -algebras where the morphisms are unital embeddings. In classical physics, one looks instead at Poisson algebras, and in perturbative quantum field theory one admits algebras which possess nontrivial representations as formal power series of Hilbert space operators. It is the leading principle of AQFT that the functor \mathcal{A} contains all physical information. In particular, two theories are equivalent if the corresponding functors are naturally equivalent.

In the analysis of the functor \mathcal{A} , a crucial role is played by natural transformations from other functors on the locality category. For instance, a field A may be defined as a natural transformation from the category of test function spaces to the category of observable algebras via their functors related to the locality category.

Quantum Field Theories as Covariant Functors

The rigorous implementation of the generally covariant locality principle uses the language of category theory.

The following two categories are used:

Loc: The class of objects $\text{obj}(\text{Loc})$ is formed by all (smooth) d -dimensional ($d \geq 2$ is held fixed), globally hyperbolic Lorentzian spacetimes M which are oriented and time oriented. Given any two such objects M_1 and M_2 , the morphisms $\psi \in \text{hom}_{\text{Loc}}(M_1, M_2)$ are taken to be the isometric embeddings $\psi: M_1 \rightarrow M_2$ of M_1 into M_2 but with the following constraints:

- (i) if $\gamma: [a, b] \rightarrow M_2$ is any causal curve and $\gamma(a), \gamma(b) \in \psi(M_1)$ then the whole curve must be in the image $\psi(M_1)$, that is, $\gamma(t) \in \psi(M_1)$ for all $t \in [a, b]$;
- (ii) any morphism preserves orientation and time orientation of the embedded spacetime.

The composition is defined as the composition of maps, the unit element in $\text{hom}_{\text{Loc}}(M, M)$ is given by the identical embedding $\text{id}_M: M \mapsto M$ for any $M \in \text{obj}(\text{Loc})$.

Obs: The class of objects $\text{obj}(\text{Obs})$ is formed by all C^* -algebras possessing unit elements, and the morphisms are faithful (injective) unit-preserving $*$ -homomorphisms. The composition is again defined as the composition of maps, the unit element in $\text{hom}_{\text{Obs}}(\mathcal{A}, \mathcal{A})$ is for any $\mathcal{A} \in \text{obj}(\text{Obs})$ given by the identical map $\text{id}_{\mathcal{A}}: \mathcal{A} \mapsto \mathcal{A}$.

The categories are chosen for definitiveness. One may envisage changes according to particular needs, as, for instance, in perturbation theory where instead of C^* -algebras general topological $*$ -algebras are better suited. Or one may use von Neumann algebras, in case particular states are selected. On the other hand, one might consider for **Loc** bundles over spacetimes, or (in conformally invariant theories) admit conformal embeddings as morphisms. In case one is interested in spacetimes which are not globally hyperbolic, one could look at the globally hyperbolic subregions (where one needs to be careful about the causal convexity condition (i) above).

The concept of locally covariant quantum field theory is defined as follows.

Definition 1

- (i) A locally covariant quantum field theory is a covariant functor \mathcal{A} from **Loc** to **Obs** and (writing α_ψ for $\mathcal{A}(\psi)$) with the covariance properties

$$\alpha_{\psi'} \circ \alpha_\psi = \alpha_{\psi' \circ \psi}, \quad \alpha_{\text{id}_M} = \text{id}_{\mathcal{A}(M)}$$

for all morphisms $\psi \in \text{hom}_{\text{Loc}}(M_1, M_2)$, all morphisms $\psi' \in \text{hom}_{\text{Loc}}(M_2, M_3)$, and all $M \in \text{obj}(\text{Loc})$.

- (ii) A locally covariant quantum field theory described by a covariant functor \mathcal{A} is called “causal” if the following holds: whenever there are morphisms $\psi_j \in \text{hom}_{\text{Loc}}(M_j, M)$, $j = 1, 2$, so that the sets $\psi_1(M_1)$ and $\psi_2(M_2)$ are causally separated in M , then one has

$$[\alpha_{\psi_1}(\mathcal{A}(M_1)), \alpha_{\psi_2}(\mathcal{A}(M_2))] = \{0\}$$

where the element-wise commutation makes sense in $\mathcal{A}(M)$.

- (iii) One says that a locally covariant quantum field theory given by the functor \mathcal{A} obeys the “time-slice axiom” if

$$\alpha_\psi(\mathcal{A}(M)) = \mathcal{A}(M')$$

holds for all $\psi \in \text{hom}_{\text{Loc}}(M, M')$ such that $\psi(M)$ contains a Cauchy surface for M' .

Thus, a quantum field theory is an assignment of C^* -algebras to (all) globally hyperbolic spacetimes so that the algebras are identifiable when the spacetimes are isometric, in the indicated way. This is a precise description of the generally covariant locality principle.

The Traditional Approach

The traditional framework of AQFT, in the Araki–Haag–Kastler sense, on a fixed globally hyperbolic spacetime can be recovered from a locally covariant quantum field theory, that is, from a covariant functor \mathcal{A} with the properties listed above.

Indeed, let M be an object in $\text{obj}(\text{Loc})$. $\mathcal{K}(M)$ denotes the set of all open subsets in M which are relatively compact and also contain, with each pair of points x and y , all g -causal curves in M connecting x and y (cf. condition (i) in the definition of **Loc**). $O \in \mathcal{K}(M)$, endowed with the metric of M restricted to O and with the induced orientation and time orientation, is a member of $\text{obj}(\text{Loc})$, and the injection map $\iota_{M,O}: O \rightarrow M$, that is, the identical map restricted to O , is an element in $\text{hom}_{\text{Loc}}(O, M)$.

With this notation, it is easy to prove the following assertion:

Theorem 1 *Let \mathcal{A} be a covariant functor with the above-stated properties, and define a map $\mathcal{K}(\mathbb{M}) \ni O \mapsto \mathcal{A}(O) \subset \mathcal{A}(\mathbb{M})$ by setting*

$$\mathcal{A}(O) := \alpha_{\iota_{\mathbb{M},O}}(\mathcal{A}(O))$$

Then the following statements hold:

(i) *The map fulfills isotony, that is,*

$$O_1 \subset O_2 \Rightarrow \mathcal{A}(O_1) \subset \mathcal{A}(O_2) \\ \text{for all } O_1, O_2 \in \mathcal{K}(\mathbb{M})$$

(ii) *If there exists a group G of isometric diffeomorphisms $\kappa : M \rightarrow M$ (so that $\kappa * \mathbf{g} = \mathbf{g}$) preserving orientation and time orientation, then there is a representation $G \ni \kappa \mapsto \tilde{\alpha}_\kappa$ of G by C^* -algebra automorphisms $\tilde{\alpha}_\kappa : \mathcal{A}(\mathbb{M}) \rightarrow \mathcal{A}(\mathbb{M})$ such that*

$$\tilde{\alpha}_\kappa(\mathcal{A}(O)) = \mathcal{A}(\kappa(O)), \quad O \in \mathcal{K}(\mathbb{M})$$

(iii) *If the theory given by \mathcal{A} is additionally causal, then it holds that*

$$[\mathcal{A}(O_1), \mathcal{A}(O_2)] = \{0\}$$

for all $O_1, O_2 \in \mathcal{K}(\mathbb{M})$ with O_1 causally separated from O_2 .

These properties are just the basic assumptions of the Araki–Haag–Kastler framework.

The Achievements of the Traditional Approach

In the Araki–Haag–Kastler approach in Minkowski spacetime \mathbb{M} , many results have been obtained in the last 40 years, some of them also becoming a source of inspiration to mathematics. A description of the achievements can be organized in terms of a length-scale basis, from the small to the large. We assume in this section that the algebra $\mathcal{A}(\mathbb{M})$ is faithfully and irreducibly represented on a Hilbert space \mathcal{H} , that the Poincaré transformations are unitarily implemented with positive energy, and that the subspace of Poincaré invariant vectors is one dimensional (uniqueness of the vacuum). Moreover, algebras corresponding to regions which are spacelike to a nonempty open region are assumed to be weakly closed (i.e., von Neumann algebras on \mathcal{H}), and the condition of weak additivity is fulfilled, that is, for all $O \in \mathcal{K}(\mathbb{M})$ the algebra generated from the algebras $\mathcal{A}(O+x), x \in \mathbb{M}$ is weakly dense in $\mathcal{A}(\mathbb{M})$.

Ultraviolet Structure and Idealized Localizations

This section deals with the problem of inspecting the theory at very small scales. In the limiting case, one is interested in idealized localizations, eventually the points of spacetimes. But the observable algebras are trivial at any point $x \in \mathbb{M}$, namely

$$\bigcap_{O \ni x} \mathcal{A}(O) = \mathbb{C}\mathbf{1}, \quad O \in \mathcal{K}(\mathbb{M})$$

Hence, pointlike localized observables are necessarily singular. Actually, the Wightman formulation of quantum field theory is based on the use of distributions on spacetime with values in the algebra of observables (as a topological $*$ -algebra). In spite of technical complications whose physical significance is unclear, this formalism is well suited for a discussion of the connection with the Euclidean theory, which allows, in fortunate cases, a treatment by path integrals; it is more directly related to models and admits, via the operator-product expansion, a study of the short-distance behavior. It is, therefore, an important question how the algebraic approach is related to the Wightman formalism. The reader is referred to the literature for exploring the results on this relation.

Whereas these results point to an essential equivalence of both formalisms, one needs in addition a criterion for the existence of sufficiently many Wightman fields associated with a given local net. Such a criterion can be given in terms of a compactness condition to be discussed in the next subsection. As a benefit, one derives an operator-product expansion which has to be assumed in the Wightman approach.

In the purely algebraic approach, the ultraviolet structure has been investigated by Buchholz and Verch. Small-scale properties of theories are studied with the help of the so-called scaling algebras whose elements can be described as orbits of observables under all possible renormalization group motions. There results a classification of theories in the scaling limit which can be grouped into three broad classes: theories for which the scaling limit is purely classical (commutative algebras), those for which the limit is essentially unique (stable ultraviolet fixed point) and not classical, and those for which this is not the case (unstable ultraviolet fixed point). This classification does not rely on perturbation expansions. It allows an intrinsic definition of confinement in terms of the so-called ultraparticles, that is, particles which are visible only in the scaling limit.

Phase-Space Analysis

As far as finite distances are concerned, there are two apparently competing principles, those of

nuclearity and modularity. The first one suggests that locally, after a cutoff in energy, one has a situation similar to that of old quantum mechanics, namely a finite number of states in a finite volume of phase space. Aiming at a precise formulation, Haag and Swieca introduced their notion of compactness, which Buchholz and Wichmann sharpened into that of nuclearity. The latter authors proposed that the set generated from the vacuum vector Ω ,

$$\{e^{-\beta H}A\Omega \mid A \in \mathcal{A}(O), \|A\| < 1\}$$

H denoting the generator of time translations (Hamiltonian), is nuclear for any $\beta > 0$, roughly stating that it is contained in the image of the unit ball under a trace class operator. The nuclear size $Z(\beta, O)$ of the set plays the role of the partition function of the model and has to satisfy certain bounds in the parameter β . The consequence of this constraint is the existence of product states, namely those normal states for which observables localized in two given spacelike separated regions are uncorrelated. A further consequence is the existence of thermal equilibrium states (KMS states) for all $\beta > 0$.

The second principle concerns the fact that, even locally, quantum field theory has infinitely many degrees of freedom. This becomes visible in the Reeh–Schlieder theorem, which states that every vector Φ which is in the range of $e^{-\beta H}$ for some $\beta > 0$ (in particular, the vacuum Ω) is cyclic and separating for the algebras $\mathcal{A}(O)$, $O \in \mathcal{K}(\mathbb{M})$, that is, $\mathcal{A}(O)\Phi$ is dense in \mathcal{H} (Φ is cyclic) and $A\Phi = 0, A \in \mathcal{A}(O)$ implies $A = 0$ (Φ is separating). The pair $(\mathcal{A}(O), \Omega)$ is then a von Neumann algebra in the so-called standard form. On such a pair, the Tomita–Takesaki theory can be applied, namely the densely defined operator

$$SA\Omega = A^*\Omega, \quad A \in \mathcal{A}(O)$$

is closable, and the polar decomposition of its closure $\bar{S} = J\Delta^{1/2}$ delivers an antiunitary involution J (the modular conjugation) and a positive self-adjoint operator Δ (the modular operator) associated with the standard pair $(\mathcal{A}(O), \Omega)$. These operators have the properties

$$J\mathcal{A}(O)J = \mathcal{A}(O)'$$

where the prime denotes the commutant, and

$$\Delta^{it}\mathcal{A}(O)\Delta^{-it} = \mathcal{A}(O), \quad t \in \mathbb{R}$$

The importance of this structure is based on the fact disclosed by Bisognano and Wichmann using Poincaré-covariant Wightman fields and local algebras generated by them, that for specific regions in Minkowski spacetime the modular operators have a

geometrical meaning. Indeed, these authors showed for the pair $(\mathcal{A}(W), \Omega)$, where W denotes the wedge region $W = \{x \in \mathbb{M} \mid |x^0| < x^1\}$, that the associated modular unitary Δ^{it} is the Lorentz boost with velocity $\tanh(2\pi t)$ in the direction 1 and that the modular conjugation J is the CP_1T symmetry operator with parity P_1 the reflection with respect to the $x^1 = 0$ plane. Later, Borchers discovered that already on the purely algebraic level a corresponding structure exists. He proved that, given any standard pair (\mathcal{A}, Φ) and a one-parameter group of unitaries $\tau \rightarrow U(\tau)$ acting on the Hilbert space \mathcal{H} with a positive generator and such that Φ is invariant and $U(\tau)\mathcal{A}U(\tau)^* \subset \mathcal{A}, \tau > 0$, then the associated modular operators Δ and J fulfill the commutation relations

$$\begin{aligned} \Delta^{it}U(\tau)\Delta^{-it} &= U(e^{-2\pi t}\tau) \\ JU(\tau)J &= U(-\tau) \end{aligned}$$

which are just the commutation relations between boosts and lightlike translations.

Surprisingly, there is a direct connection between the two concepts of nuclearity and modularity. Indeed, in the nuclearity condition, it is possible to replace the Hamiltonian operator by a specific function of the modular operator associated with a slightly larger region. Furthermore, under mild conditions, nuclearity and modularity together determine the structure of local algebras completely; they are isomorphic to the unique hyperfinite type III₁ von Neumann algebra.

Sectors, Symmetries, Statistics, and Particles

Large scales are appropriate for discussing global issues like superselection sectors, statistics and symmetries as far as large spacelike distances are concerned, and scattering theory, with the resulting notions of particles and infraparticles, as far as large timelike distances are concerned.

In purely massive theories, where the vacuum sector has a mass gap and the mass shell of the particles are isolated, a very satisfactory description of the multiparticle structure at large times can be given. Using the concept of almost local particle generators,

$$\Psi = A(t)\Omega$$

where Ψ is a single-particle state (i.e., an eigenstate of the mass operator), $A(t)$ is a family of almost local operators essentially localized in the kinematical region accessible from a given point by a motion with the velocities contained in the spectrum of Ψ , one obtains the multiparticle states as limits of products $A_1(t) \cdots A_n(t)\Omega$ for disjoint velocity supports. The corresponding closed subspaces are

invariant under Poincaré transformations and are unitarily equivalent to the Fock spaces of noninteracting particles.

For massless particles, no almost-local particle generators can be expected to exist. In even dimensions, however, one can exploit Huygens principle to construct asymptotic particle generators which are in the commutant of the algebra of the forward or backward lightcone, respectively. Again, their products can be determined and multiparticle states obtained.

Much less well understood is the case of massive particles in a theory which also possesses massless particles. Here, in general, the corresponding states are not eigenstates of the mass operator. Since quantum electrodynamics (QED) as well as the standard model of elementary particles have this problem, the correct treatment of scattering in these models is still under discussion. One attempt to a correct treatment is based on the concept of the so-called particle weights, that is, unbounded positive functionals on a suitable algebra. This algebra is generated by positive almost-local operators annihilating the vacuum and interpreted as counters.

The structure at large spacelike scales may be analyzed by the theory of superselection sectors. The best-understood case is that of locally generated sectors which are the objects of the DHR theory. Starting from a distinguished representation π_0 (vacuum representation) which is assumed to fulfill the Haag duality,

$$\pi_0(\mathcal{A}(O)) = \pi_0(\mathcal{A}(O'))'$$

for all double cones O , one may look at all representations which are equivalent to the vacuum representation if restricted to the observables localized in double cones in the spacelike complement of a given double cone. Such representations give rise to endomorphisms of the algebra of observables, and the product of endomorphisms can be interpreted as a product of sectors (“fusion”). In general, these representations violate the Haag duality, but there is a subclass of the so-called finite statistics sectors where the violation of Haag duality is small, in the sense that the nontrivial inclusion

$$\pi(\mathcal{A}(O)) \subset \pi(\mathcal{A}(O'))'$$

has a finite Jones index. These sectors form (in at least three spacetime dimensions) a symmetric tensor category with some further properties which can be identified, in a generalization of the Tannaka–Krein theorem, as the dual of a unique compact group. This group plays the role of a global gauge group. The symmetry of the category is expressed in terms of a

representation of the symmetric group. One may then enlarge the algebra of observables and obtain an algebra of operators which transform covariantly under the global gauge group and satisfy Bose or Fermi commutation relations for spacelike separation.

In two spacetime dimensions, one obtains instead braided tensor categories. They have been classified under additional conditions (conformal symmetry, central charge $c < 1$) in a remarkable work by Kawahigashi and Longo. Moreover, in their paper, one finds that by using completely new methods (Q-systems) a new model is unveiled, apparently inaccessible by methods used by others. To some extent, these categories can be interpreted as duals of generalized quantum groups.

The question arises whether all representations describing elementary particles are, in the massive case, DHR representations. One can show that in the case of a representation with an isolated mass shell there is an associated vacuum representation which becomes equivalent to the particle representation after restriction to observables localized spacelike to a given infinitely extended spacelike cone. This property is weaker than the DHR condition but allows, in four spacetime dimensions, the same construction of a global gauge group and of covariant fields with Bose and Fermi commutation relations, respectively, as the DHR condition. In three space dimensions, however, one finds a braided tensor category, which has similar properties as those known from topological field theories in three dimensions.

The sector structure in massless theories is not well understood, due to the infrared problem. This is in particular true for QED.

Fields as Natural Transformations

In order to be able to interpret the theory in terms of measurements, one has to be able to compare observables associated with different regions of spacetime, or, even different spacetimes. In the absence of nontrivial isometries, such a comparison can be made in terms of locally covariant fields. By definition, these are natural transformations from the functor of quantum field theory to another functor on the category of spacetimes **Loc**.

The standard case is the functor which associates with every spacetime M its space $\mathcal{D}(M)$ of smooth compactly supported test functions. There, the morphisms are the pushforwards $\mathcal{D}\psi \equiv \psi_*$.

Definition 2 A locally covariant quantum field Φ is a natural transformation between the functors \mathcal{D} and \mathcal{A} , that is, for any object M in $\text{obj}(\text{Loc})$ there exists a morphism $\Phi_M: \mathcal{D}(M) \rightarrow \mathcal{A}(M)$ such that for

any pair of objects M_1 and M_2 and any morphism ψ between them, the following diagram commutes:

$$\begin{array}{ccc} \mathcal{D}(M_1) & \xrightarrow{\Phi_{M_1}} & \mathcal{A}(M_1) \\ \psi_* \downarrow & & \downarrow \alpha_\psi \\ \mathcal{D}(M_2) & \xrightarrow{\Phi_{M_2}} & \mathcal{A}(M_2) \end{array}$$

The commutativity of the diagram means, explicitly, that

$$\alpha_\psi \circ \Phi_{M_1} = \Phi_{M_2} \circ \psi_*$$

which is the requirement sought for the covariance of fields. It contains, in particular, the standard covariance condition for spacetime isometries.

Fields in the above sense are not necessarily linear. Examples for fields which are also linear are the scalar massive free Klein–Gordon fields on all globally hyperbolic spacetimes and its locally covariant Wick polynomials. In particular, the energy–momentum tensors can be constructed as locally covariant fields, and they provide a crucial tool for discussing the back-reaction problem for matter fields.

An example for the more general notion of a field are the local S -matrices in the Stückelberg–Bogolubov–Epstein–Glaser sense. These are unitaries $S_M(\lambda)$ with $M \in \text{obj}(\text{Loc})$ and $\lambda \in \mathcal{D}(M)$ which satisfy the conditions

$$S_M(0) = 1$$

$$S_M(\lambda + \mu + \nu) = S_M(\lambda + \mu)S_M(\mu)^{-1}S_M(\mu + \nu)$$

for $\lambda, \mu, \nu \in \mathcal{D}(M)$ such that the supports of λ and ν can be separated by a Cauchy surface of M with $\text{supp } \lambda$ in the future of the surface.

The importance of these S -matrices relies on the fact that they can be used to define a new quantum field theory. The new theory is locally covariant if the original theory is and if the local S -matrices satisfy the condition of the locally covariant field above. A perturbative construction of interacting quantum field theory on globally hyperbolic spacetimes was completed in this way by Hollands and Wald, based on previous work by Brunetti and Fredenhagen.

See also: Axiomatic Quantum Field Theory; Constructive Quantum Field Theory; Current Algebra; Deformation Quantization and Representation Theory; Dispersion Relations; Indefinite Metric; Integrability and Quantum Field Theory; Operads; Perturbative Renormalization Theory and BRST; Quantum Central Limit Theorems; Quantum Field Theory: A Brief Introduction; Quantum Field Theory in Curved Spacetime; Quantum Fields with Indefinite Metric: Non-Trivial Models; Quantum Fields with Topological Defects; Quantum Geometry and its Applications; Scattering in Relativistic Quantum

Field Theory: Fundamental Concepts and Tools; Scattering in Relativistic Quantum Field Theory: The Analytic Program; Spin Foams; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Tomita–Takesaki Modular Theory; Two-Dimensional Models; von Neumann Algebras: Introduction, Modular Theory and Classification Theory; von Neumann Algebras: Subfactor Theory.

Further Reading

- Araki H (1999) *Mathematical Theory of Quantum Fields*. Oxford: Oxford University Press.
- Baumgärtel H and Wolleberg M (1992) *Causal Nets of Operator Algebras*. Berlin: Akademie Verlag.
- Borchers HJ (1996) *Translation Group and Particle Representation in Quantum Field Theory*, Lecture Notes in Physics. New Series m: Monographs, 40. Berlin: Springer.
- Borchers HJ (2000) On revolutionizing quantum field theory with Tomita’s modular theory. *Journal of Mathematical Physics* 41: 3604–3673.
- Bratteli O and Robinson DW (1987) *Operator Algebras and Quantum Statistical Mechanics*, vol. 1 Berlin: Springer.
- Brunetti R and Fredenhagen K (2000) Microlocal analysis and interacting quantum field theories: renormalization on physical backgrounds. *Communications in Mathematical Physics* 208: 623–661.
- Brunetti R, Fredenhagen K, and Verch R (2003) The generally covariant locality principle – a new paradigm for local quantum field theory. *Communications in Mathematical Physics* 237: 31–68.
- Buchholz D and Haag R (2000) The quest for understanding in relativistic quantum physics. *Journal of Mathematical Physics* 41: 3674–3697.
- Dixmier J (1964) *Les C^* -algèbres et leurs représentations*. Paris: Gauthier-Villars.
- Doplicher S, Haag R, and Roberts JE (1971) Local observables and particle statistics I. *Communications in Mathematical Physics* 23: 199–230.
- Doplicher S, Haag R, and Roberts JE (1974) Local observables and particle statistics II. *Communications in Mathematical Physics* 35: 49–85.
- Evans DE and Kawahigashi Y (1998) *Quantum Symmetries on Operator Algebras*. New York: Clarendon Press.
- Haag R (1996) *Local Quantum Physics*, 2nd edn. Berlin: Springer.
- Haag R and Kastler D (1964) An algebraic approach to quantum field theory. *Journal of Mathematical Physics* 5: 848–861.
- Hollands S and Wald RM (2001) Local Wick polynomials and time ordered products of quantum fields in curved spacetime. *Communications in Mathematical Physics* 223: 289–326.
- Hollands S and Wald RM (2002) Existence of local covariant time ordered products of quantum field in curved spacetime. *Communications in Mathematical Physics* 231: 309–345.
- Kastler D (ed.) (1990) *The Algebraic Theory of Superselection Sectors. Introductions and Recent Results*. Singapore: World Scientific.
- Kawahigashi Y and Longo R (2004) Classification of local conformal nets. Case $c < 1$. *Annals of Mathematics* 160: 1–30.
- Takesaki M (2003) *Theory of Operator Algebras I, II, III*, Encyclopedia of Mathematical Sciences, vols. 124, 125, 127. Berlin: Springer.
- Wald RM (1994) *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.

Anomalies

S L Adler, Institute for Advanced Study, Princeton, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Synopsis

Anomalies are the breaking of classical symmetries by quantum mechanical radiative corrections, which arise when the regularizations needed to evaluate small fermion loop Feynman diagrams conflict with a classical symmetry of the theory. They have important implications for a wide range of issues in quantum field theory, mathematical physics, and string theory.

Chiral Anomalies, Abelian and Nonabelian

Consider quantum electrodynamics, with the fermionic Lagrangian density

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - e_0\gamma^\mu B_\mu - m_0)\psi \quad [1a]$$

where $\bar{\psi} = \psi^\dagger\gamma^0$, e_0 and m_0 are the bare charge and mass, and B_μ is the electromagnetic gauge potential. (We reserve the notation A for axial-vector quantities.) Under a chiral transformation

$$\psi \rightarrow e^{i\lambda\gamma_5}\psi \quad [1b]$$

with constant λ , the kinetic term in eqn [1a] is invariant (because γ_5 commutes with $\gamma^0\gamma^\mu$), whereas the mass term is not invariant. Therefore, naive application of Noether's theorem would lead one to expect that the axial-vector current

$$j_\mu^5 = \bar{\psi}\gamma_\mu\gamma_5\psi \quad [1c]$$

obtained from the Lagrangian density by applying a chiral transformation with spatially varying λ , should have a divergence given by the change under chiral transformation of the mass term in eqn [1a]. Up to tree approximation, this is indeed true, but when one computes the AVV Feynman diagram with one axial-vector and two vector vertices (see **Figure 1**), and insists on conservation of the vector current $j_\mu = \bar{\psi}\gamma_\mu\psi$, one finds that to order e_0^2 , the classical Noether theorem is modified to read

$$\partial^\mu j_\mu^5(x) = 2im_0j^5(x) + \frac{e_0^2}{16\pi^2}F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho} \quad [2]$$

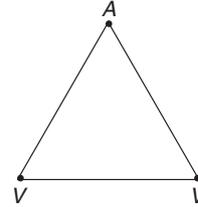


Figure 1 The AVV triangle diagram responsible for the abelian chiral anomaly.

with $F^{\xi\sigma}(x) = \partial^\sigma B^\xi(x) - \partial^\xi B^\sigma(x)$ the electromagnetic field strength tensor. The second term in eqn [2], which would be unexpected from the application of the classical Noether theorem, is the abelian axial-vector anomaly (often called the Adler–Bell–Jackiw (or ABJ) anomaly after the seminal papers on the subject). Since vector current conservation, together with the axial-vector current anomaly, implies that the left- and right-handed chiral currents $j_\mu \pm j_\mu^5$ are also anomalous, the axial-vector anomaly is frequently called the “chiral anomaly,” and we shall use the terms interchangeably in this article.

There are a number of different ways to understand why the extra term in eqn [2] appears. (1) Working through the formal Feynman diagrammatic Ward identity proof of the Noether theorem, one finds that there is a step where the closed fermion loop contributions are eliminated by a shift of the loop-integration variable. For Feynman diagrams that are convergent, this is not a problem, but the AVV diagram is linearly divergent. The linear divergence vanishes under symmetric integration, but the shift then produces a finite residue, which gives the anomaly. (2) If one defines the AVV diagram by Pauli–Villars regularization with regulator mass M_0 that is allowed to approach infinity at the end of the calculation, one finds a classical Noether theorem in the regulated theory,

$$\partial^\mu j_\mu^5|_{m_0} - \partial^\mu j_\mu^5|_{M_0} = 2im_0j^5|_{m_0} - 2iM_0j^5|_{M_0} \quad [3a]$$

with the subscripts m_0 and M_0 indicating that fermion loops are to be calculated with fermion mass m_0 and M_0 , respectively. Taking the vacuum to two-photon matrix element of eqn [3a], one finds that the matrix element $\langle 0|j_\mu^5|_{M_0}|\gamma\gamma\rangle$, which is unambiguously computable after imposing vector-current conservation, falls off only as M_0^{-1} as the regulator mass approaches infinity. Thus, the product of $2iM_0$ with this matrix element has a finite limit, which gives the anomaly. (3) If the

gauge-invariant axial-vector current is defined by point-splitting

$$j_\mu^5(x) = \bar{\psi}(x + \epsilon/2)\gamma_\mu\gamma_5\psi(x - \epsilon/2)e^{-ie_0\epsilon^\sigma B_\sigma(x)} \quad [3b]$$

with $\epsilon \rightarrow 0$ at the end of the calculation, one observes that the divergence of eqn [3b] contains an extra term with a factor of ϵ . On careful evaluation, one finds that the coefficient of this factor is an expression that behaves as ϵ^{-1} , which gives the anomaly in the limit of vanishing ϵ . (4) Finally, if the field theory is defined by a functional integral over the classical action, the standard Noether analysis shows that the classical action is invariant under the chiral transformation of eqn [1b], apart from the contribution of the mass term, which gives the naive axial-vector divergence. However, as pointed out by Fujikawa, the chiral transformation must also be applied to the functional integration measure, and since the measure is an infinite product, it must be regularized to be well defined. Careful calculation shows that the regularized measure is not chiral invariant, but contributes an extra term to the axial-vector Ward identity that is precisely the chiral anomaly.

A key feature of the anomaly is that it is irreducible: a local polynomial counter term cannot be added to the AVV diagram that preserves vector-current conservation and eliminates the anomaly. More generally, one can show that there is no way of modifying quantum electrodynamics so as to eliminate the chiral anomaly, without spoiling either vector-current conservation (i.e., electromagnetic gauge invariance), renormalizability, or unitarity. Thus, the chiral anomaly is a new physical effect in renormalizable quantum field theory, which is not present in the prequantization classical theory.

The abelian chiral anomaly is the simplest case of the anomaly phenomenon. It was extended to nonabelian gauge theories by Bardeen using a point-splitting method to compute the divergence, followed by adding polynomial counter terms to remove as many of the residual terms as possible. The resulting irreducible divergence is the nonabelian chiral anomaly, which in terms of Yang–Mills field strengths for vector and axial-vector gauge potentials V^μ and A^μ ,

$$\begin{aligned} F_V^{\mu\nu}(x) &= \partial^\mu V^\nu(x) - \partial^\nu V^\mu(x) - i[V^\mu(x), V^\nu(x)] \\ &\quad - i[A^\mu(x), A^\nu(x)] \\ F_A^{\mu\nu}(x) &= \partial^\mu A^\nu(x) - \partial^\nu A^\mu(x) - i[V^\mu(x), A^\nu(x)] \\ &\quad - i[A^\mu(x), V^\nu(x)] \end{aligned} \quad [4a]$$

is given by

$$\begin{aligned} \partial^\mu j_{5\mu}^a(x) &= \text{normal divergence term} \\ &\quad + (1/4\pi^2)\epsilon_{\mu\nu\sigma\tau}\text{tr}\lambda_A^a[(1/4)F_V^{\mu\nu}(x)F_V^{\sigma\tau}(x) \\ &\quad + (1/12)F_A^{\mu\nu}(x)F_A^{\sigma\tau}(x) \\ &\quad + (2/3)iA^\mu(x)A^\nu(x)F_V^{\sigma\tau}(x) \\ &\quad + (2/3)iF_V^{\mu\nu}(x)A^\sigma(x)A^\tau(x) \\ &\quad + (2/3)iA^\mu(x)F_V^{\nu\sigma}(x)A^\tau(x) \\ &\quad - (8/3)A^\mu(x)A^\nu(x)A^\sigma(x)A^\tau(x)] \end{aligned} \quad [4b]$$

In eqn [4b], “tr” denotes a trace over internal degrees of freedom, and λ_A^a is the internal symmetry matrix associated with the axial-vector external field. In the abelian case, where there is no internal symmetry structure, the terms involving two or four factors of A^μ, A^ν, \dots vanish by antisymmetry of $\epsilon_{\mu\nu\sigma\tau}$, and one recovers the AVV triangle anomaly, as well as a kinematically related anomaly in the AAA triangle diagram. In the nonabelian case, with nontrivial internal symmetry structure, there are also box- and pentagon-diagram anomalies.

In addition to coupling to spin-1 gauge fields, fermions can also couple to spin-2 gauge fields, associated with the graviton. When the coupling of fermions to gravitation is taken into account, the axial-vector current $\bar{\psi}T\gamma_\mu\gamma_5\psi$, with T an internal symmetry matrix, has an additional anomalous contribution to its divergence proportional to

$$\text{tr} T\epsilon_{\xi\sigma\tau\rho}R^{\xi\sigma\alpha\beta}R^{\tau\rho}_{\alpha\beta} \quad [4c]$$

where $R_{\xi\sigma\tau\rho}$ is the Riemann curvature tensor of the gravitational field.

Chiral Anomaly Nonrenormalization

A salient feature of the chiral anomaly is the fact that it is not renormalized by higher-order radiative corrections. In other words, the one-loop expressions of eqns [2] and [4b] give the exact anomaly coefficient without modification in higher orders of perturbation theory. In gauge theories such as quantum electrodynamics and quantum chromodynamics, this result (the Adler–Bardeen theorem) can be understood heuristically as follows. Write down a modified Lagrangian, in which regulators are included for all gauge-boson fields. Since the gauge-boson regulators do not influence the chiral-symmetry properties of the theory, the divergences of the chiral currents are not affected by their inclusion, and so the only sources of anomalies in the regularized theory are small single-fermion loops, giving the anomaly expressions of eqns [2] and [4b]. Since the renormalized theory is obtained as the limit of

the regularized theory as the regulator masses approach infinity, this result applies to the renormalized theory as well.

The above argument can be made precise, and extends to nongauge theories such as the σ -model as well. For both gauge theories and the σ -model, cancellation of radiative corrections to the anomaly coefficient has been explicitly demonstrated in fourth-order calculations. Nonperturbative demonstrations of anomaly renormalization have also been given using the Callan–Symanzik equations. For example, in quantum electrodynamics, Zee, and Lowenstein and Schroer, showed that a factor f that gives the ratio of the true anomaly to its one-loop value obeys the differential equation

$$\left(m \frac{\partial}{\partial m} + \alpha \beta(\alpha) \frac{\partial}{\partial \alpha}\right) f = 0 \quad [5]$$

Since f is dimensionless, it can have no dependence on the mass m , and since $\beta(\alpha)$ is nonzero this implies $\partial f / \partial \alpha = 0$. Thus, f has no dependence on α , and so $f = 1$.

Applications of Chiral Anomalies

Chiral anomalies have numerous applications in the standard model of particle physics and its extensions, and we describe here a few of the most important ones.

Neutral Pion Decay $\pi^0 \rightarrow \gamma\gamma$

As a result of the abelian chiral anomaly, the partially conserved axial-vector current (PCAC) equation relevant to neutral pion decay is modified to read

$$\partial^\mu \mathcal{F}_{3\mu}^5(x) = \left(f_\pi \mu_\pi^2 / \sqrt{2}\right) \phi_\pi(x) + S \frac{\alpha_0}{4\pi} F^{\xi\sigma}(x) F^{\tau\rho}(x) \epsilon_{\xi\sigma\tau\rho} \quad [6a]$$

with μ_π the pion mass, $f_\pi \simeq 131$ MeV the charged-pion decay constant, and S a constant determined by the constituent fermion charges and axial-vector couplings. Taking the matrix element of eqn [6a] between the vacuum state and a two-photon state, and using the fact that the left-hand side has a kinematic zero (the Sutherland–Veltman theorem), one sees that the $\pi^0 \rightarrow \gamma\gamma$ amplitude F is completely determined by the anomaly term, giving the formula

$$F = -(\alpha/\pi) 2S\sqrt{2}/f_\pi \quad [6b]$$

For a single set of fractionally charged quarks, the amplitude F is a factor of three too small to agree with experiment; for three fractionally charged

quarks (or an equivalent Han–Nambu triplet), eqn [6b] gives the correct neutral pion decay rate. This calculation was one of the first pieces of evidence for the color degree of freedom of quarks.

Anomaly Cancellation in Gauge Theories

In quantum electrodynamics, the gauge particle (the photon) couples to the vector current, and so the anomalous conservation properties of the axial-vector current have no effect. The same statement holds for the gauge gluons in quantum chromodynamics, when treated in isolation from the other interactions. However, in the electroweak theory that embeds quantum electrodynamics in a theory of the weak force, the gauge particles (the W^\pm and Z intermediate bosons) couple to chiral currents, which are left- or right-handed linear combinations of the vector and axial-vector currents. In this case, the chiral anomaly leads to problems with the renormalizability of the theory, unless the anomalies cancel between different fermion species. Writing all fermions as left-handed, the condition for anomaly cancellation is

$$\text{tr}\{T_\alpha, T_\beta\} T_\gamma = \text{tr}(T_\alpha T_\beta + T_\beta T_\alpha) T_\gamma = 0 \quad [7]$$

for all α, β, γ

with T_α the coupling matrices of gauge bosons to left-handed fermions. These conditions are obeyed in the standard model, by virtue of three nontrivial sum rules on the fermion gauge couplings being satisfied (four sum rules, if one includes the gravitational contribution to the chiral anomaly given in eqn [4c], which also cancels in the standard model). Note that anomaly cancellation in the locally gauged currents of the standard model does not imply anomaly cancellation in global-flavor currents. Thus, the flavor axial-vector current anomaly that gives the $\pi^0 \rightarrow \gamma\gamma$ matrix element remains anomalous in the full electroweak theory. Anomaly cancellation imposes important constraints on the construction of grand unified models that combine the electroweak theory with quantum chromodynamics. For instance, in $SU(5)$ the fermions are put into a $\bar{5}$ and 10 representation, which together, but not individually, are anomaly free. The larger unification groups $SO(10)$ and E_6 satisfy eqn [7] for all representations, and so are automatically anomaly free.

Instanton Physics and the Theta Vacuum

The theory of anomalies is intimately tied to the physics associated with instanton classical Yang–Mills theory solutions. Since the instanton field

strength is self-dual, the nonvanishing instanton Euclidean action

$$S_E = \int d^4x \frac{1}{4} F_{\mu\nu} F^{\mu\nu} = 8\pi^2 \quad [8a]$$

implies that the integral of the pseudoscalar density $F_{\mu\nu} F_{\lambda\sigma} \epsilon^{\mu\nu\lambda\sigma}$ over the instanton is also nonzero,

$$\int d^4x F_{\mu\nu} F_{\lambda\sigma} \epsilon^{\mu\nu\lambda\sigma} = 64\pi^2 \quad [8b]$$

Referring back to eqn [4b], this means that the integral of the nonabelian chiral anomaly for fermions in the background field of an instanton is an integer, which in the Minkowski space continuation has the interpretation of a topological winding number change produced by the instanton tunneling solution. This fact has a number of profound consequences. Since a vacuum with a definite winding number $|\nu\rangle$ is unstable under instanton tunneling, careful analysis shows that the nonabelian vacuum that has correct clustering properties is a Fourier superposition

$$|\theta\rangle = \sum_{\nu} e^{i\theta\nu} |\nu\rangle \quad [8c]$$

giving rise to the θ -vacuum of quantum chromodynamics, and a host of issues associated with (the lack of) strong CP violation, the Peccei–Quinn mechanism, and axion physics. Also, the fact that the integral of eqn [8b] is nonzero means that the $U(1)$ chiral symmetry of quantum chromodynamics is broken by instantons, which as shown by 't Hooft resolves the longstanding “ $U(1)$ problem” of strong interactions, that of explaining why the flavor singlet pseudoscalar meson η' is not light, unlike its flavor octet partners.

Anomaly Matching Conditions

The anomaly structure of a theory, as shown by 't Hooft, leads to important constraints on the formation of massless composite bound states. Consider a theory with a set of left-handed fermions ψ^{if} , with i a “color” index acted on by a nonabelian gauge force, and f an ungauged family or “flavor” index. Suppose that the family multiplet structure is such that the global chiral symmetries associated with the flavor index have nonvanishing anomalies $\text{tr}\{T_{\alpha}, T_{\beta}\}T_{\gamma}$. Then the 't Hooft condition asserts that if the color forces result in the formation of composite massless bound states of the original completely confined fermions, and if there is no spontaneous breaking of the original global flavor symmetries, then these bound states must contain left-handed spin-1/2 composites with a representation structure S that

has the same anomaly coefficient as that in the underlying theory. In other words, we must have

$$\text{tr}\{S_{\alpha}, S_{\beta}\}S_{\gamma} = \text{tr}\{T_{\alpha}, T_{\beta}\}T_{\gamma} \quad [9]$$

To prove this, one adjoins to the theory a set of right-handed spectator fermions ψ^f with the same flavor structure as the original set, but which are not acted on by the color force. These right-handed fermions cancel the original anomaly, making the underlying theory anomaly free at zero color coupling; since dynamics cannot spontaneously generate anomalies, the theory, when the color dynamics is turned on, must also have no global chiral anomalies. This implies that the bound-state spectrum must conspire to cancel the anomalies associated with the right-handed spectators; in other words, the bound-state anomaly structure must match that of the original fermions. This anomaly matching condition has found applications in the study of the possible compositeness of quarks and leptons. It has also been applied to the derivation of nonperturbative dynamical results in whole classes of supersymmetric theories, where the combined tools of holomorphicity, instanton physics, and anomaly matching have given incisive results.

Global Structure of Anomalies

We noted earlier that chiral anomalies are irreducible, in that they cannot be eliminated by adding a local polynomial counter-term to the action. However, anomalies can be described by a nonlocal effective action, obtained by integrating out the fermion field dynamics, and this point of view proves very useful in the nonabelian case. Starting with the abelian case for orientation, we note that if A^{μ} is an external axial-vector field, and we write an effective action $\Gamma[A]$, then the axial-vector current j_{μ}^5 associated with A^{μ} is given (up to an overall constant) by the variational derivative expression

$$j_{\mu}^5(x) = \frac{\delta\Gamma[A]}{\delta A^{\mu}(x)} \quad [10a]$$

and the abelian anomaly appears as the fact that the expression

$$\partial^{\mu} j_{\mu}^5 = X\Gamma[A] = G \neq 0, \quad X = \partial^{\mu} \frac{\delta}{\delta A^{\mu}(x)} \quad [10b]$$

is nonvanishing even when the theory is classically chiral invariant. Turning now to the nonabelian case, the variational derivative appearing in eqns [10a] and [10b] must be replaced by an appropriate

covariant derivative. In terms of the internal-symmetry component fields A_μ^a and V_μ^a of the Yang–Mills potentials of eqn [4a], one introduces operators

$$\begin{aligned} -X^a(x) &= \partial^\mu \frac{\delta}{\delta A_\mu^a(x)} + f_{abc} V_\mu^b \frac{\delta}{\delta A_\mu^c(x)} \\ &\quad + f_{abc} A_\mu^b \frac{\delta}{\delta V_\mu^c(x)} \\ -Y^a(x) &= \partial^\mu \frac{\delta}{\delta V_\mu^a(x)} + f_{abc} V_\mu^b \frac{\delta}{\delta V_\mu^c(x)} \\ &\quad + f_{abc} A_\mu^b \frac{\delta}{\delta A_\mu^c(x)} \end{aligned} \quad [11a]$$

with f_{abc} the antisymmetric nonabelian group structure constants. The operators X^a and Y^a are easily seen to obey the commutation relations

$$\begin{aligned} [X^a(x), X^b(y)] &= f_{abc} \delta(x-y) Y_c(x) \\ [X^a(x), Y^b(y)] &= f_{abc} \delta(x-y) X_c(x) \\ [Y^a(x), Y^b(y)] &= f_{abc} \delta(x-y) Y_c(x) \end{aligned} \quad [11b]$$

Let $\Gamma[V, A]$ be the effective action as a functional of the fields V^μ, A^μ , constructed so that the vector currents are covariantly conserved, as expressed formally by

$$Y^a \Gamma[V, A] = 0 \quad [12a]$$

Then the nonabelian axial-vector current anomaly is given by

$$X^a \Gamma[V, A] = G^a \quad [12b]$$

From eqns [12a] and [12b] and the first line of eqn [11b], we have

$$\begin{aligned} X^b G^a - X^a G^b &= (X^b X^a - X^a X^b) \Gamma[V, A] \\ &\quad + f_{abc} Y^c \Gamma[V, A] = 0 \end{aligned} \quad [12c]$$

which is the Wess–Zumino consistency condition on the structure of the anomaly G^a . It can be shown that this condition uniquely fixes the form of the nonabelian anomaly to be that of eqn [4b], up to an overall constant, which can be determined by comparison with the simplest anomalous AVV triangle graph. A physical consequence of the consistency condition is that the $\pi^0 \rightarrow \gamma\gamma$ decay amplitude determines uniquely certain other anomalous amplitudes, such as $2\gamma \rightarrow 3\pi, \gamma \rightarrow 3\pi$, and a five pseudoscalar vertex.

Although the action $\Gamma[V, A]$ is necessarily non-local, Wess and Zumino were able to write down a local action, involving an auxiliary pseudoscalar field, that obeys the anomalous Ward identities and

the consistency conditions. Subsequently, Witten gave a new construction of this local action, in terms of the integral of a fifth-rank antisymmetric tensor over a five-dimensional disk which has a four-dimensional space as its boundary. He also showed that requiring $e^{i\Gamma}$ to be independent of the choice of the spanning disk requires, in analogy with Dirac’s quantization condition for monopole charge, the condition that the overall coefficient in the nonabelian anomaly be quantized in integer multiples. Comparison with the lowest-order triangle diagram shows that in the case of $SU(N_c)$ gauge theory, this integer is just the number of colors N_c . Thus, global considerations tightly constrain the nonabelian chiral anomaly structure, and dictate that up to an integer-proportionality constant, it must have the form given in eqns [4a] and [4b].

Trace Anomalies

The discovery of chiral anomalies inspired the search for other examples of anomalous behavior. First indications of a perturbative trace anomaly obtained in a study of broken scale invariance by Coleman and Jackiw were shown by Crewther, and by Chanowitz and Ellis, to correspond to an anomaly in the three-point function $\theta_\sigma^\mu V_\mu V_\nu$, where θ_μ^μ is the energy–momentum tensor. Letting $\Delta_{\mu\nu}(p)$ be the momentum space expression for this three-point function, and $\Pi_{\mu\nu}$ the corresponding $V_\mu V_\nu$ two-point function, the trace anomaly equation in quantum electrodynamics reads

$$\begin{aligned} \Delta_{\mu\nu}(p) &= \left(2 - p_\sigma \frac{\partial}{\partial p_\sigma} \right) \Pi_{\mu\nu}(p) \\ &\quad - \frac{R}{6\pi^2} (p_\mu p_\nu - \eta_{\mu\nu} p^2) \end{aligned} \quad [13a]$$

with the first term on the right-hand side the naive divergence, and the second term the trace anomaly, with anomaly coefficient R given by

$$R = \sum_{i, \text{spin } \frac{1}{2}} Q_i^2 + \frac{1}{4} \sum_{i, \text{spin } 0} Q_i^2 \quad [13b]$$

The fact that there should be a trace anomaly can readily be inferred from a trace analog of the Pauli–Villars regulator argument for the chiral anomaly given in eqn [3a]. Letting $j = \bar{\psi}\psi$ be the scalar current in abelian electrodynamics, one has

$$\theta_\mu^\mu|_{m_0} - \theta_\mu^\mu|_{M_0} = m_0 j|_{m_0} - M_0 j|_{M_0} \quad [13c]$$

Taking the vacuum to two-photon matrix element of this equation, and imposing vector-current conservation, one finds that the matrix element $\langle 0|j|_{M_0}|\gamma\gamma\rangle$ is proportional to $M_0^{-1} \langle 0|F_{\lambda\sigma} F^{\lambda\sigma}|\gamma\gamma\rangle_{M_0}$ for a large regulator mass, and so makes a

nonvanishing contribution to the right-hand side of eqn [13c], giving the lowest-order trace anomaly.

Unlike the chiral anomaly, the trace anomaly is renormalized in higher orders of perturbation theory; heuristically, the reason is that whereas boson field regulators do not affect the chiral symmetry properties of a gauge theory (which are determined just by the fermionic terms in the Lagrangian), they do alter the energy–momentum tensor, since gravitation couples to all fields, including regulator fields. An analysis using the Callan–Symanzik equations shows, however, that the trace anomaly is computable to all orders in terms of various renormalization group functions of the coupling. For example, in abelian electrodynamics, defining $\beta(\alpha)$ and $\delta(\alpha)$ by $\beta(\alpha) = (m/\alpha)\partial\alpha/\partial m$ and $1 + \delta(\alpha) = (m/m_0)\partial m_0/\partial m$, the trace of the energy–momentum tensor is given to all orders by

$$\theta_{\mu}^{\mu} = [1 + \delta(\alpha)]m_0\bar{\psi}\psi + \frac{1}{4}\beta(\alpha)N[F_{\lambda\sigma}F^{\lambda\sigma}] + \dots \quad [14]$$

with $N[\]$ specifying conditions that make the division into two terms in eqn [14] unique, and with the ellipsis \dots indicating terms that vanish by the equations of motion. A similar relation holds in the nonabelian case, again with the β function appearing as the coefficient of the anomalous $\text{tr} N[F_{\lambda\sigma}F^{\lambda\sigma}]$ term.

Just as in the chiral anomaly case, when spin-0, spin-1/2, or spin-1 fields propagate on a background spacetime, there are curvature-dependent contributions to the trace anomaly, in other words, gravitational anomalies. These typically take the form of complicated linear combinations of terms of the form R^2 , $R_{\mu\nu}R^{\mu\nu}$, $R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma}$, $R_{,\mu}{}^{;\mu}$, with coefficients depending on the matter fields involved.

In supersymmetric theories, the axial-vector current and the energy–momentum tensor are both components of the supercurrent, and so their anomalies imply the existence of corresponding supercurrent anomalies. The issue of how the nonrenormalization of chiral anomalies (which have a supercurrent generalization given by the Konishi anomaly), and the renormalization of trace anomalies, can coexist in supersymmetric theories originally engendered considerable confusion. This apparent puzzle is now understood in the context of a perturbatively exact expression for the β function in supersymmetric field theories (the so-called NSVZ, for Novikov, Shifman, Vainshtein, and Zakharov, β function). Supersymmetry anomalies can be used to infer the structure of effective actions in supersymmetric theories, and these in turn have important implications for possibilities for dynamical supersymmetry breaking. Anomalies may also play a role, through anomaly mediation, in communicating supersymmetry breaking in “hidden

sectors” of a theory, which do not contain the physical fields that we directly observe, to the “physical sector” containing the observed fields.

Further Anomaly Topics

The above discussion has focused on some of the principal features and applications of anomalies. There are further topics of interest in the physics and mathematics of anomalies that are discussed in detail in the references cited in the “Further reading” section. We briefly describe a few of them here.

Anomalies in Other Spacetime Dimensions and in String Theory

The focus above has been on anomalies in four-dimensional spacetime, but anomalies of various types occur both in lower-dimensional quantum field theories (such as theories in two- and three-dimensional spacetimes) and in quantum field theories in higher-dimensional spacetimes (such as $N = 1$ supergravity in ten-dimensional spacetime). Anomalies also play an important role in the formulation and consistency of string theory. The bosonic string is consistent only in 26-dimensional spacetime, and the analogous supersymmetric string only in ten-dimensional spacetime, because in other dimensions both these theories violate Lorentz invariance after quantization. In the Polyakov path-integral formulation of these string theories, these special dimensions are associated with the cancellation of the Weyl anomaly, which is the relevant form of the trace anomaly discussed above. Yang–Mills, gravitational, and mixed Yang–Mills gravitational anomalies make an appearance both in $N = 1$ ten-dimensional supergravity and in superstring theory, and again special dimensions play a role. In these theories, only when the associated internal symmetry groups are either $SO(32)$ or $E_8 \times E_8$ is elimination of all anomalies possible, by cancellation of hexagon-diagram anomalies with anomalous tree diagrams involving exchange of a massless antisymmetric two-form field. This mechanism, due to Green and Schwarz, requires the factorization of a sixth-order trace invariant that appears in the hexagon anomaly in terms of lower-order invariants, as well as two numerical conditions on the adjoint representation generator structure, restricting the allowed gauge groups to the two noted above.

Covariant versus Consistent Anomalies; Descent Equations

The nonabelian anomaly of eqns [4a] and [4b] is called the “consistent anomaly,” because it obeys the

Wess–Zumino consistency conditions of eqn [12c]. This anomaly, however, is not gauge covariant, as can be seen from the fact that it involves not only the Yang–Mills field strengths $F_{V,A}^{\mu\nu}$, but the potentials V^μ, A^μ as well. It turns out to be possible, by adding appropriate polynomials to the currents, to transform the consistent anomaly to a form, called the “covariant anomaly,” which is gauge covariant under gauge transformations of the potentials V^μ, A^μ . This anomaly, however, does not obey the Wess–Zumino consistency conditions, and cannot be obtained from variation of an effective action functional.

The consistent anomalies (but not the covariant anomalies) obey a remarkable set of relations, called the Stora–Zumino descent equations, which relate the abelian anomaly in $2n + 2$ spacetime dimensions to the nonabelian anomaly in $2n$ spacetime dimensions. This set of equations has been interpreted physically by Callan and Harvey as reflecting the fact that the Dirac equation has chiral zero modes in the presence of strings in $2n + 2$ dimensions and of domain walls in $2n + 1$ dimensions.

Anomalies and Fermion Doubling in Lattice Gauge Theories

A longstanding problem in lattice formulations of gauge field theories is that when fermions are introduced on the lattice, the process of discretization introduces an undesirable doubling of the fermion particle modes. In particular, when an attempt is made to put chiral gauge theories, such as the electroweak theory, on the lattice, one finds that the doublers eliminate the chiral anomalies, by cancellation between modes with positive and negative axial-vector charge. Thus, for a long time, it appeared doubtful whether chiral gauge theories could be simulated on the lattice. However, recent work has led to formulations of lattice fermions that use a mathematical analog of a domain wall to successfully incorporate chiral fermions and the chiral anomaly into lattice gauge theory calculations.

Relation of Anomalies to the Atiyah–Singer Index Theorem

The singlet ($\lambda_A^q = 1$) anomaly of eqn [4b] is closely related to the Atiyah–Singer index theorem. Specifically, the Euclidean spacetime integral of the singlet anomaly constructed from a gauge field can be shown to give the index of the related Dirac operator for a fermion moving in that background gauge field, where the index is defined as the difference between the numbers of right- and left-handed zero-eigenvalue normalizable solutions of the Dirac equation. Since the index is a topological invariant, this again implies that the Euclidean

spacetime integral of the anomaly is a topological invariant, as noted above in our discussion of instanton-related applications of anomalies.

Retrospect

The wide range of implications of anomalies has surprised – even astonished – the founders of the subject. New anomaly applications have appeared within the last few years, and very likely the future will see continued growth of the area of quantum field theory concerned with the physics and mathematics of anomalies.

Acknowledgment

This work is supported, in part, by the Department of Energy under grant #DE-FG02-90ER40542.

See also: Bosons and Fermions in External Fields; BRST Quantization; Effective Field Theories; Gauge Theories from Strings; Gerbes in Quantum Field Theory; Index Theorems; Lagrangian Dispersion (Passive Scalar); Lattice Gauge Theory; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Electrodynamics and Its Precision Tests; Quillen Determinant; Renormalization: General Theory; Seiberg–Witten Theory.

Further Reading

- Adler SL (1969) Axial-vector vertex in spinor electrodynamics. *Physical Review* 177: 2426–2438.
- Adler SL (1970) Perturbation theory anomalies. In: Deser S, Grisaru M, and Pendleton H (eds.) *Lectures on Elementary Particles and Quantum Field Theory*, vol. 1, pp. 3–164. Cambridge, MA: MIT Press.
- Adler SL (2005) Anomalies to all orders. In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*, pp. 187–228. Singapore: World Scientific.
- Adler SL and Bardeen WA (1969) Absence of higher order corrections in the anomalous axial-vector divergence equation. *Physical Review* 182: 1517–1536.
- Bardeen W (1969) Anomalous ward identities in spinor field theories. *Physical Review* 184: 1848–1859.
- Bell JS and Jackiw R (1969) A PCAC puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ -model. *Nuovo Cimento A* 60: 47–61.
- Bertlmann RA (1996) *Anomalies in Quantum Field Theory*. Oxford: Clarendon.
- De Azcárraga JA and Izquierdo JM (1995) *Lie Groups, Lie Algebras, Cohomology and Some Applications in Physics*, ch. 10. Cambridge: Cambridge University Press.
- Fujikawa K and Suzuki H (2004) *Path Integrals and Quantum Anomalies*. Oxford: Oxford University Press.
- Golterman M (2001) Lattice chiral gauge theories. *Nuclear Physics Proceeding Supplements* 94: 189–203.
- Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*. vol. 2, sects. 13.3–13.5. Cambridge: Cambridge University Press.
- Hasenfratz P (2005) Chiral symmetry on the lattice. In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*, pp. 377–398. Singapore: World Scientific.

- Jackiw R (1985) Field theoretic investigations in current algebra and topological investigations in quantum gauge theories. In: Treiman S, Jackiw R, Zumino B, and Witten E (eds.) *Current Algebra and Anomalies*. Singapore: World Scientific and Princeton: Princeton University Press.
- Jackiw R (2005) Fifty years of Yang–Mills theory and our moments of triumph. In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*, pp. 229–251. Singapore: World Scientific.
- Makeenko Y (2002) *Methods of Contemporary Gauge Theory*, ch. 3. Cambridge: Cambridge University Press.
- Neuberger H (2000) Chiral fermions on the lattice. *Nuclear Physics Proceeding Supplements* 83: 67–76.
- Polchinski J (1999) *String Theory*, vol. 1, sect. 3.4; vol. 2, sect. 12.2. Cambridge: Cambridge University Press.
- Shifman M (1997) Non-perturbative dynamics in supersymmetric gauge theories. *Progress in Particle and Nuclear Physics* 39: 1–116.
- van Nieuwenhuizen P (1988) *Anomalies in Quantum Field Theory: Cancellation of Anomalies in $d=10$ Supergravity*. Leuven: Leuven University Press.
- Volovik GE (2003) *The Universe in a Helium Droplet*, ch. 18. Oxford: Clarendon.
- Weinberg S (1996) *The Quantum Theory of Fields, Vol. II Modern Applications*, ch. 22. Cambridge: Cambridge University Press.
- Zee A (2003) *Quantum Field Theory in a Nutshell*, sect. IV.7. Princeton: Princeton University Press.

Arithmetic Quantum Chaos

J Marklof, University of Bristol, Bristol, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The central objective in the study of quantum chaos is to characterize universal properties of quantum systems that reflect the regular or chaotic features of the underlying classical dynamics. Most developments of the past 25 years have been influenced by the pioneering models on statistical properties of eigenstates (Berry 1977) and energy levels (Berry and Tabor 1977, Bohigas *et al.* 1984). Arithmetic quantum chaos (AQC) refers to the investigation of quantum systems with additional arithmetic structures that allow a significantly more extensive analysis than is generally possible. On the other hand, the special number-theoretic features also render these systems nongeneric, and thus some of the expected universal phenomena fail to emerge. Important examples of such systems include the modular surface and linear automorphisms of tori (“cat maps”) which will be described below.

The geodesic motion of a point particle on a compact Riemannian surface \mathcal{M} of constant negative curvature is the prime example of an Anosov flow, one of the strongest characterizations of dynamical chaos. The corresponding quantum eigenstates φ_j and energy levels λ_j are given by the solution of the eigenvalue problem for the Laplace–Beltrami operator Δ (or Laplacian for short)

$$(\Delta + \lambda)\varphi = 0, \quad \|\varphi\|_{L^2(\mathcal{M})} = 1 \quad [1]$$

where the eigenvalues

$$\lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots \rightarrow \infty \quad [2]$$

form a discrete spectrum with an asymptotic density governed by Weyl’s law

$$\#\{j : \lambda_j \leq \lambda\} \sim \frac{\text{Area}(\Gamma \backslash \mathbb{H})}{4\pi} \lambda, \quad \lambda \rightarrow \infty \quad [3]$$

We rescale the sequence by setting

$$X_j = \frac{\text{Area}(\Gamma \backslash \mathbb{H})}{4\pi} \lambda_j \quad [4]$$

which yields a sequence of asymptotic density 1. One of the central conjectures in AQC says that, if \mathcal{M} is an arithmetic hyperbolic surface (see the next section for examples of this very special class of surfaces of constant negative curvature), the eigenvalues of the Laplacian have the same local statistical properties as independent random variables from a Poisson process (see, e.g., the surveys by Sarnak (1995) and Bogomolny *et al.* (1997)). This means that the probability of finding k eigenvalues X_j in randomly shifted interval $[X, X + L]$ of fixed length L is distributed according to the Poisson law $L^k e^{-L}/k!$. The gaps between eigenvalues have an exponential distribution,

$$\frac{1}{N} \#\{j \leq N : X_{j+1} - X_j \in [a, b]\} \rightarrow \int_a^b e^{-s} ds \quad [5]$$

as $N \rightarrow \infty$, and thus eigenvalues are likely to appear in clusters. This is in contrast to the general expectation that the energy level statistics of generic chaotic systems follow the distributions of random matrix ensembles; Poisson statistics are usually associated with quantized integrable systems. Although we are at present far from a proof of [5], the deviation from random matrix theory is well understood (see the section “Eigenvalue statistics and Selberg trace formula”).

Highly excited quantum eigenstates $\varphi_j(j \rightarrow \infty)$ (cf. **Figure 1**) of chaotic systems are conjectured to behave locally like random wave solutions of [1],

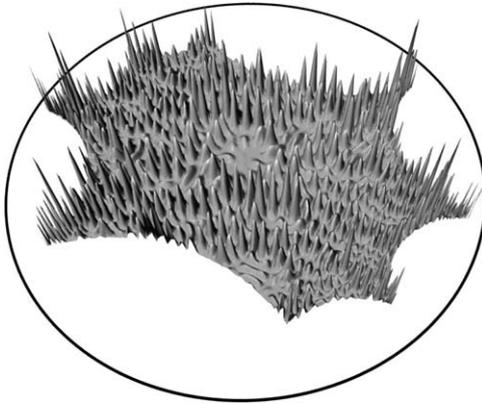


Figure 1 Image of the absolute-value-squared of an eigenfunction $\varphi_j(z)$ for a nonarithmetic surface of genus 2. The surface is obtained by identifying opposite sides of the fundamental region. Reproduced from Aurich and Steiner (1993) Statistical properties of highly excited quantum eigenstates of a strongly chaotic system. *Physica D* 64(1–3): 185–214, with permission from R Aurich.

where boundary conditions are ignored. This hypothesis was put forward by Berry in 1977 and tested numerically, for example, in the case of certain arithmetic and nonarithmetic surfaces of constant negative curvature (Hejhal and Rackner 1992, Aurich and Steiner 1993). One of the implications is that eigenstates should have uniform mass on the surface \mathcal{M} , that is, for any bounded continuous function $g: \mathcal{M} \rightarrow \mathbb{R}$

$$\int_{\mathcal{M}} |\varphi_j|^2 g \, dA \rightarrow \int_{\mathcal{M}} g \, dA, \quad j \rightarrow \infty \quad [6]$$

where dA is the Riemannian area element on \mathcal{M} . This phenomenon, referred to as quantum unique ergodicity (QUE), is expected to hold for general surfaces of negative curvature, according to a conjecture by Rudnick and Sarnak (1994). In the case of arithmetic hyperbolic surfaces, there has been substantial progress on this conjecture in the works of Lindenstrauss, Watson, and Luo–Sarnak (discussed later in this article; see also the review by Sarnak (2003)). For general manifolds with ergodic geodesic flow, the convergence in [6] is so far established only for subsequences of eigenfunctions of density 1 (Schnirelman–Zelditch–Colin de Verdière theorem, see Quantum Ergodicity and Mixing of Eigenfunctions), and it cannot be ruled out that exceptional subsequences of eigenfunctions have singular limit, for example, localized on closed geodesics. Such “scarring” of eigenfunctions, at least in some weak form, has been suggested by numerical experiments in Euclidean domains, and the existence of singular quantum limits is a matter of controversy

in the current physics and mathematics literature. A first rigorous proof of the existence of scarred eigenstates has recently been established in the case of quantized toral automorphisms. Remarkably, these quantum cat maps may also exhibit QUE. A more detailed account of results for these maps is given in the section “Quantum eigenstates of cat maps”; see also Rudnick (2001) and De Bièvre (to appear).

There have been a number of other fruitful interactions between quantum chaos and number theory, in particular the connections of spectral statistics of integrable quantum systems with the value distribution properties of quadratic forms, and analogies in the statistical behavior of energy levels of chaotic systems and the zeros of the Riemann zeta function. We refer the reader to Marklof (2006) and Berry and Keating (1999), respectively, for information on these topics.

Hyperbolic Surfaces

Let us begin with some basic notions of hyperbolic geometry. The hyperbolic plane \mathbb{H} may be abstractly defined as the simply connected two-dimensional Riemannian manifold with Gaussian curvature -1 . A convenient parametrization of \mathbb{H} is provided by the complex upper-half plane, $\mathfrak{H} = \{x + iy: x \in \mathbb{R}, y > 0\}$, with Riemannian line and volume elements

$$ds^2 = \frac{dx^2 + dy^2}{y^2}, \quad dA = \frac{dx \, dy}{y^2} \quad [7]$$

respectively. The group of orientation-preserving isometries of \mathbb{H} is given by fractional linear transformations

$$\begin{aligned} \mathfrak{H} \rightarrow \mathfrak{H}, \quad z \mapsto \frac{az + b}{cz + d} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{R}) \end{aligned} \quad [8]$$

where $\text{SL}(2, \mathbb{R})$ is the group of 2×2 matrices with unit determinant. Since the matrices 1 and -1 represent the same transformation, the group of orientation-preserving isometries can be identified with $\text{PSL}(2, \mathbb{R}) := \text{SL}(2, \mathbb{R}) / \{\pm 1\}$. A finite-volume hyperbolic surface may now be represented as the quotient $\Gamma \backslash \mathbb{H}$, where $\Gamma \subset \text{PSL}(2, \mathbb{R})$ is a Fuchsian group of the first kind. An arithmetic hyperbolic surface (such as the modular surface) is obtained, if Γ has, loosely speaking, some representation in $n \times n$ matrices with integer coefficients, for some suitable n .

This is evident in the case of the modular surface, where the fundamental group is the modular group

$$\Gamma = \text{PSL}(2, \mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{PSL}(2, \mathbb{R}) : a, b, c, d \in \mathbb{Z} \right\} / \{\pm 1\}$$

A fundamental domain for the action of the modular group $\text{PSL}(2, \mathbb{Z})$ on \mathfrak{H} is the set

$$\mathcal{F}_{\text{PSL}(2, \mathbb{Z})} = \left\{ z \in \mathfrak{H} : |z| > 1, -\frac{1}{2} < \text{Re } z < \frac{1}{2} \right\} \quad [9]$$

(see **Figure 2**). The modular group is generated by the translation

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} : z \mapsto z + 1$$

and the inversion

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} : z \mapsto -1/z$$

These generators identify sections of the boundary of $\mathcal{F}_{\text{PSL}(2, \mathbb{Z})}$. By gluing the fundamental domain along identified edges, we obtain a realization of the modular surface, a noncompact surface with one cusp at $z \rightarrow \infty$, and two conic singularities at $z = i$ and $z = 1/2 + i\sqrt{3}/2$.

An interesting example of a compact arithmetic surface is the “regular octagon,” a hyperbolic surface of genus 2. Its fundamental domain is shown in **Figure 3** as a subset of the Poincaré disk $\mathfrak{D} = \{z \in \mathbb{C} : |z| < 1\}$, which yields an alternative parametrization of the hyperbolic plane \mathbb{H} . In these coordinates, the Riemannian line and volume element read

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}, \quad dA = \frac{4dx dy}{(1 - x^2 - y^2)^2} \quad [10]$$

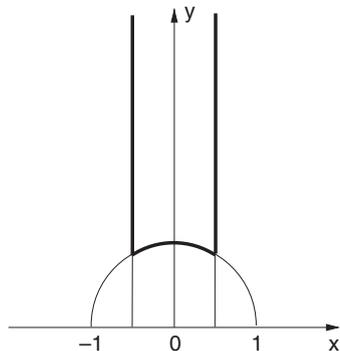


Figure 2 Fundamental domain of the modular group $\text{PSL}(2, \mathbb{Z})$ in the complex upper-half plane.

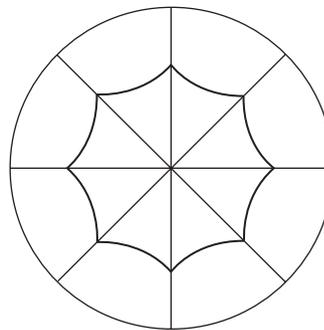


Figure 3 Fundamental domain of the regular octagon in the Poincaré disk.

The group of orientation-preserving isometries is now represented by $\text{PSU}(1, 1) = \text{SU}(1, 1) / \{\pm 1\}$, where

$$\text{SU}(1, 1) = \left\{ \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 - |\beta|^2 = 1 \right\} \quad [11]$$

acting on \mathfrak{D} as above via fractional linear transformations. The fundamental group of the regular octagon surface is the subgroup of all elements in $\text{PSU}(1, 1)$ with coefficients of the form

$$\alpha = k + l\sqrt{2}, \quad \beta = (m + n\sqrt{2})\sqrt{1 + \sqrt{2}} \quad [12]$$

where $k, l, m, n \in \mathbb{Z}[i]$, that is, Gaussian integers of the form $k_1 + ik_2, k_1, k_2 \in \mathbb{Z}$. Note that not all choices of $k, l, m, n \in \mathbb{Z}[i]$ satisfy the condition $|\alpha|^2 - |\beta|^2 = 1$. Since all elements $\gamma \neq 1$ of Γ act fix-point free on \mathbb{H} , the surface $\Gamma \backslash \mathbb{H}$ is smooth without conic singularities.

In the following, we will restrict our attention to a representative case, the modular surface with $\Gamma = \text{PSL}(2, \mathbb{Z})$.

Eigenvalue Statistics and Selberg Trace Formula

The statistical properties of the rescaled eigenvalues X_j (cf. [4]) of the Laplacian can be characterized by their distribution in small intervals

$$\mathcal{N}(x, L) := \#\{j : x \leq X_j \leq x + L\} \quad [13]$$

where x is uniformly distributed, say, in the interval $[X, 2X]$, X large. Numerical experiments by Bogomolny, Georgeot, Giannoni, and Schmit, as well as Bolte, Steil, and Steiner (see references in

Bogomolny (1997)) suggest that the X_j are asymptotically Poisson distributed:

Conjecture 1 For any bounded function $g: \mathbb{Z}_{\geq 0} \rightarrow \mathbb{C}$ we have

$$\frac{1}{X} \int_X^{2X} g(\mathcal{N}(x, L)) dx \rightarrow \sum_{k=0}^{\infty} g(k) \frac{L^k e^{-L}}{k!} \quad [14]$$

as $T \rightarrow \infty$.

One may also consider larger intervals, where $L \rightarrow \infty$ as $X \rightarrow \infty$. In this case, the assumption on the independence of the X_j predicts a central-limit theorem. Weyl’s law [3] implies that the expectation value is asymptotically, for $T \rightarrow \infty$,

$$\frac{1}{X} \int_X^{2X} \mathcal{N}(x, L) dx \sim L \quad [15]$$

This asymptotics holds for any sequence of L bounded away from zero (e.g., L constant, or $L \rightarrow \infty$).

Define the variance by

$$\Sigma^2(X, L) = \frac{1}{X} \int_X^{2X} (\mathcal{N}(x, L) - L)^2 dx \quad [16]$$

In view of the above conjecture, one expects $\Sigma^2(X, L) \sim L$ in the limit $X \rightarrow \infty, L/\sqrt{X} \rightarrow 0$ (the variance exhibits a less universal behavior in the range $L \gg \sqrt{X}$ (the notation $A \ll B$ means there is a constant $c > 0$ such that $A \leq cB$), cf. Sarnak (1995), and a central-limit theorem for the fluctuations around the mean:

Conjecture 2 For any bounded function $g: \mathbb{R} \rightarrow \mathbb{C}$ we have

$$\begin{aligned} & \frac{1}{X} \int_X^{2X} g\left(\frac{\mathcal{N}(x, L) - L}{\sqrt{\Sigma^2(x, L)}}\right) dx \\ & \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t) e^{-(1/2)t^2} dt \end{aligned} \quad [17]$$

as $X, L \rightarrow \infty, L \ll X$.

The main tool in the attempts to prove the above conjectures has been the Selberg trace formula. It relates sums over eigenvalues of the Laplacians to sums over lengths of closed geodesics on the hyperbolic surface. The trace formula is in its simplest form in the case of compact hyperbolic surfaces; we have

$$\begin{aligned} \sum_{j=0}^{\infty} b(\rho_j) &= \frac{\text{Area}(\mathcal{M})}{4\pi} \int_{-\infty}^{\infty} b(\rho) \tanh(\pi\rho) \rho d\rho \\ &+ \sum_{\gamma \in H_*} \sum_{n=1}^{\infty} \frac{\ell_{\gamma} g(n\ell_{\gamma})}{2 \sinh(n\ell_{\gamma}/2)} \end{aligned} \quad [18]$$

where H_* is the set of all primitive oriented closed geodesics γ , and ℓ_{γ} their lengths. The quantity ρ_j is related to the eigenvalue λ_j by the equation $\lambda_j = \rho_j^2 + 1/4$. The trace formula [18] holds for a large class of even test functions b . For example, it is sufficient to assume that b is infinitely differentiable, and that the Fourier transform of b ,

$$g(t) = \frac{1}{2\pi} \int_{\mathbb{R}} b(\rho) e^{-i\rho t} d\rho \quad [19]$$

has compact support. The trace formula for non-compact surfaces has additional terms from the parabolic elements in the corresponding group, and includes also sums over the resonances of the continuous part of the spectrum. The noncompact modular surface behaves in many ways like a compact surface. In particular, Selberg showed that the number of eigenvalues embedded in the continuous spectrum satisfies the same Weyl law as in the compact case (Sarnak 2003).

Setting

$$b(\rho) = \chi_{[X, X+L]} \left(\frac{\text{Area}(\mathcal{M})}{4\pi} \left(\rho^2 + \frac{1}{4} \right) \right) \quad [20]$$

where $\chi_{[X, X+L]}$ is the characteristic function of the interval $[X, X+L]$, we may thus view $\mathcal{N}(X, L)$ as the left-hand side of the trace formula. The above test function b is, however, not admissible, and requires appropriate smoothing. Luo and Sarnak (cf. Sarnak (2003)) developed an argument of this type to obtain a lower bound on the average number variance,

$$\frac{1}{L} \int_0^L \Sigma^2(X, L') dL' \gg \frac{\sqrt{X}}{(\log X)^2} \quad [21]$$

in the regime $\sqrt{X}/\log X \ll L \ll \sqrt{X}$, which is consistent with the Poisson conjecture $\Sigma^2(X, L) \sim L$. Bogomolny, Levyraz, and Schmit suggested a remarkable limiting formula for the two-point correlation function for the modular surface (cf. Bogomolny *et al.* (1997) and Bogomolny (2006)), based on an analysis of the correlations between multiplicities of lengths of closed geodesics. A rigorous analysis of the fluctuations of multiplicities is given by Peter (cf. Bogomolny (2006)). Rudnick (2005) has recently established a smoothed version of Conjecture 2 in the regime

$$\frac{\sqrt{X}}{L} \rightarrow \infty, \quad \frac{\sqrt{X}}{L \log X} \rightarrow 0 \quad [22]$$

where the characteristic function in [20] is replaced by a certain class of smooth test functions.

All of the above approaches use the Selberg trace formula, exploiting the particular properties of the

distribution of lengths of closed geodesics in arithmetic hyperbolic surfaces. These will be discussed in more detail in the next section, following the work of Bogomolny, Georgeot, Giannoni and Schmit, Bolte, and Luo and Sarnak (see [Bogomolny et al. \(1997\)](#) and [Sarnak \(1995\)](#) for references).

Distribution of Lengths of Closed Geodesics

The classical prime geodesic theorem asserts that the number $N(\ell)$ of primitive closed geodesics of length less than ℓ is asymptotically

$$N(\ell) \sim \frac{e^\ell}{\ell} \tag{23}$$

One of the significant geometrical characteristics of arithmetic hyperbolic surfaces is that the number of closed geodesics with the same length ℓ grows exponentially with ℓ . This phenomenon is most easily explained in the case of the modular surface, where the set of lengths ℓ appearing in the lengths spectrum is characterized by the condition

$$2 \cosh(\ell/2) = |\text{tr } \gamma| \tag{24}$$

where γ runs over all elements in $\text{SL}(2, \mathbb{Z})$ with $|\text{tr } \gamma| > 2$. It is not hard to see that any integer $n > 2$ appears in the set $\{|\text{tr } \gamma| : \gamma \in \text{SL}(2, \mathbb{Z})\}$, and hence the set of distinct lengths of closed geodesics is

$$\mathcal{L} = \{2 \operatorname{arcosh}(n/2) : n = 3, 4, 5, \dots\} \tag{25}$$

Therefore, the number of distinct lengths less than ℓ is asymptotically (for large ℓ)

$$N'(\ell) = \#\mathcal{L} \cap [0, \ell] \sim e^{\ell/2} \tag{26}$$

[Equations \[26\] and \[23\]](#) say that on average the number of geodesics with the same lengths is at least $\asymp e^{\ell/2}/\ell$.

The prime geodesic theorem [\[23\]](#) holds equally for all hyperbolic surfaces with finite area, while [\[26\]](#) is specific to the modular surface. For general arithmetic surfaces, we have the upper bound

$$N'(\ell) \leq c e^{\ell/2} \tag{27}$$

for some constant $c > 0$ that may depend on the surface. Although one expects $N'(\ell)$ to be asymptotic to $(1/2)N(\ell)$ for generic surfaces (since most geodesics have a time-reversal partner which thus has the same length, and otherwise all lengths are distinct), there are examples of nonarithmetic Hecke triangles where numerical and heuristic arguments suggest $N'(\ell) \sim c_1 e^{c_2 \ell}/\ell$ for suitable constants $c_1 > 0$ and $0 < c_2 < 1/2$ (cf. [Bogomolny \(2006\)](#)). Hence

exponential degeneracy in the length spectrum seems to occur in a weaker form also for nonarithmetic surfaces.

A further useful property of the length spectrum of arithmetic surfaces is the bounded clustering property: there is a constant C (again surface dependent) such that

$$\#(\mathcal{L} \cap [\ell, \ell + 1]) \leq C \tag{28}$$

for all ℓ . This fact is evident in the case of the modular surface; the general case is proved by Luo and Sarnak (cf. [Sarnak \(1995\)](#)).

Quantum Unique Ergodicity

The unit tangent bundle of a hyperbolic surface $\Gamma \backslash \mathbb{H}$ describes the physical phase space on which the classical dynamics takes place. A convenient parametrization of the unit tangent bundle is given by the quotient $\Gamma \backslash \text{PSL}(2, \mathbb{R})$ – this may be seen by means of the Iwasawa decomposition for an element $g \in \text{PSL}(2, \mathbb{R})$,

$$g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \times \begin{pmatrix} \cos \theta/2 & \sin \theta/2 \\ -\sin \theta/2 & \cos \theta/2 \end{pmatrix} \tag{29}$$

where $x + iy \in \mathfrak{H}$ represents the position of the particle in $\Gamma \backslash \mathbb{H}$ in half-plane coordinates, and $\theta \in [0, 2\pi)$ the direction of its velocity. Multiplying the matrix [\[29\]](#) from the left by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and writing the result again in the Iwasawa form [\[29\]](#), one obtains the action

$$(z, \phi) \mapsto \left(\frac{az + b}{cz + d}, \theta - 2 \arg(cz + d) \right) \tag{30}$$

which represents precisely the geometric action of isometries on the unit tangent bundle.

The geodesic flow Φ^t on $\Gamma \backslash \text{PSL}(2, \mathbb{R})$ is represented by the right translation

$$\Phi^t : \Gamma g \mapsto \Gamma g \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \tag{31}$$

The Haar measure μ on $\text{PSL}(2, \mathbb{R})$ is thus trivially invariant under the geodesic flow. It is well known that μ is not the only invariant measure, that is, Φ^t is not uniquely ergodic, and that there is in fact an abundance of invariant measures. The simplest examples are those with uniform mass on one, or a countable collection of, closed geodesics.

To test the distribution of an eigenfunction ψ_j in phase space, one associates with a function

$a \in C^\infty(\Gamma \backslash \text{PSL}(2, \mathbb{R}))$ the quantum observable $\text{Op}(a)$, a zeroth order pseudodifferential operator with principal symbol a . Using semiclassical techniques based on Friedrich’s symmetrization, one can show that the matrix element

$$\nu_j(a) = \langle \text{Op}(a)\varphi_j, \varphi_j \rangle \tag{32}$$

is asymptotic (as $j \rightarrow \infty$) to a positive functional that defines a probability measure on $\Gamma \backslash \text{PSL}(2, \mathbb{R})$. Therefore, if \mathcal{M} is compact, any weak limit of ν_j represents a probability measure on $\Gamma \backslash \text{PSL}(2, \mathbb{R})$. Egorov’s theorem (see Quantum Ergodicity and Mixing of Eigenfunctions) in turn implies that any such limit must be invariant under the geodesic flow, and the main challenge in proving QUE is to rule out all invariant measures apart from Haar.

Conjecture 3 (Rudnick and Sarnak (1994); see Sarnak (1995, 2003)). For every compact hyperbolic surface $\Gamma \backslash \mathbb{H}$, the sequence ν_j converges weakly to μ .

Lindenstrauss has proved this conjecture for compact arithmetic hyperbolic surfaces of congruence type (such as the second example in the section “Hyperbolic surfaces”) for special bases of eigenfunctions, using ergodic-theoretic methods. These will be discussed in more detail in the next section. His results extend to the noncompact case, that is, to the modular surface where $\Gamma = \text{PSL}(2, \mathbb{Z})$. Here he shows that any weak limit of subsequences of ν_j is of the form $c\mu$, where c is a constant with values in $[0, 1]$. One believes that $c = 1$, but with present techniques it cannot be ruled out that a proportion of the mass of the eigenfunction escapes into the noncompact cusp of the surface. For the modular surface, $c = 1$ can be proved under the assumption of the generalized Riemann hypothesis (see the section “Eigenfunctions and L -functions” and Sarnak (2003)). QUE also holds for the continuous part of the spectrum, which is furnished by the Eisenstein series $E(z, s)$, where $s = 1/2 + ir$ is the spectral parameter. Note that the measures associated with the matrix elements

$$\nu_r(a) = \langle \text{Op}(a)E(\cdot, 1/2 + ir), E(\cdot, 1/2 + ir) \rangle \tag{33}$$

are not probability measures but only Radon measures, since $E(z, s)$ is not square-integrable. Luo and Sarnak, and Jakobson have shown that

$$\lim_{r \rightarrow \infty} \frac{\nu_r(a)}{\nu_r(b)} = \frac{\mu(a)}{\mu(b)} \tag{34}$$

for suitable test functions $a, b \in C^\infty(\Gamma \backslash \text{PSL}(2, \mathbb{R}))$ (cf. Sarnak (2003)).

Hecke Operators, Entropy and Measure Rigidity

For compact surfaces, the sequence of probability measures approaching the matrix elements ν_j is relatively compact. That is, every infinite sequence contains a convergent subsequence. Lindenstrauss’ central idea in the proof of QUE is to exploit the presence of Hecke operators to understand the invariance properties of possible quantum limits. We will sketch his argument in the case of the modular surface (ignoring issues related to the non-compactness of the surface), where it is most transparent.

For every positive integer n , the Hecke operator T_n acting on continuous functions on $\Gamma \backslash \mathbb{H}$ with $\Gamma = \text{SL}(2, \mathbb{Z})$ is defined by

$$T_n f(z) = \frac{1}{\sqrt{n}} \sum_{\substack{a,d=1 \\ ad=n}}^n \sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right) \tag{35}$$

The set M_n of matrices with integer coefficients and determinant n can be expressed as the disjoint union

$$M_n = \bigcup_{\substack{a,d=1 \\ ad=n}}^n \bigcup_{b=0}^{d-1} \Gamma \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \tag{36}$$

and hence the sum in [35] can be viewed as a sum over the cosets in this decomposition. We note the product formula

$$T_m T_n = \sum_{d|\text{gcd}(m,n)} T_{mn/d^2} \tag{37}$$

The Hecke operators are normal, form a commuting family, and in addition they commute with the Laplacian Δ . In the following, we consider an orthonormal basis of eigenfunctions φ_j of Δ that are simultaneously eigenfunctions of all Hecke operators. We will refer to such eigenfunctions as Hecke eigenfunctions. The above assumption is automatically satisfied, if the spectrum of Δ is simple (i.e., no eigenvalues coincide), a property conjectured by Cartier and supported by numerical computations. Lindenstrauss’ work is based on the following two observations. Firstly, all quantum limits of Hecke eigenfunctions are geodesic-flow invariant measures of positive entropy, and secondly, the only such measure of positive entropy that is recurrent under Hecke correspondences is the Lebesgue measure.

The first property is proved by Bourgain and Lindenstrauss (2003) and refines arguments of Rudnick and Sarnak (1994) and Wolpert (2001) on the distribution of Hecke points (see Sarnak (2003) for

references to these papers). For a given point $z \in \mathbb{H}$ the set of Hecke points is defined as

$$T_n(z) := M_n z \tag{38}$$

For most primes, the set $T_{p^k}(z)$ comprises $(p + 1)p^{k-1}$ distinct points on $\Gamma \backslash \mathbb{H}$. For each z , the Hecke operator T_n may now be interpreted as the adjacency matrix for a finite graph embedded in $\Gamma \backslash \mathbb{H}$, whose vertices are the Hecke points $T_n(z)$. Hecke eigenfunctions φ_j with

$$T_n \varphi_j = \lambda_j(n) \varphi_j \tag{39}$$

give rise to eigenfunctions of the adjacency matrix. Exploiting this fact, Bourgain and Lindenstrauss show that for a large set of integers n

$$|\varphi_j(z)|^2 \ll \sum_{w \in T_n(z)} |\varphi_j(w)|^2 \tag{40}$$

that is, pointwise values of $|\varphi_j|^2$ cannot be substantially larger than its sum over Hecke points. This, and the observation that Hecke points for a large set of integers n are sufficiently uniformly distributed on $\Gamma \backslash \mathbb{H}$ as $n \rightarrow \infty$, yields the estimate of positive entropy with a quantitative lower bound.

Lindenstrauss’ proof of the second property, which shows that Lebesgue measure is the only quantum limit of Hecke eigenfunctions, is a result of a currently very active branch of ergodic theory: measure rigidity. Invariance under the geodesic flow alone is not sufficient to rule out other possible limit measures. In fact, there are uncountably many measures with this property. As limits of Hecke eigenfunctions, all quantum limits possess an additional property, namely recurrence under Hecke correspondences. Since the explanation of these is rather involved, let us recall an analogous result in a simpler setup. The map $\times 2 : x \mapsto 2x \pmod 1$ defines a hyperbolic dynamical system on the unit circle with a wealth of invariant measures, similar to the case of the geodesic flow on a surface of negative curvature. Furstenberg conjectured that, up to trivial invariant measures that are localized on finitely many rational points, Lebesgue measure is the only $\times 2$ -invariant measure that is also invariant under action of $\times 3 : x \mapsto 3x \pmod 1$. This fundamental problem is still unsolved and one of the central conjectures in measure rigidity. Rudolph, however, showed that Furstenberg’s conjecture is true if one restricts the statement to $\times 2$ -invariant measures of positive entropy (cf. Lindenstrauss (to appear)). In Lindenstrauss’ work, $\times 2$ plays the role of the geodesic flow, and $\times 3$ the role of the Hecke correspondences. Although here it might also be interesting to ask whether an analog of Furstenberg’s conjecture

holds, it is inessential for the proof of QUE due to the positive entropy of quantum limits discussed in the previous paragraph.

Eigenfunctions and L-Functions

An even eigenfunction $\varphi_j(z)$ for $\Gamma = \text{SL}(2, \mathbb{Z})$ has the Fourier expansion

$$\varphi_j(z) = \sum_{n=1}^{\infty} a_j(n) y^{1/2} K_{i\rho_j}(2\pi n y) \cos(2\pi n x) \tag{41}$$

We associate with $\varphi_j(z)$ the Dirichlet series

$$L(s, \varphi_j) = \sum_{n=1}^{\infty} a_j(n) n^{-s} \tag{42}$$

which converges for $\text{Re } s$ large enough. These series have an analytic continuation to the entire complex plane \mathbb{C} and satisfy a functional equation,

$$\Lambda(s, \varphi_j) = \Lambda(1 - s, \varphi_j) \tag{43}$$

where

$$\Lambda(s, \varphi_j) = \pi^{-s} \Gamma\left(\frac{s + i\rho_j}{2}\right) \Gamma\left(\frac{s - i\rho_j}{2}\right) L(s, \varphi_j) \tag{44}$$

If $\varphi_j(z)$ is in addition an eigenfunction of all Hecke operators, then the Fourier coefficients in fact coincide (up to a normalization constant) with the eigenvalues of the Hecke operators

$$a_j(m) = \lambda_j(m) a_j(1) \tag{45}$$

If we normalize $a_j(1) = 1$, the Hecke relations [37] result in an Euler product formula for the L -function,

$$L(s, \varphi_j) = \prod_{p \text{ prime}} (1 - a_j(p) p^{-s} + p^{-1-2s})^{-1} \tag{46}$$

These L -functions behave in many other ways like the Riemann zeta or classical Dirichlet L -functions. In particular, they are expected to satisfy a Riemann hypothesis, that is, all nontrivial zeros are constrained to the critical line $\text{Im } s = 1/2$.

Questions on the distribution of Hecke eigenfunctions, such as QUE or value distribution properties, can now be translated to analytic properties of L -functions. We will discuss two examples.

The asymptotics in [6] can be established by proving [6] for the choices $g = \varphi_k, k = 1, 2, \dots$, that is,

$$\int_{\mathcal{M}} |\varphi_j|^2 \varphi_k \, dA \rightarrow 0 \tag{47}$$

Watson discovered the remarkable relation (Sarnak 2003)

$$\left| \int_{\mathcal{M}} \varphi_{j_1} \varphi_{j_2} \varphi_{j_3} dA \right|^2 = \frac{\pi^4 \Lambda(\frac{1}{2}, \varphi_{j_1} \times \varphi_{j_2} \times \varphi_{j_3})}{\Lambda(1, \text{sym}^2 \varphi_{j_1}) \Lambda(1, \text{sym}^2 \varphi_{j_2}) \Lambda(1, \text{sym}^2 \varphi_{j_3})} \quad [48]$$

The L -functions $\Lambda(s, g)$ in Watson’s formula are more advanced cousins of those introduced earlier (see Sarnak (2003) for details). The Riemann hypothesis for such L -functions then implies, via [48], a precise rate of convergence to QUE for the modular surface,

$$\int_{\mathcal{M}} |\varphi_j|^2 g dA = \int_{\mathcal{M}} g dA + O(\lambda_j^{-1/4+\epsilon}) \quad [49]$$

for any $\epsilon > 0$, where the implied constant depends on ϵ and g .

A second example on the connection between statistical properties of the matrix elements $\nu_j(a) = \langle \text{Op}(a)\varphi_j, \varphi_j \rangle$ (for fixed a and random j) and values L -functions has appeared in the work of Luo and Sarnak (cf. Sarnak (2003)). Define the variance

$$V_\lambda(a) = \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\nu_j(a) - \mu(a)|^2 \quad [50]$$

with $N(\lambda) = \#\{j: \lambda_j \leq \lambda\}$; cf. [3]. Following a conjecture by Feingold–Peres and Eckhardt *et al.* (see Sarnak (2003) for references) for “generic” quantum chaotic systems, one expects a central-limit theorem for the statistical fluctuations of the $\nu_j(a)$, where the normalized variance $N(\lambda)^{1/2} V_\lambda(a)$ is asymptotic to the classical autocorrelation function $C(a)$, see eqn [54].

Conjecture 4 For any bounded function $g: \mathbb{R} \rightarrow \mathbb{C}$ we have

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} g \left(\frac{\nu_j(a) - \mu(a)}{\sqrt{V_\lambda(a)}} \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t) e^{-(1/2)t^2} dt \quad [51]$$

as $\lambda \rightarrow \infty$.

Luo and Sarnak prove that in the case of the modular surface the variance has the asymptotics

$$\lim_{\lambda \rightarrow \infty} N(\lambda)^{1/2} V_\lambda(a) = \langle Ba, a \rangle \quad [52]$$

where B is a non-negative self-adjoint operator which commutes with the Laplacian Δ and all Hecke operators T_n . In particular, we have

$$B\varphi_j = \frac{1}{2} L(\frac{1}{2}, \varphi_j) C(\varphi_j) \varphi_j \quad [53]$$

where

$$C(a) := \int_{\mathbb{R}} \int_{\Gamma \backslash \text{PSL}(2, \mathbb{R})} a(\Phi^t(g)) \overline{a(g)} d\mu(g) dt \quad [54]$$

is the classical autocorrelation function for the geodesic flow with respect to the observable a (Sarnak 2003). Up to the arithmetic factor $(1/2)L(1/2, \varphi_j)$, eqn [53] is consistent with the Feingold–Peres prediction for the variance of generic chaotic systems. Furthermore, recent estimates of moments by Rudnick and Soundararajan (2005) indicate that Conjecture 4 is not valid in the case of the modular surface.

Quantum Eigenstates of Cat Maps

Cat maps are probably the simplest area-preserving maps on a compact surface that are highly chaotic. They are defined as linear automorphisms on the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$,

$$\Phi_A : \mathbb{T}^2 \rightarrow \mathbb{T}^2 \quad [55]$$

where a point $\xi \in \mathbb{R}^2(\text{mod } \mathbb{Z}^2)$ is mapped to $A\xi(\text{mod } \mathbb{Z}^2)$; A is a fixed matrix in $\text{GL}(2, \mathbb{Z})$ with eigenvalues off the unit circle (this guarantees hyperbolicity). We view the torus \mathbb{T}^2 as a symplectic manifold, the phase space of the dynamical system. Since \mathbb{T}^2 is compact, the Hilbert space of quantum states is an N -dimensional vector space \mathcal{H}_N , N integer. The semiclassical limit, or limit of small wavelengths, corresponds here to $N \rightarrow \infty$.

It is convenient to identify \mathcal{H}_N with $L^2(\mathbb{Z}/N\mathbb{Z})$, with the inner product

$$\langle \psi_1, \psi_2 \rangle = \frac{1}{N} \sum_{Q \text{ mod } N} \psi_1(Q) \overline{\psi_2(Q)} \quad [56]$$

For any smooth function $f \in C^\infty(\mathbb{T}^2)$, define a quantum observable

$$\text{Op}_N(f) = \sum_{n \in \mathbb{Z}^2} \widehat{f}(n) T_N(n)$$

where $\widehat{f}(n)$ are the Fourier coefficients of f , and $T_N(n)$ are translation operators

$$T_N(n) = e^{\pi i n_1 n_2 / N} t_2^{n_2} t_1^{n_1} \quad [57]$$

$$\begin{aligned} [t_1 \psi](Q) &= \psi(Q + 1) \\ [t_2 \psi](Q) &= e^{2\pi i Q / N} \psi(Q) \end{aligned} \quad [58]$$

The operators $\text{Op}_N(a)$ are the analogs of the pseudodifferential operators discussed in the section “Quantum unique ergodicity.”

A quantization of Φ_A is a unitary operator $U_N(A)$ on $L^2(\mathbb{Z}/N\mathbb{Z})$ satisfying the equation

$$U_N(A)^{-1} \text{Op}_N(f) U_N(A) = \text{Op}_N(f \circ \Phi_A) \quad [59]$$

for all $f \in C^\infty(\mathbb{T}^2)$. There are explicit formulas for $U_N(A)$ when A is in the group

$$\Gamma = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{Z}) : ab \equiv cd \equiv 0 \pmod{2} \right\} \quad [60]$$

These may be viewed as analogs of the Shale–Weil or metaplectic representation for $\text{SL}(2)$. For example, the quantization of

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix} \quad [61]$$

yields

$$U_N(A)\psi(Q) = N^{-1/2} \sum_{Q' \pmod{N}} \exp \left[\frac{2\pi i}{N} (Q^2 - QQ' + Q'^2) \right] \psi(Q') \quad [62]$$

In analogy with [1], we are interested in the statistical features of the eigenvalues and eigenfunctions of $U_N(A)$, that is, the solutions to

$$U_N(A)\varphi = \lambda\varphi, \quad \|\varphi\|_{L^2(\mathbb{Z}/N\mathbb{Z})} = 1 \quad [63]$$

Unlike typical quantum-chaotic maps, the statistics of the N eigenvalues

$$\lambda_{N1}, \lambda_{N2}, \dots, \lambda_{NN} \in S^1 \quad [64]$$

do not follow the distributions of unitary random matrices in the limit $N \rightarrow \infty$, but are rather singular (Keating 1991). In analogy with the Selberg trace formula for hyperbolic surfaces [18], there is an exact trace formula relating sums over eigenvalues of $U_N(A)$ with sums over fixed points of the classical map (Keating 1991).

As in the case of arithmetic surfaces, the eigenfunctions of cat maps appear to behave more generically. The analog of the Schnirelman–Zelditch–Colin de Verdière theorem states that, for any orthonormal basis of eigenfunctions $\{\varphi_{Nj}\}_{j=1}^N$ we have, for all $f \in C^\infty(\mathbb{T}^2)$,

$$\langle \text{Op}(f)\varphi_{Nj}, \varphi_{Nj} \rangle \rightarrow \int_{\mathbb{T}^2} f(\xi) d\xi \quad [65]$$

as $N \rightarrow \infty$, for all j in an index set J_N of full density, that is, $\#J_N \sim N$. Kurlberg and Rudnick (see Rudnick (2001)) have characterized special bases of eigenfunctions $\{\varphi_{Nj}\}_{j=1}^N$ (termed Hecke eigenbases, in analogy with arithmetic surfaces) for which QUE holds, generalizing earlier work of Degli Esposti,

Graffi, and Isola (1995). That is, [65] holds for all $j=1, \dots, N$. Rudnick and Kurlberg, and more recently Gurevich and Hadani, have established results on the rate of convergence analogous to [49]. These results are unconditional. Gurevich and Hadani use methods from algebraic geometry based on those developed by Deligne in his proof of the Weil conjectures (an analog of the Riemann hypothesis for finite fields).

In the case of quantum-cat maps, there are values of N for which the number of coinciding eigenvalues can be large, a major difference to what is expected for the modular surface. Linear combinations of eigenstates with the same eigenvalue are as well eigenstates, and may lead to different quantum limits. Indeed, Faure, Nonnenmacher, and De Bièvre (see De Bièvre (to appear)) have shown that there are subsequences of values of N , so that, for all $f \in C^\infty(\mathbb{T}^2)$,

$$\langle \text{Op}(f)\varphi_{Nj}, \varphi_{Nj} \rangle \rightarrow \frac{1}{2} \int_{\mathbb{T}^2} f(\xi) d\xi + \frac{1}{2} f(0) \quad [66]$$

that is, half of the mass of the quantum limit localizes on the hyperbolic fixed point of the map. This is the first, and to date the only, rigorous result concerning the existence of scarred eigenfunctions in systems with chaotic classical limit.

Acknowledgment

The author is supported by an EPSRC Advanced Research Fellowship.

See also: Quantum Ergodicity and Mixing of Eigenfunctions; Random Matrix Theory in Physics.

Further Reading

Aurich R and Steiner F (1993) Statistical properties of highly excited quantum eigenstates of a strongly chaotic system. *Physica D* 64(1–3): 185–214.
 Berry MV and Keating JP (1999) The Riemann zeros and eigenvalue asymptotics. *SIAM Review* 41(2): 236–266.
 Bogomolny EB (2006) Quantum and arithmetical chaos. In: Cartier PE, Julia B, Moussa P, and Vanhove P (eds.) *Frontiers in Number Theory, Physics and Geometry on Random Matrices, Zeta Functions, and Dynamical Systems*, Springer Lecture Notes. Les Houches.
 Bogomolny EB, Georgeot B, Giannoni M-J, and Schmit C (1997) Arithmetical chaos. *Physics Reports* 291(5–6): 219–324.
 De Bièvre S *Recent Results on Quantum Map Eigenstates*, Proceedings of QMATH9, Giens 2004 (to appear).
 Hejhal DA and Rackner BN (1992) On the topography of Maass waveforms for $\text{PSL}(2, \mathbb{Z})$. *Experiment. Math.* 1(4): 275–305.
 Keating JP (1991) The cat maps: quantum mechanics and classical motion. *Nonlinearity* 4(2): 309–341.

- Lindenstrauss E Rigidity of multi-parameter actions. *Israel Journal of Mathematics* (Furstenberg Special Volume) (to appear).
- Marklof J (2006) Energy level statistics, lattice point problems and almost modular functions. In: Cartier PE, Julia B, Moussa P, and Vanhove P (eds.) *Frontiers in Number Theory, Physics and Geometry on Random Matrices, Zeta Functions, and Dynamical Systems*, Springer Lecture Notes. Les Houches.
- Rudnick Z (2001) On quantum unique ergodicity for linear maps of the torus. In: *European Congress of Mathematics*, (Barcelona, 2000), Progr. Math., vol. 202, pp. 429–437. Basel: Birkhäuser.
- Rudnick Z (2005) A central limit theorem for the spectrum of the modular group, Park city lectures. *Annales Henri Poincaré* 6: 863–883.
- Sarnak P Arithmetic quantum chaos. The Schur lectures (1992) (Tel Aviv), Israel Math. Conf. Proc., 8, pp. 183–236. Bar-Ilan Univ., Ramat Gan, 1995.
- Sarnak P (2003) Spectra of hyperbolic surfaces. *Bulletin of the American Mathematical Society (N.S.)* 40(4): 441–478.

Asymptotic Structure and Conformal Infinity

J Frauendiener, Universität Tübingen, Tübingen, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

A major motivation for studying the asymptotic structure of spacetimes has been the need for a rigorous description of what should be understood by an “isolated system” in Einstein’s theory of gravity. As an example, consider a gravitating system somewhere in our universe (e.g., a galaxy, a cluster of galaxies, a binary system, or a star) evolving according to its own gravitational interaction, and possibly reacting to gravitational radiation impinging on it from the outside. Thereby it will emit gravitational radiation. We are interested in describing these waves because they provide us with important information about the physics governing the system.

To adequately describe this situation, we need to idealize the real situation in an appropriate way, since it is hopeless to try to analyze the behavior of the system in its interaction with the rest of the universe. We are mainly interested in the behavior of the system, and not so much in other processes taking place at large distances from the system. Since we would like to ignore those regions, we need a way to isolate the system from their influence.

The notion of an isolated system allows us to select individual subsystems of the universe and describe their properties regardless of the rest of the universe so that we can assign to each subsystem such physical attributes as its energy–momentum, angular momentum, or its emitted radiation field. Without this notion, we would always have to take into account the interaction of the system with its environment in full detail.

In general relativity (GR) it turns out to be a rather difficult task to describe an isolated system and the reason is – as always in Einstein’s theory – the fact that the metric acts both as the physical field and as

the background. In other theories, like electrodynamics, the physical field, such as the Maxwell field, is very different from the background field, the flat metric of Minkowski space. The fact that the metric in GR plays a dual role makes it difficult to extract physical meaning from the metric because there is no nondynamical reference point.

Imagine a system alone in the universe. As we recede from the system we would expect its influence to decrease. So we expect that the spacetime which models this situation mathematically will resemble the flat Minkowski spacetime and it will approximate it even better the farther away we go. This implies that one needs to impose fall-off conditions for the curvature and that the manifold will be asymptotically flat in an appropriate sense. However, there is the problem that fall-off conditions necessarily imply the use of coordinates and it is awkward to decide which coordinates should be “good ones.” Thus, it is not clear whether the notion of an asymptotically flat spacetime is an invariant concept.

What is needed, therefore, is an invariant definition of asymptotically flat spacetimes. The key observation in this context is that “infinity” is far away with respect to the spacetime metric. This means that geodesics heading away from the system should be able to “run forever,” that is, be defined for arbitrary values of their affine parameter s . “Infinity” will be reached for $s \rightarrow \infty$. However, suppose we do not use the spacetime metric g but a metric \hat{g} which is scaled down with respect to g , that is, in such a way that $\hat{g} = \Omega^2 g$ for some function Ω . Then it might be possible to arrange Ω in such a way that geodesics for the metric \hat{g} cover the same events (strictly speaking, this holds only for null geodesics, but this is irrelevant for the present plausibility argument) as those for the metric g yet that their affine parameter \hat{s} (which is also scaled down with respect to s) approaches a finite value \hat{s}_0 for $s \rightarrow \infty$. Then we could attach a boundary to the spacetime manifold consisting of all the limit points corresponding to the events with $\hat{s} = \hat{s}_0$ on the \hat{g} -geodesics.

This boundary would have to be interpreted as “infinity” for the spacetime because it takes infinitely long for the g -geodesics to get there.

We arrived at this idea of attaching a boundary by considering the metric structure only “up to arbitrary scaling,” that is, by looking at metrics which differ only by a factor. This is the conformal structure of the spacetime manifold in question. By considering the spacetime only from the point of view of its conformal structure we obtain a picture of the spacetime which is essentially finite but which leaves its causal properties unchanged, and hence in particular the properties of wave propagation. This is exactly what is needed for a rigorous treatment of radiation emitted by the system.

Infinity for Minkowski Spacetime

The above discussion suggests that we should consider the spacetime metric only up to scale, that is, to focus on the conformal structure of the spacetime in question. Since we are interested in systems which approach Minkowski spacetime at large distances from the source, it is illuminating to study Minkowski spacetime as a preliminary example. So consider the manifold $\mathbb{M} = \mathbb{R}^4$ equipped with the flat metric

$$g = dt^2 - dr^2 - r^2 d\sigma^2 \quad [1]$$

where r is the standard radial coordinate defined by $r^2 = x^2 + y^2 + z^2$ and

$$d\sigma^2 = d\theta^2 + \sin^2 \theta d\phi^2$$

is the standard metric on the unit sphere S^2 . We now introduce retarded and advanced time coordinates, which are adapted to the null cone and hence to the conformal structure of g by the definition

$$u = t - r, \quad v = t + r$$

and obtain the metric in the form

$$g = du dv - \frac{1}{4}(v - u)^2 d\sigma^2$$

The coordinates u and v both take arbitrary real values but they are restricted by the relation $v - u = 2r \geq 0$. In order to see what happens “at infinity,” we introduce the coordinates U and V by the relations

$$u = \tan U, \quad v = \tan V$$

Then U and V both take values in the open interval $(-\pi/2, \pi/2)$ with $V \geq U$ and the metric is transformed to

$$g = \frac{1}{4 \cos^2 U \cos^2 V} [4dU dV - \sin^2(V - U) d\sigma^2] \quad [2]$$

Clearly, the metric is undefined at events with $\cos U = 0$ or $\cos V = 0$. These would correspond to events with $u = \pm\infty$ or $v = \pm\infty$ which do not lie in \mathbb{M} . However, by defining the function

$$\Omega = 2 \cos U \cos V$$

we find that the metric $\hat{g} = \Omega^2 g$ with

$$\hat{g} = 4dU dV - \sin^2(V - U) d\sigma^2 \quad [3]$$

is conformally equivalent to g and is regular for all values of U and V (keeping $V \geq U$). In fact, by defining the coordinates

$$T = V + U, \quad R = V - U$$

this metric takes the form

$$\hat{g} = dT^2 - dR^2 - \sin^2 R d\sigma^2 \quad [4]$$

the metric of the static Einstein universe \mathbb{E} . Thus, we may regard the Minkowski spacetime as the part of the Einstein cylinder defined by restricting the coordinates T and R to the region $|T| + R < \pi$ as illustrated in **Figure 1**. Although \mathbb{M} can be considered as being diffeomorphic to the shaded part in **Figure 1**, these two manifolds are not isometric. This is obvious from considering the properties of the events lying on

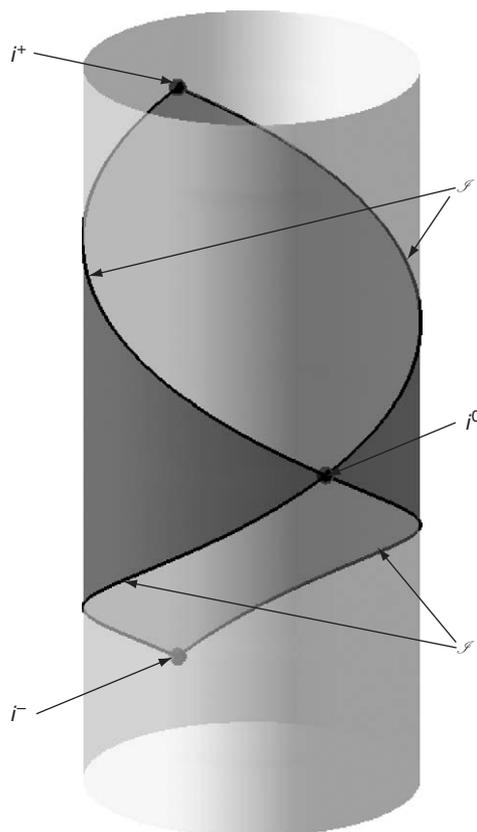


Figure 1 The embedding of Minkowski spacetime into the Einstein cylinder.

the boundary $\partial\mathbb{M}$ of \mathbb{M} in \mathbb{E} . Fix a point P inside \mathbb{M} and follow a null geodesic with respect to the metric \hat{g} from P toward the future. It will intersect $\partial\mathbb{M}$ after a finite amount of its affine parameter has elapsed. When we follow a null geodesic with respect to g from P in the same direction, we find that it does not reach $\partial\mathbb{M}$ for any value of its affine parameter. Thus, the boundary is at infinity for the metric g but at a finite location with respect to the metric \hat{g} . When we consider all possible kinds of geodesics for the metric g we find that $\partial\mathbb{M}$ consists of five qualitatively different pieces. The future pointing timelike geodesics all approach the point i^+ given by $(T, R) = (\pi, 0)$, while the past-pointing geodesics approach i^- with coordinates $(-\pi, 0)$. All spacelike geodesics come arbitrarily close to a point i^0 with coordinates $(0, \pi)$ (located on the front of the cylinder in **Figure 1**). Null geodesics, however, are different. For any point $(T, \pi - |T|)$ with $T \neq 0, \pm\pi$ on $\partial\mathbb{M}$ there are g -null-geodesics which come arbitrarily close.

In this sense, we may regard $\partial\mathbb{M}$ as consisting of limit points obtained by tracing-geodesics for infinite values of their affine parameters. According to the causal character of the geodesics the set of their respective limit points is called future/past timelike infinity i^\pm , spacelike infinity i^0 or future/past null-infinity, denoted by \mathcal{I}^\pm . These two parts of null-infinity are three-dimensional regular submanifolds of the embedding manifold \mathbb{E} , while the points i^\pm, i^0 are regular points in \mathbb{E} in the sense that the metric \hat{g} is regular there. This is not automatic, considering the fact that infinitely many geodesics converge to a single point. However, the flatness of Minkowski spacetime guarantees that the geodesics approach at just the appropriate rate for the limit points to be regular.

This example shows that the structure of the boundary is determined entirely by the metric g of Minkowski spacetime. If we had chosen a different function $\Omega' = \omega\Omega$ with $\omega > 0$ then we would not have obtained the Einstein cylinder but some different Lorentzian manifold (\mathcal{M}', g') . Yet, the boundary of \mathbb{M} in \mathcal{M}' would have had the same properties.

Asymptotically Flat Spacetimes

The physical idea of an isolated system is captured mathematically by an asymptotically flat spacetime. Since such a spacetime \mathcal{M} is expected to approach Minkowski spacetime asymptotically, the asymptotic structure of \mathcal{M} is also expected to be similar to that of \mathbb{M} . This expectation is expressed in

Definition 1 A spacetime (\mathcal{M}, g_{ab}) is called “asymptotically simple” if there exists a manifold-with-boundary $\widehat{\mathcal{M}}$ with metric \hat{g}_{ab} and scalar field Ω on $\widehat{\mathcal{M}}$ and boundary $\mathcal{I} = \partial\mathcal{M}$ such that the following conditions hold:

1. \mathcal{M} is the interior of $\widehat{\mathcal{M}}$: $\mathcal{M} = \text{int } \widehat{\mathcal{M}}$;
2. $\hat{g}_{ab} = \Omega^2 g_{ab}$ on \mathcal{M} ;
3. Ω and \hat{g}_{ab} are smooth on all of $\widehat{\mathcal{M}}$;
4. $\Omega > 0$ on \mathcal{M} ; $\Omega = 0, \nabla_a \Omega \neq 0$ on \mathcal{I} ; and
5. each null geodesic acquires both future and past endpoints on \mathcal{I} .

This definition formalizes the construction which was explicitly performed above, by which one attaches a regular (nonempty) boundary to a spacetime after suitably rescaling its metric. Asymptotically simple spacetimes are exactly those for which this process of conformal compactification is possible. The purpose of condition 5 is to exclude pathological cases. There are spacetimes which do not satisfy this condition (e.g., the Schwarzschild spacetime, where some of the null geodesics enter the event horizon and cannot escape to infinity). Yet, one would like to include them as being asymptotically simple in a sense, because they clearly describe isolated systems. For these cases, there exists the notion of weakly asymptotically simple spacetimes.

In order to arrive at asymptotically flat spacetimes, one needs to make certain assumptions about the behavior of the curvature near the boundary, thus:

Definition 2 An asymptotically simple spacetime is called “asymptotically flat” if its Ricci tensor $\text{Ric}[g]$ vanishes in a neighborhood of \mathcal{I} .

Note that this definition imposes a rather strong restriction on the Ricci curvature; less restrictive assumptions are possible. This condition applies only near \mathcal{I} . Thus, it is possible to consider spacetimes which contain matter fields as long as these fields do not extend to infinity.

Other asymptotically simple spacetimes which are not asymptotically flat are the de Sitter and anti-de Sitter spacetimes which are solutions of the Einstein equations with nonvanishing cosmological constant λ . It is a simple consequence of the definition that the boundary \mathcal{I} is a regular three-dimensional hypersurface of the embedding spacetime $\widehat{\mathcal{M}}$ which is timelike, spacelike, or null depending on the sign of λ . In particular, for the Minkowski spacetime ($\lambda = 0$) the boundary is necessarily a null hypersurface, as noted above.

The requirement that the vacuum Einstein equations hold near \mathcal{I} has several important

consequences. First, \mathcal{S} is a null hypersurface with the special property of being shear-free. This means that any cross section of a bundle of its null generators does not suffer any distortions when moved along the generators. Only expansion or contraction can occur. The global structure of \mathcal{S} is the same as the one from the example above. Null infinity consists of two connected components, \mathcal{S}^\pm , each of which is diffeomorphic to $S^2 \times \mathbb{R}$. Thus, topologically, \mathcal{S}^\pm are cylinders. The cone-like appearance as seen in [Figure 1](#) is artificial. It depends on the particular conformal factor Ω chosen for the conformal compactification. Furthermore, it is only in very exceptional cases that the metric \hat{g} is regular at i^0 or i^\pm .

The most important consequence, however, concerns the conformal Weyl tensor $C^a{}_{bcd}$. This is the part of the full Riemann curvature tensor $R^a{}_{bcd}$ which is trace-free. It is invariant under conformal rescalings of the metric. Thus, on \mathcal{M} , $C^a{}_{bcd} = \hat{C}^a{}_{bcd}$. When the vanishing of the Ricci tensor near \mathcal{S} is assumed then it turns out that the Weyl tensor necessarily vanishes on \mathcal{S} . This is the ultimate justification for calling such manifolds asymptotically flat because the entire curvature vanishes on \mathcal{S} .

Some Consequences

There are several consequences of the existence of the conformal boundary \mathcal{S} . They all can be traced back to the fact that this boundary can be used to separate the geometric fields into a universal background field and dynamical fields which propagate on it. The background is given by the boundary points attached to an asymptotically flat spacetime which always form a three-dimensional null hypersurface \mathcal{S} with two connected components (in the sequel, we restrict our attention to \mathcal{S}^+ only; \mathcal{S}^- is treated similarly), each with the topology of a cylinder. And in each case, \mathcal{S} is shear-free.

The BMS Group

Since the structure of null-infinity is universal over all asymptotically flat spacetimes, it is obvious that its symmetry group should also possess a universal meaning. This group, the so-called Bondi–Metzner–Sachs (BMS) group is in many respects similar to the Poincaré group, the symmetry group of \mathbb{M} . It is the semidirect product of the Lorentz group with an abelian group which, however, is not the four-dimensional translation group but an infinite-dimensional group of supertranslations. This group is a normal subgroup, so the factor group is isomorphic to the Lorentz group.

In physical terms, the supertranslations arise because there are infinitely many directions from which observers at infinity (whose world lines coincide with the null generators of \mathcal{S} in a certain limit) can observe the system and because each observer is free to choose its own origin of proper time u . The observers surrounding the system are not synchronized, because under the assumptions made there is no natural way to fix a unique common origin. Hence, a supertranslation is a shift of the parameter along each null generator of \mathcal{S}^+ corresponding to a change of origin for each individual observer. It can be given as a map $S^2 \rightarrow \mathbb{R}$. A choice of origin on each null generator of \mathcal{S}^+ is referred to as a “cut” of \mathcal{S}^+ . It is a two-dimensional surface of spherical topology which intersects each null generator exactly once. It is an open question whether one can always synchronize the observers by imposing canonical conditions at i^0 or i^\pm , thereby reducing the BMS group to the smaller Poincaré group.

The supertranslations contain a unique four-dimensional normal subgroup. In \mathbb{M} these special supertranslations are the ones which are induced by the translations of Minkowski spacetime in the following way. Take the future light cone of some event P and follow it out to \mathcal{S}^+ , where its intersection defines an origin for each observer located there. Now consider the light cone of another event Q obtained from P by a translation in a spatial direction. Then the light emitted from Q will arrive at \mathcal{S}^+ earlier than that from P for observers in the direction of the translation, while it will be delayed for observers in the opposite direction. This change in arrival time defines a specific supertranslation. Similarly, for a translation in a temporal direction, the light from Q will arrive later than that from P for all observers. Thus, every translation in \mathbb{M} defines a particular supertranslation on \mathcal{S}^+ . These can be characterized in a different way, which is intrinsic to \mathcal{S}^+ and which can be used in the general case even though there will be no Killing vectors present in a general asymptotically flat spacetime. In an appropriate coordinate system, the asymptotic translations are given as linear combinations of the first four spherical harmonics $Y_{00}, Y_{10}, Y_{1\pm 1}$. The space of asymptotic translations \mathbb{T} is in a natural way isometric to \mathbb{M} .

The Peeling Property

Now consider the Weyl tensor $C^a{}_{bcd}$ on $\widehat{\mathcal{M}}$. Since it vanishes on \mathcal{S} where $\Omega = 0$ we may form the quotient

$$K^a{}_{bcd} = \Omega^{-1} C^a{}_{bcd}$$

which can be shown to be smooth on \mathcal{S}^+ . The physical interpretation of this tensor field is based on the following properties. In source-free regions the field satisfies the spin-2 zero-rest-mass equation

$$\widehat{\nabla}_a K^a{}_{bcd} = 0$$

which is very similar to the Maxwell equations for the electromagnetic (spin-1) Faraday tensor. Thus, $K^a{}_{bcd}$ is interpreted as the gravitational field, which describes the gravitational waves contained inside the system. The zero-rest-mass equation for $K^a{}_{bcd}$ and the fact that the field is smooth on \mathcal{S} implies that the Weyl tensor satisfies the “peeling” property. This is a characteristic conspiracy between the fall-off behavior of certain components of the Weyl tensor along outgoing g -null-geodesics approaching \mathcal{S}^+ in \mathcal{M} with respect to an affine parameter s for $s \rightarrow \infty$ and their algebraic type. Symbolically, the Weyl tensor has the following behavior as $s \rightarrow \infty$ along the null geodesic:

$$C = \frac{[4]}{s} + \frac{[31]}{s^2} + \frac{[211]}{s^3} + \frac{[1111]}{s^4} + O(s^{-5}) \quad [5]$$

where the numerator of each component indicates its Petrov type. The repeated principal null direction (PND) in the first three components and one of the PNDs in the fourth component are aligned with the tangent vector of the geodesic. This implies that the farthest reaching component of the Weyl tensor, which is $O(1/s)$, has the Petrov type of a radiation field. It is customary to combine the components which are $O(1/s^i)$ into one complex function and denote it by ψ_{5-i} . When expressed in terms of the field $K^a{}_{bcd}$ on \mathcal{M} , this fall-off behavior implies that of all components of $K^a{}_{bcd}$ only ψ_4 does not necessarily vanish on \mathcal{S}^+ .

In special cases like the Minkowski, Schwarzschild, Kerr, and more generally in all asymptotically flat stationary spacetimes, even ψ_4 vanishes on \mathcal{S}^+ . For these reasons, ψ_4 is called the radiation field of the system, that is, that part of the gravitational field which can be registered by the observers at infinity. It describes the outgoing radiation which is being emitted by the system during its evolution.

The Bondi–Sachs Mass-Loss Formula

Gravitational waves carry away energy from the system. This is a consequence of the Bondi–Sachs mass-loss formula. The Bondi–Sachs energy-momentum is related to a weighted integral over a cut \mathcal{C} ,

$$P_{\mathcal{C}}[W] = -\frac{1}{4\pi G} \int_{\mathcal{C}} W[\psi_2 + \sigma\dot{\sigma}] d^2S \quad [6]$$

The quantity in brackets, the mass aspect, is a combination of the scalar ψ_2 which in a sense measures the strength of the Coulomb-like part of the gravitational field on \mathcal{S}^+ and the complex quantity σ . In a so-called Bondi coordinate system, this quantity is related to the radiation field ψ_4 by the relation

$$\psi_4 = -\ddot{\sigma}$$

the dot indicating differentiation with respect to the affine parameter along the null generators. Thus, σ is essentially the second time integral of the radiation field. The mass aspect is integrated against a function W which is an asymptotic translation, that is, a linear combination of the first four spherical harmonics. Thus, one can view the expression [6] as defining a linear map $\mathbb{T} \rightarrow \mathbb{R}$. Since \mathbb{T} and \mathbb{M} are isometric this defines a covector P_a on \mathbb{M} , which can always be shown to be timelike, $P_a P^a \geq 0$. This positivity property together with the fact that in the special cases of Schwarzschild and Kerr spacetimes the integral yields the mass parameters when evaluated for a time translation ($W=1$) motivates the interpretation of $P_{\mathcal{C}}$ as the energy–momentum 4-vector of the spacetime at the instant defined by the cut \mathcal{C} . In particular, for $W=1$ the integral gives the time component of $P_{\mathcal{C}}$, the Bondi–Sachs energy E .

The interpretation of [6] as energy–momentum is strengthened by the fact that $P_{\mathcal{C}}$ arises as dual to the translations which is familiar from Lagrangian field theories where energy and momentum appear as generators for time and space translations. In fact, one can set up a Hamiltonian framework where the role of the Bondi–Sachs energy–momentum as generator of asymptotic translations is made explicit.

This point of view suggests that one should also be able to define a notion of angular momentum for asymptotically flat spacetimes because angular momentum arises as the generator of rotations, which can also be defined asymptotically. However, while there is a unique notion of translation on \mathcal{S}^+ , this is not the case for rotations (and boosts). The reason is hidden in the structure of the BMS group where the Lorentz group appears naturally as a factor group but not as a unique subgroup. In physical terms, the angular momentum depends on an origin but there is no natural way to choose an origin on \mathcal{S}^+ . This ambiguity in the choice of origin leads to several nonequivalent expressions for angular momentum in the literature.

Consider now two cuts \mathcal{C} and \mathcal{C}' , with \mathcal{C}' later than \mathcal{C} . Then we may compute the difference $\Delta E = E - E'$ of the Bondi–Sachs energies with respect to the two

cuts. It turns out that this difference can be expressed as an integral over the (three-dimensional) piece Σ of \mathcal{S}^+ which is bounded by the two cuts (i.e., $\partial\Sigma = C' - C$):

$$E' - E = -\frac{1}{4\pi G} \int_{\Sigma} \dot{\sigma} \dot{\bar{\sigma}} d^3V \quad [7]$$

This result means that the Bondi–Sachs energy of the system decreases, since $E' < E$ and the rate of decrease is given by the (positive-definite) amount of gravitational radiation which leaves the system during the period defined by the two cuts.

It is necessary to point out that in this article the structure of null infinity has been postulated based on physical reasonings. The Einstein equations have been used only in a very weak sense, namely only in a neighborhood of \mathcal{S} . It is an entirely different question whether the field equations are compatible with this postulated structure. To answer it, one needs to show that there are global solutions of the Einstein equations which exhibit the postulated behavior in the asymptotic region. This question has been settled recently in the affirmative: there are many global spacetimes which are asymptotically flat in the sense described here.

This article discussed has the notion of null infinity, that is, of spacetimes which are asymptotically flat in lightlike directions. Spacetimes which are asymptotically flat in spacelike directions have not been covered. The latter is a notion which has been developed largely independently of null infinity since it is essentially a property of an initial data set and not of the entire four-dimensional spacetime. Ultimately, these two notions should coincide, in the sense that if one has an initial data set which is asymptotically flat in spatial directions in an appropriate sense then its Cauchy development will be an asymptotically flat spacetime. However, as of yet, it is not clear what the appropriate conditions should be because the structure of the gravitational field in

the neighborhood of spacelike infinity i^0 is not sufficiently well understood so far.

See also: Black Hole Mechanics; Boundaries for Spacetimes; Canonical General Relativity; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; General Relativity: Overview; Gravitational Waves; Quantum Entropy; Spacetime Topology, Causal Structure and Singularities; Stability of Minkowski Space; Stationary Black Holes.

Further Reading

- Ashtekar A (1987) *Asymptotic Quantization*. Naples: Bibliopolis.
- Bondi H, van der Burg MGJ, and Metzner AWK (1962) Gravitational waves in general relativity VII. Waves from axi-symmetric isolated systems. *Proceedings of the Royal Society of London, Series A* 269: 21–52.
- Frauenhauer J (2004) Conformal infinity. *Living Reviews in Relativity*, vol. 3. <http://relativity.livingreviews.org/Articles/lrr-2004-1/index.html>.
- Friedrich H (1992) Asymptotic structure of space-time. In: Janis AI and Porter JR (eds.) *Recent Advances in General Relativity*. Boston: Birkhäuser.
- Friedrich H (1998a) Einstein's equation and conformal structure. In: Huggett SA, Mason LJ, Tod KP, Tsou SS, and Woodhouse NMJ (eds.) *The Geometric Universe: Science, Geometry and the Work of Roger Penrose*. Oxford: Oxford University Press.
- Friedrich H (1998b) Gravitational fields near space-like and null infinity. *Journal of Geometry and Physics* 24: 83–163.
- Geroch R (1977) Asymptotic structure of space-time. In: Esposito FP and Witten L (eds.) *Asymptotic Structure of Space-Time*. New York: Plenum.
- Hawking S and Ellis GFR (1973) *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
- Penrose R (1965) Zero rest-mass fields including gravitation: asymptotic behaviour. *Proceedings of the Royal Society of London, Series A* 284: 159–203.
- Penrose R (1968) Structure of space-time. In: DeWitt CM and Wheeler JA (eds.) *Battelle Rencontres*. New York: W. A. Benjamin.
- Penrose R and Rindler W (1984, 1986) *Spinors and Space-Time*, Cambridge: Cambridge University Press.
- Sachs RK (1962) Gravitational waves in general relativity VIII. Waves in asymptotically flat space-time. *Proceedings of the Royal Society of London, Series A* 270: 103–127.

Averaging Methods

A I Neishtadt, Russian Academy of Sciences, Moscow, Russia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Averaging methods are the methods of perturbation theory that are based on the averaging principle and the idea of dividing the dynamics into slow drift and

fast oscillations. The most common field of applications of averaging methods is the analysis of the behavior of dynamical systems that differ from integrable systems by small perturbations.

Averaging Principle

Equations of motion of a system that differ from an integrable system by small perturbations often can be written in the form

$$\begin{aligned} \dot{I} &= \varepsilon g(I, \varphi, \varepsilon), \quad \dot{\varphi} = \omega(I) + \varepsilon f(I, \varphi, \varepsilon) \\ I &= (I_1, \dots, I_n) \in \mathbb{R}^n \\ \varphi &= (\varphi_1, \dots, \varphi_m) \in \mathbb{T}^m \text{ modd } 2\pi, 0 < \varepsilon \ll 1 \end{aligned} \quad [1]$$

The small parameter ε characterizes the amplitude of the perturbation. For $\varepsilon=0$ one gets the unperturbed system. The equation $I=\text{const.}$ singles out an invariant m -dimensional torus of the unperturbed system. The motion on this torus is quasiperiodic with frequency vector $\omega(I)$; components of vector I are called “slow variables” whereas components of vector φ are called “fast variables” or “phases.” The right-hand sides of system [1] are 2π -periodic with respect to all φ_j . It is assumed that they are smooth enough functions of all arguments. It is also assumed that components of the frequency vector are not linearly dependent over the ring of integer numbers identically with respect to I . System [1] is called a “system with rotating phases.”

In applications, one is often interested mainly in the behavior of slow variables. The “averaging principle” (or method) consists in replacing the system of perturbed equations [1] by the “averaged system”

$$\dot{J} = \varepsilon G(J), \quad G(J) = (2\pi)^{-m} \oint_{\mathbb{T}^m} g(J, \varphi, 0) d\varphi \quad [2]$$

for the purpose of providing an approximate description of the evolution of the slow variables over time intervals of order $1/\varepsilon$ or longer. Here, $d\varphi = d\varphi_1 \cdots d\varphi_m$. System [2] contains only slow variables and, therefore, is much simpler for investigation than system [1]. When passing from system [1] to system [2], one ignores the terms $g(I, \varphi, 0) - G(I)$ on the right-hand side of [1]. The averaging principle is based on the idea that these terms oscillate and lead only to small oscillations which are superimposed on the drift described by the averaged system. To justify the averaging principle, one should establish a relation between the behavior of the solutions of systems [1] and [2]. This problem is still far from being completely solved.

Another version of the averaging principles is used in the case when frequencies are approximately in resonance. This means that one or several relations of the form $(k, \omega) = 0$ approximately are valid with irreducible integer coefficient vectors $k \neq 0$; here, (k, ω) is the standard scalar product in \mathbb{R}^m . Let Γ be a sublattice of the integer lattice \mathbb{Z}^m generated by these vectors. Let $r = \text{rank } \Gamma$ and $k^{(1)}, k^{(2)}, \dots, k^{(m)}$ be a basis in \mathbb{Z}^m ,

the first r vectors of which belong to Γ . Instead of φ , one can introduce new variables:

$$\begin{aligned} \vartheta &= (\vartheta_1, \dots, \vartheta_r) \in \mathbb{T}^r \text{ modd } 2\pi \\ \chi &= (\chi_1, \dots, \chi_{m-r}) \in \mathbb{T}^{m-r} \text{ modd } 2\pi \\ \vartheta_i &= (k^{(i)}, \varphi), \quad \chi_j = (k^{(r+j)}, \varphi) \end{aligned}$$

Let R be an $r \times m$ matrix whose rows are vectors $k^{(i)}, 1 \leq i \leq r$. For an approximate description of the behavior of variables I, ϑ , the averaging principle prescribes replacing system [1] by the system

$$\begin{aligned} \dot{J} &= \varepsilon G_\Gamma(J, \gamma), \quad \dot{\gamma} = R\omega(J) + \varepsilon R F_\Gamma(J, \gamma) \\ G_\Gamma(J, \vartheta) &= (2\pi)^{-(m-r)} \oint_{\mathbb{T}^{m-r}} g(J, \varphi, 0) d\chi \\ F_\Gamma(J, \vartheta) &= (2\pi)^{-(m-r)} \oint_{\mathbb{T}^{m-r}} f(J, \varphi, 0) d\chi \end{aligned} \quad [3]$$

(one should express g, f through ϑ, χ and then integrate over $\chi, d\chi = d\chi_1 \cdots d\chi_{m-r}$). System [3] is called “partially averaged system” for resonances in Γ . Functions G_Γ, F_Γ can be obtained from Fourier series expansions of functions g, f for $\varepsilon=0$ by throwing away harmonics $\exp(i(k, \varphi)), k \notin \Gamma$ (nonresonant harmonics). Passing from system [1] to system [3] is based on the idea that the ignored nonresonant harmonics oscillate fast and do not affect essentially the evolution of the slow variables.

Now let system [1] be a Hamiltonian system close to an integrable one. The Hamiltonian function has the form

$$H = H_0(p) + \varepsilon H_1(p, \varphi, y, x, \varepsilon)$$

where φ, x are coordinates and p, y are conjugated to them. The equations of motion have the same form as [1], with $I = (p, y, x)$:

$$\begin{aligned} \dot{p} &= -\varepsilon \frac{\partial H_1}{\partial \varphi}, \quad \dot{y} = -\varepsilon \frac{\partial H_1}{\partial x} \\ \dot{x} &= \varepsilon \frac{\partial H_1}{\partial y}, \quad \dot{\varphi} = \frac{\partial H_0}{\partial I} + \varepsilon \frac{\partial H_1}{\partial I} \end{aligned} \quad [4]$$

The averaging principle in the case when there are no resonant relations leads to the system

$$\begin{aligned} \dot{p} &= 0, \quad \dot{y} = -\varepsilon \frac{\partial \mathcal{H}_1}{\partial x}, \quad \dot{x} = \varepsilon \frac{\partial \mathcal{H}_1}{\partial y} \\ \mathcal{H}_1 &= (2\pi)^{-m} \oint_{\mathbb{T}^m} H_1(p, \varphi, y, x, 0) d\varphi \end{aligned} \quad [5]$$

Therefore, in this case there is no drift in p , and the behavior of y, x is described by the Hamiltonian system, which contains p as a parameter. Equations of motion of planets around the Sun can be reduced to the form [4]. The issue of the absence of the evolution of momenta p is known in this problem as

the Lagrange–Laplace theorem, about the absence of the evolution of semimajor axes of planetary orbits.

Elimination of Fast Variables, Decoupling of Slow and Fast Motions

The basic role in the averaging method is played by the idea that the exact system can be in the principal approximation transformed into the averaged system by means of a transformation of variables close to the identical one. The extension of this idea is the idea that similar transformation of variables allows one to eliminate, up to an arbitrary degree of accuracy, the fast phases from the right-hand sides of the equations of perturbed motion and in this way decouple the slow motion from the fast one. For system [1], provided there are no resonant relations between frequencies, the elimination of fast variables is performed as follows. The desirable transformation of variables $(I, \varphi) \mapsto (J, \psi)$ is sought as a formal series

$$\begin{aligned} I &= J + \varepsilon u_1(J, \psi) + \varepsilon^2 u_2(J, \psi) + \dots \\ \varphi &= \psi + \varepsilon v_1(J, \psi) + \varepsilon^2 v_2(J, \psi) + \dots \end{aligned} \quad [6]$$

where functions u_j, v_j are 2π -periodic in ψ . The transformation [6] should be chosen in such a way that in the new variables the right-hand sides of equations of motion do not contain fast variables, that is, the equations of motion should have the form

$$\begin{aligned} \dot{J} &= \varepsilon G_0(J) + \varepsilon^2 G_1(J) + \dots \\ \dot{\psi} &= \omega(J) + \varepsilon F_0(J) + \varepsilon^2 F_1(J) + \dots \end{aligned} \quad [7]$$

Substituting [6] into [7], taking into account [1], and equating the terms of the same order in ε , we obtain the following set of relations:

$$\begin{aligned} G_0(J) &= g(J, \psi, 0) - \frac{\partial u_1}{\partial \psi} \omega \\ F_0(J) &= f(J, \psi, 0) + \frac{\partial \omega}{\partial J} u_1 - \frac{\partial v_1}{\partial \psi} \omega \\ G_i(J) &= X_i(J, \psi) - \frac{\partial u_{i+1}}{\partial \psi} \omega \\ F_i(J) &= Y_i(J, \psi) + \frac{\partial \omega}{\partial J} u_{i+1} - \frac{\partial v_{i+1}}{\partial \psi} \omega, \quad i \geq 1 \end{aligned} \quad [8]$$

The functions X_i, Y_i are uniquely determined by the terms $u_1, v_1, \dots, u_i, v_i$ in expansion [6]. The first equation in [8] implies that

$$\begin{aligned} G_0(J) &= g_0(J) = G(J) \\ u_1(J, \psi) &= \sum_{k \neq 0} \frac{g_k}{i(k, \omega)} \exp(i(k, \psi)) + u_1^0(J) \end{aligned} \quad [9]$$

where $g_k, k \in \mathbb{Z}^m$, are Fourier coefficients of function g at $\varepsilon = 0$, and u_1^0 is an arbitrary function of J . It is assumed that the denominators in [9] do not vanish, and that the series in [9] converges and determines a smooth function. In the same way, from the other equations in [8] one can sequentially determine $F_0, v_1, \dots, G_i, u_{i+1}, F_i, v_{i+1}, i \geq 1$.

On truncating the series in [6] and [7] at the terms of order ε^l , we obtain a truncated system of the l th approximation. The equation for J is decoupled from the other equations and can be solved separately. Then the behavior of ψ is determined by means of quadrature. The behavior of original variable I in this approximation is a slow drift (described by the equation for J), on which small oscillations (described by transformation of variables) are superimposed. The behavior of φ can be represented as a rotation with slowly varying frequency, on which oscillations are also superimposed. For $l = 1$, the truncated system coincides with the averaged system [2].

If the sublattice $\Gamma \subset \mathbb{Z}^m$ specifying possible resonant relations is given, then in an analogous manner one can construct a formal transformation of variables $(I, \varphi) \mapsto (J, \psi)$ such that, in the new variables, the fast phase ψ will appear on the right-hand sides of the differential equations for the new variables only in combinations (k, ψ) , with $k \in \Gamma$ (see, e.g., Arnol'd *et al.* (1988)). Again, on truncating the series on the right-hand sides of the differential equations for the new variables at the terms of order ε^l , we obtain a truncated system of the l th approximation. At $l = 1$, this truncated system coincides with the partially averaged system [3] (for some special choice of arbitrary functions that are contained in the formulas for transformation of variables). If the original system is a Hamiltonian system of the form [4], then the transformation of variables eliminating the fast phases from the right-hand sides of the differential equations can be chosen to be symplectic. The corresponding procedures are called “Lindstedt method” and “Newcomb method” (nonresonant case for $n = m$), “Delaunay method” (resonant case for $n = m$), and “von Zeipel method” (resonant case for $n \geq m$) (see Poincaré (1957) and Arnol'd *et al.* (1988)).

The calculation of high-order terms in the procedures of elimination of fast variables is rather cumbersome. There are versions of these procedures which are convenient for symbolic processors (especially for Hamiltonian systems, e.g., the Deprit–Hori method; Giacaglia 1972).

The averaging method consists in using the averaged system for the description of motion in the first approximation and the truncated systems

obtained by means of the procedures of elimination of fast variables in the higher approximations, together with the corresponding transformations of variables.

Justification of the Averaging Method

To justify the averaging method, one should establish conditions under which the deviation of the slow variables along the solutions of the exact system from the solutions of the averaged system with appropriate initial data on time intervals of order $1/\varepsilon$ or longer tends to 0 as $\varepsilon \rightarrow 0$. It is desirable to have estimates from the above for these deviations. The estimates of deviations of the solutions of the exact system from the solutions of the truncated systems obtained by means of the procedure of elimination of fast phases are important as well. It can happen that there are “bad” initial data for which the slow component of the solution of the exact system deviates from the solution of the averaged system by a value of order 1 over time of order $1/\varepsilon$. In this case, one should have estimates from above for the measure of the set of such “bad” initial data; on the complementary set of initial data, one should have estimates from above for the deviation of slow variables along the solutions of the exact system from the solution of the averaged system. These problems are currently far from being completely solved. Some general results are described in the following.

Let functions ω, f, g on the right-hand side of system [1] be defined and bounded together with a sufficient number of derivatives in the domain $D\{I\} \times \mathbb{T}^m\{\varphi\} \times [0, \varepsilon_0]$. Let $J(t)$ be the solution of the averaged system [2] with initial condition $I_0 \in D$. Let $(I(t), \varphi(t))$ be the solution of the exact system [1] with initial conditions (I_0, φ_0) . So, $I(0) = J(0)$. It is assumed that the solution $J(t)$ is defined and stays at a positive distance from the boundary of D on the time interval $0 \leq t \leq K/\varepsilon, K = \text{const} > 0$.

If system [1] is a one-frequency system ($m = 1$), and the frequency ω does not vanish in D , then for $0 \leq t \leq K/\varepsilon$ the solution $(I(t), \varphi(t))$ is well defined, and $|I(t) - J(t)| < C\varepsilon, C = \text{const} > 0$. For $\omega = 1$, this assertion was proved by P Fatou (1928) and, by a different method, by L I Mandel'shtam and L D Papaleksi (1934). This was historically the first result on the justification of the averaging method (Mintropol'skii 1971). There is a proof based on the elimination of fast variables (see, e.g., Arnol'd (1983)). For a one-frequency system, higher approximations of the procedure of elimination of fast variables allow the description of the dynamics with an accuracy of the order of any power in ε on

time intervals of order $1/\varepsilon$ (Bogolyubov and Mitropol'skii 1961).

If system [1] is a multifrequency system ($m \geq 2$), but the vector of frequencies is constant and nonresonant, then for any $\rho > 0$ and small enough $\varepsilon < \varepsilon_0(\rho)$ it holds that $|I(t) - J(t)| < \rho$ for $0 \leq t \leq K/\varepsilon$ (Bogolyubov 1945, Bogolyubov and Mitropol'skii 1961). If, in addition, the frequencies satisfy the Diophantine condition $|(k, \omega)| > \text{const} |k|^{-\nu}$ for all $k \in \mathbb{Z}^m \setminus \{0\}$ and some $\nu > 0$, then one can choose $\rho = O(\varepsilon)$. In this case, higher approximations of the procedure of elimination of fast variables allow one to describe the dynamics with an accuracy of the order of any power in ε on time intervals of order $1/\varepsilon$ (see, e.g., Arnol'd *et al.* (1988)).

If the system is a multifrequency system, and frequencies are not constant (but depend on the slow variables I), then due to the evolution of slow variables the frequencies themselves are evolving slowly. At certain time moments, they can satisfy certain resonant relations. One of the phenomena that can take place here is a capture into a resonance; this capture leads to a large deviation of the solutions of the exact and averaged systems. However, the general Anosov averaging theorem (Anosov 1960) implies that if the frequencies ω are nonresonant for almost all I , then for any $\rho > 0$, the inequality $|I(t) - J(t)| < \rho$ is satisfied for $0 \leq t \leq K/\varepsilon$ for all initial data outside a set $E(\rho, \varepsilon)$ whose measure tends to 0 as $\varepsilon \rightarrow 0$. In many cases, it turns out that $\text{mes} E(\rho, \varepsilon) = O(\sqrt{\varepsilon}/\rho)$ (in particular, the sufficient condition for the last estimate is that $\text{rank}(\partial\omega/\partial I) = m$) (Arnol'd *et al.* (1988)).

The knowledge about averaging in two-frequency systems ($m = 2$) on time intervals, of order $1/\varepsilon$, is relatively more complete (see Arnol'd (1983), Arnol'd *et al.* (1988), and Lochak and Meunier (1988)). For Hamiltonian and reversible systems, the justification of the averaging method is a by-product of Kolmogorov–Arnold–Moser (KAM) theory. The KAM theory provides estimates of the difference between the solutions of the exact and averaged systems for majority of initial data on infinite time interval $-\infty < t < +\infty$. For remaining data this difference can grow because of Arnol'd diffusion, but, in general, very slowly. According to the Nekhoroshev theorem, this difference is small on time intervals whose length grows exponentially when the perturbation decays linearly (for an analytic Hamiltonian if the unperturbed Hamiltonian is a generic function, the so-called steep function).

Another aspect of justification of the averaging method is establishing relations between invariant manifolds of the exact and averaged systems. Consider, in particular, the case of a one-frequency

system and a multifrequency system with constant Diophantine frequencies. Suppose that the averaged system has an equilibrium such that real parts of all its eigenvalues are different from 0, or a limit cycle such that the absolute values of all but one of its multipliers are different from 1. Then the exact system has an invariant torus, respectively, m - or $(m + 1)$ -dimensional, whose projection onto the space of the slow variables is $O(\varepsilon)$ -close to the equilibrium (cycle) of the averaged system. This torus is stable or unstable together with the equilibrium (cycle) of the averaged system. For Hamiltonian and reversible systems, the problem of invariant manifolds is considered in the framework of the KAM theory.

Averaging in Bogolyubov's Systems

Systems in the standard form of Bogolyubov (1945) are of the form

$$\dot{x} = \varepsilon X(t, x, \varepsilon), \quad x \in \mathbb{R}^p, \quad 0 < \varepsilon \ll 1 \quad [10]$$

It is assumed that the function X , besides the usual smoothness conditions, satisfies the condition of uniform average: the limit (time average)

$$X_0(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t, x, 0) dt \quad [11]$$

exists uniformly in x . The averaging principle of Bogolyubov consists of the replacement of the original system in standard form by the averaged system

$$\dot{\xi} = \varepsilon X_0(\xi) \quad [12]$$

with a goal to provide an approximate description of the behavior of x . This approach generalizes the approach of the section “Averaging principle” for the case of constant frequencies ($\omega = \text{const}$). Upon introducing in the given system with constant frequencies the deviation from uniform rotation $\alpha = \varphi - \omega t$ and denoting $x = (I, \alpha)$, we obtain a system in the standard form [10]. Here the condition of uniform average is fulfilled because $X(t, x, 0)$ is a quasiperiodic function of time t . The averaged system [12] for nonresonant frequencies coincides with the averaged system [2]; for resonant frequencies, it coincides with the partially averaged system [3] (one should only supply systems [2] and [3] with equations for some components of the vector $\varphi - \omega t$ that do not enter into the right-hand side of the averaged system).

The averaging principle of Bogolyubov is justified by three Bogolyubov theorems. According to the

first theorem, if $\xi(t), 0 \leq t \leq K/\varepsilon$, is a solution of the averaged system, and $x(t)$ is a solution of the exact system with initial condition $x(0) = \xi(0)$, then for any $\rho > 0$ there exists $\varepsilon_0(\rho) > 0$ such that $|x(t) - \xi(t)| < \rho$ for $0 \leq t \leq K/\varepsilon$ and $0 < \varepsilon < \varepsilon_0(\rho)$. The second and the third Bogolyubov theorems describe the motion in the neighborhoods of equilibria and the limit cycles of the averaged system. In particular, if for an equilibrium real parts of all its eigenvalues are different from 0, or, for a limit cycle, the absolute values of all but one multipliers are different from 1, then the exact system has a solution which eternally stays near this equilibrium (cycle). The stability properties of this solution are the same as the stability properties of the corresponding equilibrium (cycle) of the averaged system.

For systems of the form [10] a procedure exists that, similarly to the procedure in the section “Elimination of fast variables, decoupling of slow and fast motions,” allows us to eliminate time t from the right-hand side of the system with an accuracy of the order of any power in ε by means of a transformation of variables. (To perform this procedure, one should assume that the conditions of uniform average are satisfied for functions that arise in the process of constructing higher approximations in this procedure (Bogolyubov and Mitropol'skii 1961).) In the first approximation, such a transformation of variables transforms the original system into the averaged one.

The condition of uniform average is very important for theory. If the limit in [11] exists, but convergence is nonuniform in x , then the time average X_0 could be, for example, a discontinuous function of x , and the averaged system would not be well defined.

Averaging in Slow-Fast Systems

Systems of the form [1] are particular cases of the systems of the form

$$\dot{x} = f(x, y, \varepsilon), \quad \dot{y} = \varepsilon g(x, y, \varepsilon) \quad [13]$$

which are called “slow-fast systems” (or systems with slow and fast motions, with slow and fast variables). The generalization of the approach of the section “Averaging principle” for these systems is the following averaging principle of Anosov (1960). In the system [6], let $x \in M, y \in \mathbb{R}^n$, where M is a smooth compact m -dimensional manifold. At $\varepsilon = 0$, the system for fast variables x contains slow variables y as parameters. Assume that this system (which is called “fast system”) has a finite smooth

invariant measure μ_y and is ergodic for almost all values of y . Introduce the averaged system

$$\dot{Y} = \varepsilon G(Y), \quad G(Y) = \frac{1}{\mu_Y(M)} \int_M g(x, Y, 0) d\mu_Y$$

According to the averaging principle, one should use the solution $Y(t)$ of the averaged system with initial condition $Y(0) = y(0)$ for approximate description of slow motion $y(t)$ in the original system. This averaging principle is justified by the following Anosov theorem [1]: *for any positive ρ the measure of the set $E(\rho, \varepsilon)$ of initial data (from a compact in the phase space) such that*

$$\max_{0 \leq t \leq 1/\varepsilon} |y(t) - Y(t)| > \rho$$

tends to 0 as $\varepsilon \rightarrow 0$.

The particular case when the original system is a Hamiltonian system depending on slowly varying parameter $\lambda = \varepsilon t$, and for almost all values of λ the motion of the system with $\lambda = \text{const}$ is ergodic on almost all energy levels, is considered in Kasuga (1961).

For the case when the has strong mixing properties, see Bakhtin (2004) and Kifer (2004).

For slow-fast systems, there is also a generalization of approach of the previous section that uses time averaging and the condition of uniform average (Volosov 1962).

Applications of the Averaging Method

The averaging method is one of the most productive methods of perturbation theory, and its applications are immense. It is widely used in celestial mechanics and space flight dynamics for the description of the evolution of motions of celestial bodies, in plasma physics and theory of accelerators for description of motion of charged particles, and in radio engineering for the description of nonlinear oscillatory regimes. There are also applications in hydrodynamics, physics of lasers, optics, acoustics, etc. (see Arnol'd *et al.* (1988), Bogolyubov and Mitropol'skii (1961), Lochak and Meunier (1988), Mitropol'skii (1971), and Volosov (1962)).

See also: Central Manifolds, Normal Forms; Diagrammatic Techniques in Perturbation Theory; Hamiltonian Systems: Stability and Instability Theory; KAM Theory and Celestial Mechanics; Multiscale Approaches; Random Walks in Random Environments; Separatrix Splitting; Stability Problems in Celestial Mechanics; Stability Theory and KAM.

Further Reading

- Anosov DV (1960) Averaging in systems of ordinary differential equations with rapidly oscillating solutions. *Izvestiya Akademii Nauk SSSR, Ser. Mat.* 24(5): 721–742 (Russian).
- Arnol'd VI (1983) *Geometrical Methods in the Theory of Ordinary Differential Equations*. New York–Berlin: Springer.
- Arnol'd VI, Kozlov VV, and Neishtadt AI (1988) *Mathematical Aspects of Classical and Celestial Mechanics*, Encyclopaedia of Mathematical Sciences, vol. 3. Berlin: Springer.
- Bakhtin VI (2004) Cramér asymptotics in the averaging method for systems with fast hyperbolic motions. *Proceedings of the Steklov Institute of Mathematics* 244(1): 79.
- Bogolyubov NN (1945) On some statistical methods in mathematical physics. *Akad. Nauk USSR. L'vov* (Russian).
- Bogolyubov NN and Mitropol'skii YuA (1961) *Asymptotic Methods in the Theory of Nonlinear Oscillations*. New York: Gordon and Breach.
- Giacaglia GEO (1972) *Perturbation Methods in Nonlinear Systems*, Applied Mathematical Science, vol. 8. Berlin: Springer.
- Kasuga T (1961) On the adiabatic theorem for the Hamiltonian system of differential equations in the classical mechanics I, II, III. *Proceedings of the Japan Academy* 37(7): 366–382.
- Kevorkian J and Cole JD (1996) *Multiple Scale and Singular Perturbations Methods*, Applied Mathematical Sciences, vol. 114. New York: Springer.
- Kifer Y (2004) Some recent advances in averaging. In: *Modern Dynamical Systems and Applications*, 403. Cambridge: Cambridge University Press.
- Lochak P and Meunier P (1988) *Multiphase Averaging for Classical Systems*, Applied Mathematical Sciences, vol. 72. New York: Springer.
- Mitropol'skii YuA (1971) *Averaging Method in Nonlinear Mechanics*. Kiev: Naukova Dumka (Russian).
- Poincaré H (1957) *Les Méthodes Nouvelles de la Mécanique Céleste*, vols. 1–3. New York: Dover.
- Sanders JA and Verhulst F (1985) *Averaging Methods in Nonlinear Dynamical Systems*, Applied Mathematical Sciences, vol. 59. New York: Springer.
- Volosov VM (1962) Averaging in systems of ordinary differential equations. *Russian Mathematical Surveys* 17(6): 1–126.

Axiomatic Approach to Topological Quantum Field Theory

C Blanchet, Université de Bretagne-Sud, Vannes, France

V Turaev, IRMA, Strasbourg, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The idea of topological invariants defined via path integrals was introduced by AS Schwartz (1977) in a special case and by E Witten (1988) in its full power. To formalize this idea, Witten (1988) introduced a notion of a topological quantum field theory (TQFT). Such theories, independent of Riemannian metrics, are rather rare in quantum physics. On the other hand, they admit a simple axiomatic description first suggested by M Atiyah (1989). This description was inspired by G Segal's (1988) axioms for a two-dimensional conformal field theory. The axiomatic formulation of TQFTs makes them suitable for a purely mathematical research combining methods of topology, algebra, and mathematical physics. Several authors explored axiomatic foundations of TQFTs (see Quinn (1995) and Turaev (1994)).

Axioms of a TQFT

An $(n + 1)$ -dimensional TQFT (V, τ) over a scalar field k assigns to every closed oriented n -dimensional manifold X a finite-dimensional vector space $V(X)$ over k and assigns to every cobordism (M, X, Y) a k -linear map

$$\tau(M) = \tau(M, X, Y) : V(X) \rightarrow V(Y)$$

Here a cobordism (M, X, Y) between X and Y is a compact oriented $(n + 1)$ -dimensional manifold M endowed with a diffeomorphism $\partial M \approx \overline{X} \amalg Y$ (the overline indicates the orientation reversal). All manifolds and cobordisms are supposed to be smooth. A TQFT must satisfy the following axioms.

1. *Naturality* Any orientation-preserving diffeomorphism of closed oriented n -dimensional manifolds $f : X \rightarrow X'$ induces an isomorphism $f_{\sharp} : V(X) \rightarrow V(X')$. For a diffeomorphism g between the cobordisms (M, X, Y) and (M', X', Y') , the following diagram is commutative:

$$\begin{array}{ccc} V(X) & \xrightarrow{(g|_X)_{\sharp}} & V(X') \\ \tau(M) \downarrow & & \downarrow \tau(M') \\ V(Y) & \xrightarrow{(g|_Y)_{\sharp}} & V(Y') \end{array}$$

2. *Functoriality* If a cobordism (W, X, Z) is obtained by gluing two cobordisms (M, X, Y) and (M', Y', Z) along a diffeomorphism $f : Y \rightarrow Y'$, then the following diagram is commutative:

$$\begin{array}{ccc} V(X) & \xrightarrow{\tau(W)} & V(Z) \\ \tau(M) \downarrow & & \downarrow \tau(M') \\ V(Y) & \xrightarrow{f_{\sharp}} & V(Y') \end{array}$$

3. *Normalization* For any n -dimensional manifold X , the linear map

$$\tau([0, 1] \times X) : V(X) \rightarrow V(X)$$

is identity.

4. *Multiplicativity* There are functorial isomorphisms

$$\begin{aligned} V(X \amalg Y) &\approx V(X) \otimes V(Y) \\ V(\emptyset) &\approx k \end{aligned}$$

such that the following diagrams are commutative:

$$\begin{array}{ccc} V((X \amalg Y) \amalg Z) &\approx & (V(X) \otimes V(Y)) \otimes V(Z) \\ \downarrow & & \downarrow \\ V(X \amalg (Y \amalg Z)) &\approx & V(X) \otimes (V(Y) \otimes V(Z)) \\ \\ V(X \amalg \emptyset) &\approx & V(X) \otimes k \\ \downarrow & & \downarrow \\ V(X) &= & V(X) \end{array}$$

Here $\otimes = \otimes_k$ is the tensor product over k . The vertical maps are respectively the ones induced by the obvious diffeomorphisms, and the standard isomorphisms of vector spaces.

5. *Symmetry* The isomorphism

$$V(X \amalg Y) \approx V(Y \amalg X)$$

induced by the obvious diffeomorphism corresponds to the standard isomorphism of vector spaces

$$V(X) \otimes V(Y) \approx V(Y) \otimes V(X)$$

Given a TQFT (V, τ) , we obtain an action of the group of diffeomorphisms of a closed oriented n -dimensional manifold X on the vector space $V(X)$. This action can be used to study this group.

An important feature of a TQFT (V, τ) is that it provides numerical invariants of compact oriented $(n + 1)$ -dimensional manifolds without boundary. Indeed, such a manifold M can be considered as a cobordism between two copies of \emptyset so that $\tau(M) \in \text{Hom}_k(k, k) = k$. Any compact oriented $(n + 1)$ -dimensional manifold M can be considered as a

cobordism between \emptyset and ∂M ; the TQFT assigns to this cobordism a vector $\tau(M)$ in $\text{Hom}_k(k, V(\partial M)) = V(\partial M)$ called the vacuum vector.

The manifold $[0, 1] \times X$, considered as a cobordism from $\bar{X} \amalg X$ to \emptyset induces a nonsingular pairing

$$V(\bar{X}) \otimes V(X) \rightarrow k$$

We obtain a functorial isomorphism $V(\bar{X}) = V(X)^* = \text{Hom}_k(V(X), k)$.

We now outline definitions of several important classes of TQFTs.

If the scalar field k has a conjugation and all the vector spaces $V(X)$ are equipped with natural nondegenerate Hermitian forms, then the TQFT (V, τ) is Hermitian. If $k = \mathbb{C}$ is the field of complex numbers and the Hermitian forms are positive definite, then the TQFT is unitary.

A TQFT (V, τ) is nondegenerate or cobordism generated if for any closed oriented n -dimensional manifold X , the vector space $V(X)$ is generated by the vacuum vectors derived as above from the manifolds bounded by X .

Fix a Dedekind domain $D \subset \mathbb{C}$. A TQFT (V, τ) over \mathbb{C} is almost D -integral if it is nondegenerate and there is $d \in D$ such that $d\tau(M) \in D$ for all M with $\partial M = \emptyset$. Given an almost integral TQFT (V, τ) and a closed oriented n -dimensional manifold X , we define $S(X)$ to be the D -submodule of $V(X)$ generated by all the vacuum vectors. This module is preserved under the action of self-diffeomorphisms of X and yields a finer “arithmetic” version of $V(X)$.

The notion of an $(n + 1)$ -dimensional TQFT over k can be reformulated in the categorical language as a symmetric monoidal functor from the category of n -manifolds and $(n + 1)$ -cobordisms to the category of finite-dimensional vector spaces over k . The source category is called the $(n + 1)$ -dimensional cobordism category. Its objects are closed oriented n -dimensional manifolds. Its morphisms are cobordisms considered up to the following equivalence: cobordisms (M, X, Y) and (M', X, Y) are equivalent if there is a diffeomorphism $M \rightarrow M'$ compatible with the diffeomorphisms $\partial M \approx \bar{X} \amalg Y \approx \partial M'$.

TQFTs in Low Dimensions

TQFTs in dimension $0 + 1 = 1$ are in one-to-one correspondence with finite-dimensional vector spaces. The correspondence goes by associating with a one-dimensional TQFT (V, τ) the vector space $V(pt)$ where pt is a point with positive orientation.

Let (V, τ) be a two-dimensional TQFT. The linear map τ associated with a pair of pants (a 2-disk with two holes considered as a cobordism between two

circles $S^1 \amalg S^1$ and one circle S^1) defines a commutative multiplication on the vector space $\mathcal{A} = V(S^1)$. The 2-disk, considered as a cobordism between S^1 and \emptyset , induces a nondegenerate trace on the algebra \mathcal{A} . This makes \mathcal{A} into a commutative Frobenius algebra (also called a symmetric algebra). This algebra completely determines the TQFT (V, τ) . Moreover, this construction defines a one-to-one correspondence between equivalence classes of two-dimensional TQFTs and isomorphism classes of finite dimensional commutative Frobenius algebras (Kock 2003).

The formalism of TQFTs was to a great extent motivated by the three-dimensional case, specifically, Witten’s Chern–Simons TQFTs. A mathematical definition of these TQFTs was first given by Reshetikhin and Turaev using the theory of quantum groups. The Witten–Reshetikhin–Turaev three-dimensional TQFTs do not satisfy exactly the definition above: the naturality and the functoriality axioms only hold up to invertible scalar factors called framing anomalies. Such TQFTs are said to be projective. In order to get rid of the framing anomalies, one has to add extra structures on the three-dimensional cobordism category. Usually one endows surfaces X with Lagrangians (maximal isotropic subspaces in $H_1(X; \mathbb{R})$). For 3-cobordisms, several competing – but essentially equivalent – additional structures are considered in the literature: 2-framings (Atiyah 1989), p_1 -structures (Blanchet *et al.* 1995), numerical weights (K Walker, V Turaev).

Large families of three-dimensional TQFTs are obtained from the so-called modular categories. The latter are constructed from quantum groups at roots of unity or from the skein theory of links. See Quantum 3-Manifold Invariants.

Additional Structures

The axiomatic definition of a TQFT extends in various directions. In dimension 2 it is interesting to consider the so-called open–closed theories involving 1-manifolds formed by circles and intervals and two-dimensional cobordisms with boundary (G Moore, G Segal). In dimension 3 one often considers cobordisms including framed links and graphs whose components (resp. edges) are labeled with objects of a certain fixed category \mathcal{C} . In such a theory, surfaces are endowed with finite sets of points labeled with objects of \mathcal{C} and enriched with tangent directions. In all dimensions one can study manifolds and cobordisms endowed with homotopy classes of mappings to a fixed space (homotopy quantum field theory, in the sense of Turaev). Additional structures on the tangent bundles – spin

structures, framings, etc. – may be also considered provided the gluing is well defined.

See also: Braided and Modular Tensor Categories; Hopf Algebras and q -Deformation Quantum Groups; Indefinite Metric; Quantum 3-Manifold Invariants; Topological Gravity, Two-Dimensional; Topological Quantum Field Theory: Overview.

Further Reading

- Atiyah M (1989) Topological Quantum Field Theories. *Publications Mathématiques de l'Ihés* 68: 175–186.
- Bakalov B and Kirillov A Jr. (2001) Lectures on Tensor Categories and Modular Functors. *University Lecture Series* vol. 21. Providence, RI: American Mathematical Society.

- Blanchet C, Habegger N, Masbaum G, and Vogel P (1995) Topological quantum field theories derived from the Kauffman bracket. *Topology* 34: 883–927.
- Kock J (2003) Frobenius Algebras and 2D Topological Quantum Field Theories. *LMS Student Texts*, vol. 59. Cambridge: Cambridge University Press.
- Quinn F (1995) Lectures on axiomatic topological quantum field Freed DS and Uhlenbeck KK (eds.) *Geometry and Quantum Field Theory*, pp. 325–453. IAS/Park City Mathematical Series, University of Texas, Austin: American Mathematical Society.
- Segal G (1988) Two-dimensional conformal field theories and modular functors. In: Simon B, Truman A, and Davies IM (eds.) *IXth International Congress on Mathematical Physics*, pp. 22–37. Bristol: Adam Hilger Ltd.
- Turaev V (1994) Quantum Invariants of Knots and 3-Manifolds. *de Gruyter Studies in Mathematics*, vol. 18. Berlin: Walter de Gruyter.
- Witten E (1988) Topological quantum field theory. *Communication in Mathematical Physics* 117(3): 353–386.

Axiomatic Quantum Field Theory

B Kuckert, Universität Hamburg, Hamburg, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The term “axiomatic quantum field theory” subsumes a collection of research branches of quantum field theory analyzing the general principles of relativistic quantum physics. The content of the results typically is structural and retrospective rather than quantitative and predictive.

The first axiomatic activities in quantum field theory date back to the 1950s, when several groups started investigating the notion of scattering and S -matrix in detail (Lehmann, Symanzik, and Zimmermann 1955 (LSZ-approach), Bogoliubov and Parasiuk 1957, Hepp and Zimmermann (BPHZ-approach), Haag 1957–59 and Ruelle 1962 (Haag–Ruelle theory) (*see* Scattering, Asymptotic Completeness and Bound States and Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools).

Wightman (1956) analyzed the properties of the vacuum expectation values used in these approaches and formulated a system of axioms that the vacuum expectation values ought to satisfy in general. Together with Gårding (1965), he later formulated a system of axioms in order to characterize general quantum fields in terms of operator-valued functionals, and the two systems have been found to be equivalent.

A couple of spectacular theorems such as the PCT theorem and the spin–statistics theorem have been obtained in this setting, but no interacting quantum fields satisfying the axioms have been found so far

(in $1 + 3$ spacetime dimensions). So, the development of alternatives and modifications of the setting got into the focus of the theory, and the axioms themselves became the objects of research. Their role as axioms – understood in the common sense – turned into the role of mere properties of quantum fields. Today, the term “axiomatic quantum field theory” is widely avoided for this reason.

In a long list of publications spread over the 1960s, Araki, Borchers, Haag, Kastler, and others worked out an algebraic approach to quantum field theory in the spirit of Segal’s “postulates for general quantum Mechanics” (1947) (*see* Algebraic Approach to Quantum Field Theory).

The Wightman setting was the basis of a framework into which the causal construction of the S -matrix developed by Stückelberg (1951) and Bogoliubov and Shirkov (1959) has been fitted by Epstein and Glaser (1973). The causality principle fixes the time-ordered products up to a finite number of parameters at each order, which are to be put in as the renormalization constants.

Already in 1949, Dyson had seen that problems in the formulation of quantum electrodynamics (QED) could be avoided by “just” multiplying the time variable and, correspondingly, the energy variable by the imaginary unit constant (“Wick rotation”). Schwinger then investigated time-ordered Green functions of QED in this Euclidean setting. This approach was formulated in terms of axioms by Osterwalder and Schrader (1973, 1975) (*see* Euclidean Field Theory).

Other extensions of the aforementioned settings are objects of current research (*see* Indefinite Metric,

Quantum Field Theory in Curved Spacetime, Symmetries in Quantum Field Theory of Lower Spacetime Dimensions, and Thermal Quantum Field Theory).

Quantum Fields

Gårding and Wightman characterized operator-valued quantum fields on the Minkowski spacetime \mathbb{R}^{1+3} by a couple of axioms. Given additional assumptions concerning the high-energy behavior, the Gårding–Wightman fields are in one–one correspondence with algebraic field theories.

Without specifying or presupposing these additional assumptions, the axioms will now be formulated and discussed in detail and compared to the corresponding conditions in the algebraic setting. Adjoint operators are marked by an asterisk, and Einstein’s summation convention is used.

Operator-valued functionals *The components of a field F are an n -tuple $F_1 \cdots F_n$ of linear maps that assign to each test function $\varphi \in C_0^\infty(\mathbb{R}^{1+3})$ linear operators $F_1(\varphi) \cdots F_n(\varphi)$ in a Hilbert space \mathcal{H} with domains of definition $D(F_1(\varphi)) \cdots D(F_n(\varphi))$. There exists a dense subspace \mathcal{D} of \mathcal{H} with $\mathcal{D} \subset D(F_\nu(\varphi)) \cap D(F_\nu(\varphi)^*)$ and $F_\nu(\varphi)\mathcal{D} \cup F_\nu(\varphi)^*\mathcal{D} \subset \mathcal{D}$ for all indices ν . Consider m such fields $F^1 \cdots F^m$ with components $F_\nu^a, 1 \leq a \leq m, 1 \leq \nu \leq n_a$. Assume there to be an involution $*$: $(1 \cdots m) \rightarrow (1 \cdots m)$ such that $F_\nu^a(\varphi) = F_\nu^a(\overline{\varphi})^*$, where $\overline{\varphi}(x) := \overline{\varphi(x)}$.*

Quantum fields cannot be operator-valued functions on \mathbb{R}^{1+3} if one wants them to exhibit (part of) the properties to follow. But point fields can be quadratic forms; typically this is the case for fields in a Fock space.

For each component F_ν^a and each open region $\mathcal{O} \subset \mathbb{R}^{1+3}$, the field operators $F_\nu^a(\varphi)$ with $\text{supp } \varphi \subset \mathcal{O}$ generate a $*$ -algebra $\mathcal{F}_\nu^a(\mathcal{O})$ of operators defined on \mathcal{D} . These operators typically are unbounded, which is one of the differences with the traditional setting of the algebraic approach. There a C^* -algebra $\mathfrak{A}(\mathcal{O})$ is assigned to each open region \mathcal{O} in such a way that $\mathcal{O} \subset \mathcal{P}$ implies $\mathfrak{A}(\mathcal{O}) \subset \mathfrak{A}(\mathcal{P})$. Each C^* -algebra is a $*$ -algebra, but in contrast to a C^* -algebra, a $*$ -algebra does not need to be endowed with a norm. The fundamental observables in quantum theory are bounded positive operators (typically, but not always, projections), and these generate a C^* -algebra.

There is no fundamental physical motivation for confining the setting to fields with a finite number of components, except that it includes most of the fields known from “daily life.”

Continuity as a distribution *For all $\Phi, \Psi \in \mathcal{D}$, the linear functionals $T_{\Phi, \Psi, \nu}$ on $C_0^\infty(\mathbb{R}^{1+3})$ defined by*

$$T_{\nu, \Phi, \Psi}^a(\varphi) := \langle \Phi, F_\nu^a(\varphi)\Psi \rangle$$

are distributions. They can be extended to tempered distributions.

The Fourier transform of a tempered distribution is well defined as a tempered distribution. It is mainly due to the importance of Fourier transformations that the preceding assumption is convenient. Bogoliubov *et al.* (1975) remark that the assumption is not a mere technicality, since it rules out nonrenormalizable quantum fields.

Microcausality (Bose–Fermi alternative) *If φ and ψ are test functions with spacelike separated support, then*

$$F_\nu^a(\varphi)F_\mu^b(\psi)|_{\mathcal{D}} = \pm F_\mu^b(\psi)F_\nu^a(\varphi)|_{\mathcal{D}}.$$

The sign depends on the statistics of the fields, it is “ $-$ ” if and only if both F^a and F^b are fermion fields.

Microcausality is closely related to Einstein causality. Einstein causality requires that any two observables located in spacelike separated regions commute in the strong sense, that is, their spectral measures commute. But fields with Fermi–Dirac statistics are not observables, and not even for Bose–Einstein fields with self-adjoint field operators does the above condition imply that the spectral projections commute, which is the criterion for commensurability. The sign on the right-hand side does, however, specify the statistics of the field.

This is a crucial difference with the algebraic approach. If \mathcal{O} and \mathcal{P} are spacelike separated open regions and if $A \in \mathfrak{A}(\mathcal{O})$ and $B \in \mathfrak{A}(\mathcal{P})$, then one assumes, like in the above case, that $AB = BA$ (locality). But being elements of C^* -algebras, A and B are bounded operators (or can be represented accordingly), so if A and B are self-adjoint, they are, indeed, commensurable.

Doplicher, Haag, and Roberts (1974) and Buchholz and Fredenhagen (1984) have derived from this input of observables a field structure of *localized particle states*, and they showed that the statistics of these fields is Bose–Einstein, Fermi–Dirac, or some corresponding parastatistics (which is, *a priori*, forbidden if one assumes microcausality).

Recall that the unimodular group $SL(2, \mathbb{C})$ is isomorphic to the universal covering group of the restricted Lorentz group L_+^\uparrow (the connected component containing the unit element). Denote by $\Lambda : SL(2, \mathbb{C}) \rightarrow L_+^\uparrow$ a covering map.

Covariance *There exist strongly continuous unitary representations U and T of $SL(2, \mathbb{C})$ and $(\mathbb{R}^{1+3}, +)$, respectively, and representations $D^1 \cdots D^m$ of $SL(2, \mathbb{C})$ in $\mathbb{C}^{n_1} \cdots \mathbb{C}^{n_m}$, respectively, such that*

$$U(g)F_\nu^a(\varphi)U(g)^* = D^a(g^{-1})_\nu^\mu F_\mu^a(\varphi(\Lambda(g)^{-1}\cdot))$$

and

$$T(y)F_\nu^a(\varphi)T(y)^* = F_\mu^a(\varphi(\cdot - y)),$$

where $D^a(g^{-1})_\nu^\mu$ are the elements of the matrix $D^a(g^{-1})$. Dropping coordinate indices, this reads

$$U(g)F^a(\varphi)U(g)^* = D^a(g^{-1})F^a(\varphi(\Lambda(g)^{-1}\cdot))$$

and

$$T(y)F^a(\varphi)T(y)^* = F^a(\varphi(\cdot - y)).$$

The representations U and T generate a representation of the universal covering of the restricted Poincaré group.

As it stands, this assumption is a very strong one, since it manifestly fixes the action of the representation on the field operators. In the algebraic approach, the covariance assumption is more modestly formulated. Namely, it is assumed that $U(g)\mathfrak{A}(\mathcal{O})U(g)^* = \mathfrak{A}(\Lambda(g)\mathcal{O})$ and $T(y)\mathfrak{A}(\mathcal{O})T(y)^* = \mathfrak{A}(\mathcal{O} + y)$, leaving open how the representation acts on the single local observables.

Vacuum vector *There exists a unique (up to a multiple) vector $\Omega \in \mathcal{D}$ that is invariant under the representations U and T and cyclic with respect to the algebra $\mathcal{F}(\mathbb{R}^{1+3})$ generated by all field operators $F_\nu^a(\varphi)$, that is, $\mathcal{F}(\mathbb{R}^{1+3})\Omega = \mathcal{H}$.*

Spectrum condition *The joint spectrum of the components of the 4-momentum, i.e., of the generators of the spacetime translations, has support in the closed forward light cone \overline{V}_+ , that is, the set $\{k^2 \geq 0, k_0 \geq 0\}$.*

The existence of an invariant ground state called the vacuum is standard in algebraic quantum field theory as well.

N-Point Functions

Consider the above fields $F^1 \cdots F^m$. For each $N \in \mathbb{N}$ and each N -tuple $(a_1 \cdots a_N)$ of natural numbers $\leq m$ (labeling fields), define families $(F^{a_1 \cdots a_N}) := (F_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N})_{\nu_i \leq n_{a_i}}$ and $(w^{a_1 \cdots a_N}) := (w_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N})_{\nu_i \leq n_{a_i}}$ of distributions on $(\mathbb{R}^{1+3})^N$ by

$$F_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_N) := F_{\nu_1}^{a_1}(\varphi_1) \cdots F_{\nu_N}^{a_N}(\varphi_N)$$

(using the nuclear theorem) and

$$w_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\psi) := \langle \Omega, F_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\psi)\Omega \rangle. \quad [1]$$

These distributions are called the “ N -point functions” of the fields $F^1 \cdots F^m$ and yield the vacuum expectation values of the theory. It is straightforward to deduce the following properties from the Gårding–Wightman axioms.

Microcausality (Bose–Fermi alternative) *If φ_i and φ_{i+1} have spacelike separated supports, then*

$$\begin{aligned} w_{\nu_1 \cdots \nu_i \nu_{i+1} \cdots \nu_N}^{a_1 \cdots a_i a_{i+1} \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_i \otimes \varphi_{i+1} \otimes \cdots \otimes \varphi_N) \\ = \pm w_{\nu_1 \cdots \nu_{i+1} \nu_i \cdots \nu_N}^{a_1 \cdots a_{i+1} a_i \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_{i+1} \otimes \varphi_i \otimes \cdots \otimes \varphi_N). \end{aligned}$$

or dropping coordinate indices,

$$\begin{aligned} w^{a_1 \cdots a_i a_{i+1} \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_i \otimes \varphi_{i+1} \otimes \cdots \otimes \varphi_N) \\ = \pm w^{a_1 \cdots a_{i+1} a_i \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_{i+1} \otimes \varphi_i \otimes \cdots \otimes \varphi_N). \end{aligned}$$

Invariance *For all $g \in SL(2, \mathbb{C})$ and $y \in \mathbb{R}^{1+3}$, one has*

$$\begin{aligned} w_{\mu_1 \cdots \mu_N}^{a_1 \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_N) \\ = D^{a_1}(g^{-1})_{\mu_1}^{\nu_1} \cdots D^{a_N}(g^{-1})_{\mu_N}^{\nu_N} \\ \times w_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\Lambda(g)\varphi_1 \otimes \cdots \otimes \Lambda(g)\varphi_N) \\ = w_{\mu_1 \cdots \mu_N}^{a_1 \cdots a_N}(\varphi_1(\cdot - y) \otimes \cdots \otimes \varphi_N(\cdot - y)) \end{aligned}$$

or dropping coordinate indices,

$$\begin{aligned} w^{a_1 \cdots a_N}(\varphi_1 \otimes \cdots \otimes \varphi_N) \\ = (D^{a_1}(g^{-1}) \otimes \cdots \otimes D^{a_N}(g^{-1})) \\ \times w^{a_1 \cdots a_N}(\Lambda(g)\varphi_1 \otimes \cdots \otimes \Lambda(g)\varphi_N) \\ = w^{a_1 \cdots a_N}(\varphi_1(\cdot - y) \otimes \cdots \otimes \varphi_N(\cdot - y)). \end{aligned}$$

By translation invariance, the N -point functions $w_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(x_1 \cdots x_N)$ only depend on the $N - 1$ relative-position vectors $\xi_1 := x_1 - x_2$, $\xi_2 := x_2 - x_3, \dots$, $\xi_{N-1} := x_{N-1} - x_N$. This means that there are distributions $W_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}$ on $(\mathbb{R}^{1+3})^{N-1}$ related to the N -point functions by the symbolic condition

$$w_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(x_1 \cdots x_N) = W_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\xi_1 \cdots \xi_{N-1}).$$

In precise notation, this reads

$$w_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\varphi) = \int_{1+3} W_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}(\varphi_x) dx,$$

where

$$\begin{aligned} \varphi_x(\xi_1 \cdots \xi_{N-1}) := \varphi(x, x - \xi_1, x - \xi_1 - \xi_2, \dots, x - \xi_1 \\ \cdots - \xi_{N-1}). \end{aligned}$$

The functions $W_{\nu_1 \cdots \nu_N}^{a_1 \cdots a_N}$ are called the *Wightman functions*, and they have the following property because of the spectrum condition of the field.

Spectrum condition *The support of the Fourier transform of each $W_{\nu_1 \dots \nu_N}^{a_1 \dots a_N}$ is contained in $(\bar{V}_+)^{N-1}$.*

The uniqueness of the vacuum vector (up to a phase) is equivalent to the following condition.

Cluster property *For $N \geq 2$, let x be a spacelike vector in \mathbb{R}^{1+3} , let L be a natural number $< N$, and let φ and ψ be tempered test functions on $(\mathbb{R}^{1+3})^L$ and $(\mathbb{R}^{1+3})^{N-L}$, respectively. then*

$$\begin{aligned} \lim_{0 < \lambda \rightarrow \infty} w_{\nu_1 \dots \nu_N}^{a_1 \dots a_N}(\varphi \otimes \psi(\cdot - \lambda x)) \\ = w_{\nu_1 \dots \nu_L}^{a_1 \dots a_L}(\varphi) w_{\nu_{L+1} \dots \nu_N}^{a_{L+1} \dots a_N}(\psi). \end{aligned}$$

On the one hand, these properties have been deduced from the Gårding–Wightman axioms via eqn [1]. Conversely, a family of distributions labeled in the above fashion and satisfying the above properties may be used to construct a Gårding–Wightman field theory provided that two more conditions – which hold for all systems of N -point functions – are satisfied. This requires some elementary notation.

Define the index sets

$$\mathcal{I}_N := \left\{ \left(\begin{array}{c} a_1 \cdots a_N \\ \nu_1 \cdots \nu_N \end{array} \right) : 1 \leq a_i \leq m, 1 \leq \nu_i \leq n_{a_i} \right. \\ \left. \text{for all } 1 \leq i \leq N \right\}, \quad N \in \mathbb{N}$$

$\mathcal{I}_0 := \{\emptyset\}$, and $\mathcal{I} := \bigcup_{N \in \mathbb{N}_0} \mathcal{I}_N$. On \mathcal{I} a concatenation \circ is defined by

$$\left(\begin{array}{c} a_1 \cdots a_N \\ \nu_1 \cdots \nu_N \end{array} \right) \circ \left(\begin{array}{c} b_1 \cdots b_M \\ \mu_1 \cdots \mu_M \end{array} \right) := \left(\begin{array}{c} a_1 \cdots a_N b_1 \cdots b_M \\ \nu_1 \cdots \nu_N \mu_1 \cdots \mu_M \end{array} \right)$$

and

$$\emptyset \circ \kappa := \kappa \circ \emptyset := \kappa$$

and an involution $*$ by

$$\left(\begin{array}{c} a_1 \cdots a_N \\ \nu_1 \cdots \nu_N \end{array} \right)^* := \left(\begin{array}{c} a_N^* \cdots a_1^* \\ \nu_N \cdots \nu_1 \end{array} \right) \quad \text{and} \quad \emptyset^* := \emptyset.$$

Define an antilinear involution $*$ on $\mathcal{S}^N := \mathcal{S}((\mathbb{R}^{1+3})^N)$ by

$$\psi(x_1 \cdots x_N) := \overline{\psi(x_N \cdots x_1)}$$

for each $N \in \mathbb{N}$. Put $\mathcal{S}^0 := \mathbb{C}$ and $z^* := \bar{z}$ for all $z \in \mathbb{C}$.

Define $\mathcal{S}^{\mathcal{I}_N} := \mathcal{S}^N \times \mathcal{I}_N$, and $\mathcal{S}^{\mathcal{I}} := \bigcup_N \mathcal{S}^{\mathcal{I}_N}$. For each $\kappa \in \mathcal{I}_N$, the set $\mathcal{S}^\kappa := \mathcal{S}((\mathbb{R}^{1+3})^N) \times \{\kappa\}$ is a linear space. On the direct sum $\mathcal{B}^{\mathcal{I}} := \bigoplus_{\kappa \in \mathcal{I}} \mathcal{S}^\kappa$ define an associative product by

$$(\psi, \kappa)(\chi, \lambda) := (\psi \otimes \chi, \kappa \circ \lambda)$$

and an antilinear involution $*$ by $(\psi, \kappa)^* := (\psi^*, \kappa^*)$. This endows $\mathcal{B}^{\mathcal{I}}$ with the structure of a nonabelian $*$ -algebra with unit element $1 = (1, \emptyset)$ (Borchers algebra).

If one defines $F_\emptyset(z) := z\mathbf{1}$, then $w_\emptyset(z) = z$, and the Wightman functions induce a \mathbb{C} -linear functional ω on $\mathcal{B}^{\mathcal{I}}$ by

$$\omega(\psi, \kappa) := w_\kappa(\psi) \quad [2]$$

ω exhibits the following two properties, which are the announced additional conditions required for reconstructing the fields from the N -point functions.

Hermiticity $\omega(\xi^*) = \overline{\omega(\xi)}$.

Positivity $\omega(\xi^* \xi) \geq 0$.

To see Hermiticity, compute

$$\begin{aligned} \omega(\psi^*, \kappa^*) &= \langle \Omega, F_{\kappa^*}(\psi^*) \Omega \rangle \\ &= \langle F_\kappa(\psi) \Omega, \Omega \rangle = \overline{\omega(\psi, \kappa)} \end{aligned}$$

and use \mathbb{C} -linearity to prove the statement for arbitrary $\xi \in \mathcal{B}$. For positivity, write any ξ as a finite sum $\xi = (\psi_1, \kappa_1) + \cdots + (\psi_M, \kappa_M)$, and compute

$$\begin{aligned} \omega(\xi^* \xi) &= \omega \left(\sum_{i,j=1}^M (\psi_i, \kappa_i)^* (\psi_j, \kappa_j) \right) \\ &= \omega \left(\sum_{ij} (\psi_i^* \otimes \psi_j, \kappa_i^* \circ \kappa_j) \right) \\ &= \sum_{ij} w_{\kappa_i^* \circ \kappa_j}(\psi_i^* \otimes \psi_j) \\ &= \sum_{ij} \langle \Omega, F_{\kappa_i^* \circ \kappa_j}(\psi_i^* \otimes \psi_j) \Omega \rangle \\ &= \sum_{ij} \langle \Omega, F_{\kappa_i^*}(\psi_i^*) F_{\kappa_j}(\psi_j) \Omega \rangle \\ &= \sum_{ij} \langle F_{\kappa_i}(\psi_i) \Omega, F_{\kappa_j}(\psi_j) \Omega \rangle \\ &= \left\| \sum_i F_{\kappa_i}(\psi_i) \Omega \right\|^2 \geq 0. \end{aligned}$$

Theorem 1 (Wightman’s reconstruction theorem). *Let m and $n_1 \cdots n_m$ be natural numbers, let $\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \dots$, and \mathcal{I} be the above index sets, and let $\mathcal{B}^{\mathcal{I}}$ be the above Borchers algebra. Let $D_1 \cdots D_m$ be matrix representations of $\text{SL}(2, \mathbb{C})$ in $\mathbb{C}^{n_1} \cdots \mathbb{C}^{n_m}$, respectively.*

For each natural number N , let $(w_\kappa)_{\kappa \in \mathcal{I}_N}$ be a family of distributions on $(\mathbb{R}^{1+3})^N$. Suppose the family $(w_\kappa)_{\kappa \in \mathcal{I}}$ defined this way satisfies microcausality, covariance, spectrum condition, and the cluster property. If the linear functional ω defined on $\mathcal{B}^{\mathcal{I}}$ by eqn [2] is Hermitian and positive, then

there is (up to unitary equivalence) a unique family $F^1 \cdots F^m$ of Gårding–Wightman fields with $n_1 \cdots n_m$ components such that eqn [1] holds.

The proof uses the GNS construction known from the theory of operator algebras. The Borchers algebra plays several roles. On the one hand, it is a linear space with an inner product. The Hilbert space \mathcal{H} and the invariant space \mathcal{D} of the field theory are constructed from this structure. On the other hand, the Borchers algebra acts on itself as an algebra of linear operators by its own algebra multiplication. This is the structure the $*$ -algebra of field operators is constructed from.

Results

The mathematical and structural analysis of quantum fields has improved the understanding of scattering theory in the different approaches mentioned above; see Bogoliubov *et al.* (1975) and the relevant articles in this encyclopedia. Apart from this, the following results deserve to be mentioned. Evidently, many others have to be omitted for practical reasons.

PCT Symmetry

An early famous result was Lüders’s proof (1957) that all fields in the above setting exhibit PCT symmetry, that is, the symmetry under reflections in all space and time variables combined with a charge conjugation. This symmetry is exhibited by all particle reactions observed so far. The proof, like several of the main results, made extensive use of the fact that the N -point functions are boundary values of analytic functions due to the spectrum condition, and that a fundamental theorem by Bargmann, Hall, and Wightman (1957) yields invariant analytic extensions.

Reeh–Schlieder Theorem

For each field F_ν^a and each bounded open region $\mathcal{O} \subset \mathbb{R}^{1+3}$, the vacuum vector is cyclic with respect to $\mathcal{F}_\nu^a(\mathcal{O})$ (Reeh and Schlieder 1961). So excitations of the vacuum vector by field operators located in \mathcal{O} are not to be considered as state vectors of a particle localized in \mathcal{O} , since they are not perpendicular to the excitations by field operators located outside \mathcal{O} .

Unruh Effect and Modular P_1 CT Symmetry

In the 1970s, Bisognano and Wichmann (1975, 1976) discovered a surprising link of symmetries to the intrinsic algebraic structure of quantum fields, which is established by the Tomita–Takesaki modular theory (see Tomita–Takesaki Modular Theory). Namely, the

unitary operators implementing the Lorentz boosts on the fields are elements of modular groups. This means that a uniformly accelerated observer perceives the vacuum as a thermal state with a temperature proportional to its acceleration, corresponding to the famous Unruh effect.

In addition, it was shown that P_1 CT symmetries (i.e., PCT combined with rotations by the angle π) are implemented by modular conjugations (modular P_1 CT symmetry). Modular P_1 CT symmetry is a consequence of the Unruh effect (Guido and Longo 1995).

Spin and Statistics

Immediately following Lüders’s PCT theorem, the spin–statistics theorem was proved for the N -point functions of the Wightman setting (Lüders and Zumino 1958, Burgoyne 1958, Dell’Antonio 1961). This was a remarkable and widely acknowledged progress. But as remarked earlier, the confinement to finite-component fields, which is used in the proof, cannot be motivated by physical first principles (i.e., in a truly axiomatic fashion). The representation D of $SL(2, \mathbb{C})$ acting on the components, however, is forced to be finite dimensional by this assumption, and since the representations D^a are objects of investigation, a considerable part of the result is assumed this way from the outset. Even more so, there are examples of fields with a “wrong” spin–statistics connection and infinitely many components.

This was one reason to continue working on the subject. At the beginning of the 1990s, it was found that the spin–statistics theorem can be derived from the symmetries discovered by Bisognano and Wichmann, and Unruh. Two approaches not referring to the number of internal degrees of freedom have been worked out: one assumes the Unruh effect (Guido and Longo 1995), the other modular P_1 CT symmetry (Kuckert 1995, 2005, Kuckert and Lorenzen 2005). The first approach has been generalized to conformal fields, the second to the case that the symmetry group’s homogeneous part is not $SL(2, \mathbb{C})$, but only $SU(2)$.

Both approaches can be applied to infinite-component fields. They yield existence theorems; a distinguished representation is constructed from the modular symmetries, and this representation exhibits Pauli’s spin–statistics connection. As mentioned before, nothing more can be expected at this level of generality. The line of argument works in both the algebraic and the Wightman setting.

A Dynamical Property of the Vacuum

One can derive the spectrum condition, the Bisognano–Wichmann symmetries/the Unruh effect, and

covariance from the condition that no (inertial or) uniformly accelerated observer can extract mechanical energy from the field *in vacuo* by means of a cyclic process (Kuckert 2002).

Interacting Fields

The examples of interacting quantum fields that fit into the above settings live in one or two spatial dimensions only, and their relevance for physics mainly consists in being such examples. This has contributed to some frustration and to doubts on whether one is not, in fact, proving theorems on pretty empty sets, or in other words, working on “the most sophisticated theory of the free field.”

The computations in quantum field theory are, like most of the computations in physics, perturbative. In order to be successful, they need to yield good agreement with experiment with reasonable computational efforts, that is, by evolution up to the second or third order. This asymptotic convergence is more important than convergence of the series as a whole. There are low-dimensional examples of interacting Wightman fields (e.g., $(\varphi^4)_2$; cf. the monograph by Glimm and Jaffe (1987)), and time will tell whether four-dimensional interacting Wightman fields exist. But there is no reason to expect convergence for general interacting fields; for example, QED does not fit into the Wightman framework.

The appropriate extension of the Wightman setting has been formulated by Epstein and Glaser (1973). It defines the *S-matrix* rather than the field itself as a (in general divergent) formal power series of operator-valued distributions.

The above results apply to this somewhat more modest setting as well, so the “axiomatic” approaches do help in understanding the known high-energy physics interactions. This even includes gauge theories (see Perturbative Renormalization Theory and BRST). The high-precision results of QED can be reproduced within this setting, and there occur no UV singularities: renormalization amounts to the need to extend distributions by fixing some parameters, that is, the renormalization constants. The infrared problem is circumvented by considering the *S-matrix* as a (position-dependent) distribution taking values in the unitary formal power series of distributions rather than as a single (global) unitary operator (or unitary power series).

Quantum Energy Inequalities

Energy densities of Wightman fields admit negative expectation values (Epstein, Glaser, and Jaffe 1965). This is in contrast to the positivity conditions that the energy–momentum tensors of classical general

(and, hence, also special) relativity have to satisfy to ensure causality. But the conflict can be solved by smearing the densities out in space or time, as has first been realized by Ford (1991). The extent to which the energy density can become negative depends on the extent to which it is smeared out: “more smearing means less violation of positivity,” so the classical positivity conditions are restored at medium and large scales. There are many ways to make this principle concrete. Quantum energy inequalities hold for thermodynamically well-behaved quantum fields on causally well-behaved classical spacetime backgrounds.

Bibliographic Notes

Important monographs on axiomatic quantum field theory are those by Streater and Wightman (1964), Jost (1965), Bogoliubov *et al.* (1975), and Bogoliubov *et al.* (1990). Note that the books of Bogoliubov *et al.* differ in setup fundamentally and that neither replaces the other. For a lecture notes volume, see also Völkel (1977), and for a review article, see Streater (1975). A valuable discussion of the Wightman axioms can also be found in the second volume of the series by Reed and Simon (1970).

The first monograph on the algebraic approach to quantum field theory is due to Haag (1992), a more recent one has been written by Araki (1999). Concerning the sufficient conditions for “switching” between the Gårding–Wightman and the algebraic approach, see Wollenberg (1988) and the Ph.D. thesis of Bostelmann (2000) and references given there. Dynamical and thermodynamical foundation of standard axioms, the Bisognano–Wichmann symmetries (Unruh effect), and the spin–statistics theorem, have been investigated by Kuckert (2002, 2005), see also the references given there for related work.

In different formulations and at differing degrees of mathematical sophistication, the causal approach to perturbation theory can be found in the monographs by Bogoliubov and Shirkov (1959), Scharf (1989, 2001), and Steinmann (2000). Two modern review articles have been written by Brunetti and Fredenhagen (2000) and by Dütsch and Fredenhagen (2004).

The reference original articles on the Euclidean axioms are those of Osterwalder and Schrader (1973, 1975). Note that the first one contains an error. (cf. also Zinoviev (1995)). A monograph on Euclidean field theory and its relations to the other axiomatic settings of quantum field theory and to statistical mechanics is that by Glimm and Jaffe (1987).

A recent review on quantum energy inequalities is due to Fewster (2003).

Acknowledgments

The author is a fellow of the Emmy-Noether Programme (DFG). Thanks for discussions are due to Professor D Arlt.

See also: Algebraic Approach to Quantum Field Theory; C^* -Algebras and Their Classification; Constructive Quantum Field Theory; Dispersion Relations; Euclidean Field Theory; Indefinite Metric; Perturbative Renormalization Theory and BRST; Quantum Field Theory: A Brief Introduction; Quantum Field Theory in Curved Spacetime; Scattering, Asymptotic Completeness and Bound States; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Scattering in Relativistic Quantum Field Theory: The Analytic Program; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Thermal Quantum Field Theory; Tomita–Takesaki Modular Theory; Two-Dimensional Models.

Further Reading

- Araki H (1999) *Mathematical Theory of Quantum Fields*. Oxford: Oxford University Press.
- Bogoliubov NN, Logunov AA, and Todorov IT (1975) *Introduction to Axiomatic Quantum Field Theory*, (Russian original edition: Nauka (Moskow) 1969). New York: Benjamin.
- Bogoliubov NN, Logunov AA, Oksak AI, and Todorov IT (1990) *General Principles of Quantum Field Theory*, (Russian original edition Nauka (Moskow) 1987). Dordrecht–Boston–London: Kluwer.
- Bostelmann H (2000) *Lokale Algebren und Operatorprodukte am Punkt* (in German). Ph.D. thesis, Göttingen.
- Brunetti R and Fredenhagen K (2004) Microlocal Analysis and Interacting Quantum Field Theories: Renormalization on Physical Backgrounds. *Communications in Mathematical Physics* 208: 623.
- Dütsch and Fredenhagen K (2004) Causal Perturbation Theory in terms of retarded products, and a proof of the Action Ward Identity, to appear in. *Rev. Math. Phys.*
- Fewster CJ (2003) *Energy Inequalities in Quantum Field Theory*. Proceedings of the International Conference on Mathematical Physics (revised version under math-ph/0501073).
- Glimm J and Jaffe A (1987) *Quantum Physics: A Functional Integral Point of View*, 2nd edn. Berlin–Heidelberg–New York: Springer.
- Guido D and Longo R (1995) An algebraic spin and statistics Theorem. *Communications in Mathematical Physics* 172: 517.
- Haag R (1992) *Local Quantum Physics*. Berlin–Heidelberg–New York: Springer.
- Jost R (1965) *The General Theory of Quantized Fields*. American Mathematical Society.
- Kuckert B (2002) Covariant thermodynamics of quantum systems: passivity, semipassivity, and the Unruh effect. *Annals of Physics* 295: 216.
- Kuckert B (2005) Spin, statistics, and reflections, I. *Annales Henri Poincaré* 6: 849.
- Kuckert B and Lorenzen R (2005) Spin, Statistics, and Reflections, II. Preprint (math-ph/0512068).
- Osterwalder K and Schrader R (1973) Axioms for Euclidean Green's functions. *Communications in Mathematical Physics* 31: 83.
- Osterwalder K and Schrader R (1975) Axioms for Euclidean Green's functions. 2. *Communications in Mathematical Physics* 42: 281.
- Reed M and Simon B (1970) *Methods of Modern Mathematical Physics*, (4 volumes). London: Academic Press.
- Scharf G (1989) *Finite Quantum Electrodynamics*. Berlin–Heidelberg–New York: Springer.
- Scharf G (2001) *Quantum Gauge Theories A True Ghost Story*. Weinheim: Wiley.
- Streater RF (1975) Outline of axiomatic quantum field theory. *Reports on Progress in Physics* 38: 771.
- Streater RF and Wightman AS (1964) *PCT, Spin & Statistics, and All That*. New York: Benjamin.
- Völkel AH (1977) Fields, Particles, and Currents. *Lecture Notes in Physics*, vol. 66. Berlin–Heidelberg–New York: Springer.
- Wollenberg M (1988) The existence of quantum fields for local nets of observables. *Journal of Mathematical Physics* 29: 2106.
- Zinoviev YM (1995) Equivalence of Euclidean and Wightman field theories. *Communications in Mathematical Physics* 174: 1.

B

Bäcklund Transformations

D Levi, Università "Roma Tre", Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Bäcklund transformations appeared for the first time in the work of the geometers of the end of the nineteenth century, for instance, Bianchi, Lie, Bäcklund, and Darboux, when studying surfaces of constant curvature. If on a surface in three-dimensional Euclidean space, the asymptotic directions are taken as coordinate directions, then the surface metric may be written as

$$ds^2 = dx^2 + 2 \cos(w) dx dy + dy^2 \quad [1]$$

where $w(x, y)$ is a function of the surface coordinates x, y . A necessary and sufficient condition for the surface to be of constant curvature is that w satisfies the nonlinear partial differential equation

$$w_{,xy} = \sin(w) \quad [2]$$

where the subscript denotes partial derivative. Equation [2] is nowadays called the sine Gordon (sG) equation. Bianchi (1879), Lie (1888, 1890, 1893), and Bäcklund (1874) introduced a transformation which allows one to pass from a solution of eqn [2] to a new solution, that is, from a surface of constant curvature to a new one. Starting from the work of Clarin (1903), this transformation has been referred to as Bäcklund transformation (BT). The BT for eqn [2] reads

$$\tilde{w}_{,x} = w_{,x} + 2a \sin\left(\frac{\tilde{w} + w}{2}\right) \quad [3a]$$

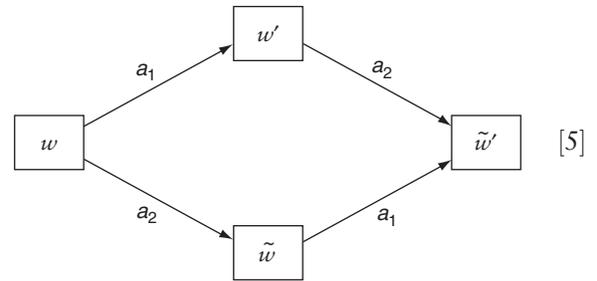
$$\tilde{w}_{,y} = -w_{,y} + \frac{2}{a} \sin\left(\frac{\tilde{w} - w}{2}\right) \quad [3b]$$

where a is a nonzero constant parameter and \tilde{w} is a different solution of eqn [2]. It is immediate to prove by appropriate differentiation of eqns [3] with respect to y and x that both w and \tilde{w} must satisfy eqn [2]. The BT [3] provides a denumerable set of exact solutions once a solution w is known. Bianchi

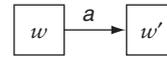
showed that four such solutions can be related in an algebraic way:

$$\tan\left(\frac{\tilde{w}' - w}{4}\right) = \frac{a_1 + a_2}{a_1 - a_2} \tan\left(\frac{w' - \tilde{w}}{4}\right) \quad [4]$$

Equation [4] is derived using the permutability theorem proved by Bianchi in his Ph.D. thesis in 1879:



whereby the diagram



we mean a BT from w to w' with parameter a .

For sG equation [2] a trivial solution is given, for example, by $w(x, y) = \pi$. Then, from eqn [3a] we get

$$\tilde{w}(x, y) = 2 \arcsin\left(\frac{1 - e^{-2[ax+\alpha(y)]}}{1 + e^{-2[ax+\alpha(y)]}}\right)$$

Introducing this result in eqn [3b], we get $\alpha_{,y} = -1/a$. So, the application of the BT [3] to sG equation gives the nontrivial solution

$$w = \pi \rightarrow \tilde{w} = 4 \arctan\left(\frac{1 - e^{-[ax-y/a]}}{1 + e^{-[ax-y/a]}}\right) \quad [6]$$

Clarin (1903) extended the results of Bäcklund to the case of a generic partial differential equation of second order,

$$F(x, y, w, w_{,x}, w_{,y}, w_{,xx}, w_{,xy}, w_{,yy}) = 0 \quad [7]$$

by assuming that

$$\begin{aligned} w_{,x} &= f(w, \tilde{w}, \tilde{w}_{,x}, \tilde{w}_{,y}) \\ w_{,y} &= g(w, \tilde{w}, \tilde{w}_{,x}, \tilde{w}_{,y}) \end{aligned} \quad [8]$$

If the compatibility of eqns [8]

$$f_{,y} - g_{,x} = 0 \tag{9}$$

is identically satisfied by eqn [7] for the variable $\tilde{w}(x,y)$, then we say that eqns [8] are an auto-Bäcklund transformation for eqn [7]. In this case, eqns [8] transform a solution of eqn [7] into a new solution of the same equation. Thus, eqns [8] simplify the problem of finding solutions of eqn [7]. Given one solution $w(x,y)$ of eqn [7], the existence of a BT reduces the problem of integrating eqn [7] into that of solving two first-order ordinary differential equations. From this point of view, the Cauchy–Riemann relations

$$w_{,x} = \tilde{w}_{,y}, \quad w_{,y} = -\tilde{w}_{,x} \tag{10}$$

for the Laplace equation

$$w_{,xx} + w_{,yy} = 0 \tag{11}$$

are a BT *ante litteram* (however, without a free parameter).

Consider the case when $\tilde{w}(x,y)$ satisfies a different partial differential equation,

$$G(x,y, \tilde{w}, \tilde{w}_{,x}, \tilde{w}_{,y}, \tilde{w}_{,xx}, \tilde{w}_{,xy}, \tilde{w}_{,yy}) = 0 \tag{12}$$

In this case, one still has a BT, but not an auto-BT. The best-known cases are when $F_1 = w_{,y} + w_{,xxx} + ww_{,x}$ and $G_1 = \tilde{w}_{,y} + \tilde{w}_{,xxx} + \tilde{w}^2 \tilde{w}_{,x}$, and $F_2 = w_{,xy} - e^w$ and $G_2 = \tilde{w}_{,xy}$ (Lamb 1976). In the first case, the BT relates the Korteweg–de Vries (KdV) equation to the modified KdV equation and this transformation paved the way to the discovery of the complete integrability of the KdV equation by Gardner *et al.* (1967). In the second case, the BT relates the Liouville equation to the wave equation, and can be used to solve it completely. Due to the first example, often a non-auto-BT is denoted as Miura transformation.

One can now state an operative definition of BT, extending the results of Bäcklund and Clariin to more general equations.

Definition 1 Consider two partial differential equations of order m_1 and m_2 :

$$F_1(\mathbf{x}, \mathbf{u}, \mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \dots, \mathbf{u}_{(m_1)}) = 0 \tag{13a}$$

$$F_2(\mathbf{x}, \tilde{\mathbf{u}}, \tilde{\mathbf{u}}_{(1)}, \tilde{\mathbf{u}}_{(2)}, \dots, \tilde{\mathbf{u}}_{(m_2)}) = 0 \tag{13b}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $(\mathbf{u}, \tilde{\mathbf{u}}) \in \mathbb{C}^p$, and \mathbf{u} is the set of k -order derivative of \mathbf{u} . The set of n equations

$$G_j(\mathbf{x}, \mathbf{u}, \mathbf{u}_{(1)}, \dots, \mathbf{u}_{(s_1)}; \tilde{\mathbf{u}}, \tilde{\mathbf{u}}_{(1)}, \dots, \tilde{\mathbf{u}}_{(s_2)}) = 0 \tag{14}$$

$j = 1, 2, \dots, n$

with $s_1 < m_1$ and $s_2 < m_2$, represents the BT of eqns [13] iff the compatibility of eqns [14] is identically satisfied on the solutions of eqns [13] and G_j depends on a set of essential arbitrary constant parameters.

The Clariin formulation [8] and the classical BT for the sG [3] are clearly special subcases of this definition. When a solution of $F_1 = 0$ is known, a solution of $F_2 = 0$ is obtained by solving a set of lower-order partial differential equations. By a proper choice of the BT parameters, once a new solution is obtained by solving the BT [14], one can use the obtained solution as a starting point to construct another one, and so on. In this way, one can construct a whole ladder of solutions, *a priori* a denumerable set of solutions. This same construction has been applied also to the case of functional equations. In particular, it has been considered for the case of differential–difference and difference–difference equations both for finite (dynamical systems (Wojciechowski 1982)) and infinite lattices (Toda 1989).

In the case when F_1 and F_2 represent the same equation, $s_1 = s_2 = 1$ and the BTs $G_j = 0$ are linear in $\mathbf{u}_{(1)}$, then Definition 1 is strictly related to the notion of nonclassical symmetry or conditional symmetry (Levi and Winternitz 1989, Olver 1993), an extension of the concept of Lie symmetry used to reduce and integrate a differential equation. In the case of the nonclassical symmetries, the known solution $\tilde{\mathbf{u}}$ is included in the arbitrary \mathbf{x} -dependent coefficients of the transformation. In this case, the BT is just a way to construct an explicit solution of the differential equation [7].

Definition 1 is often too general to be able to get explicit results. It is constructive for any partial differential equation, linear or nonlinear, but if one is not able to get a nontrivial BT this does not mean that a BT does not exist. As noted later, the existence of an auto-BT is associated to the existence of an infinity of symmetries, and this is a condition for the exact integrability of eqn [13] (Fokas 1980, Ibragimov and Shabat 1980). So, the existence of a BT is closely related to the integrability of eqn [13].

Bäcklund via Integrability

One can derive the BT from the integrability properties of eqn [13a]. Equation [13a] is said to be integrable if it can be written as the compatibility condition of an overdetermined system of linear partial differential equations for an auxiliary function depending on a free parameter belonging to the

complex C plane. The prototype of such a situation is given by the Lax pair for the KdV equation

$$u_{,t} + u_{,xxx} - 6uu_{,x} = 0 \quad [15]$$

introduced by Lax (1968):

$$L\psi = k^2\psi, \quad L = -\partial_x^2 + u(x, t) \quad [16a]$$

$$\psi_{,t} = -M\psi, \quad M = 4\partial_{xxx} - 3(u\partial_x + \partial_x u) \quad [16b]$$

where k is a free parameter and $\psi = \psi(x, t; k)$. As eqn [16a] is nothing else but the stationary Schrödinger equation, the function ψ can be interpreted as a wave function, and k^2 is the spectral parameter corresponding to the potential $u(x, t)$. The condition for the existence of a solution ψ of the overdetermined system of eqns [16] is given by the operator equation

$$L_{,t} = [L, M] \quad [17]$$

the so-called Lax equation. In the case of asymptotically bounded potentials, eqn [16a] defines the spectrum unique. Introducing the following asymptotic boundary conditions for the wave function ψ ,

$$\begin{aligned} \psi(x, t; k) &\xrightarrow{x \rightarrow -\infty} T(k, t)e^{-ikx} \\ \psi(x, t; k) &\xrightarrow{x \rightarrow +\infty} e^{-ikx} + R(k, t)e^{ikx} \end{aligned} \quad [18]$$

where $R(k, t)$ and $T(k, t)$ are, respectively, the reflection and the transmission coefficient, the spectrum is defined in the complex plane of the variable k by

$$S[u] \equiv \{R(k, t), -\infty < k < \infty; p_n, c_n(t), j = 1, 2, \dots, N\} \quad [19]$$

where p_n are the bound state parameters corresponding to isolated singularities of the reflection coefficients on the imaginary positive k -axis corresponding to a solution $\phi_n(x, t; p_n)$ of the spectral problem vanishing for $x \rightarrow -\infty$ and such that

$$\lim_{x \rightarrow +\infty} [e^{p_n x} \phi_n(x, t; p_n)] = 1 \quad [20]$$

and c_n are some functions of t related to the residues of $R(k, t)$ at the poles p_n . There is a one-to-one correspondence between the evolution of the potential $u(x, t)$ in eqn [15] and that of the spectrum $S[u]$ of the Schrödinger spectral problem [16a]. In particular, for the KdV, taking into account eqn [16b], the evolution of the reflection coefficient $R(k, t)$ is given by

$$\frac{dR(k, t)}{dt} = 8ik^3 R(k, t) \quad [21]$$

In eqn [21] and henceforth, d/dt denotes the total derivative with respect to t .

In the following, for the sake of the simplicity of exposition and for the concreteness of the presentation, all the results presented on the BT will be derived for the KdV equation. Similar results can be obtained and have been obtained in the literature for many classes of integrable partial differential equations in two and three dimensions and for differential–difference and difference–difference equations. For a partial review of the available recent literature on the subject, see Rogers and Shadwick (1982) and Coley *et al.* (2001)

A more general form of introducing the nonlinear partial differential equation as a compatibility of an overdetermined system of linear equations has been provided by Zaharov and Shabat (1979) with the dressing method (DM). In the DM, the differential equations [16] are substituted by a matrix system of linear equations

$$\Psi_{,x} = U(u(x, t), k)\Psi \quad [22a]$$

$$\Psi_{,t} = V(u(x, t), k)\Psi \quad [22b]$$

where $\Psi = \Psi(x, t; k)$ and U and V are matrix functions. The existence of a nonsingular solution of the system of linear equations [22] requires that the matrix functions U and V satisfy the equation

$$U_{,t} - V_{,x} + [U, V] = 0 \quad [23]$$

often called zero-curvature condition. The KdV equation [15] in the DM is obtained by choosing

$$U(u(x, t), k) = \begin{pmatrix} ik & u(x, t) \\ 1 & -ik \end{pmatrix}$$

$$\begin{aligned} V(u(x, t), k) &= \begin{pmatrix} 2u + 4k^2 & -u_x - 2iku - 4ik^3 \\ u_x + 2iku + 4ik^3 & 2u(u + 2k^2) - 2iku_x - u_{,xx} \end{pmatrix} \\ & \quad [24] \end{aligned}$$

The existence of an auto-BT implies the existence of a differential equation (see Definition 1) which relates two solutions of the same nonlinear equation. The new solution $\tilde{u}(x, t)$ of eqn [15] will be associated to a different Lax operator and a different spectral problem (but of the same operational form)

$$\tilde{L} = -\partial_{xx} + \tilde{u}(x, t) \quad [25a]$$

$$\tilde{L}\tilde{\psi} = k^2\tilde{\psi} \quad [25b]$$

The existence of a relation between the potentials $u(x, t)$ and $\tilde{u}(x, t)$ thus implies that there must be a $(u, \tilde{u}; k)$ -dependent operator D such that

$$\tilde{\psi} = D\psi \quad [26]$$

The compatibility of eqns [16a], [25b], and [26] implies that $\tilde{L}D\psi = Dk^2\psi$, that is,

$$\tilde{L}D = DL \quad [27]$$

Equation [27] is the auto-BT in the Lax formalism. If \tilde{L} and L are two different spectral problems related to two different nonlinear partial differential equations, then eqn [27] will provide a Miura transformation. In the DM, the requirement of the existence of a BT is given again by eqn [26] with ψ and $\tilde{\psi}$ substituted by Ψ and $\tilde{\Psi}$ and the operator D substituted by a matrix function \mathcal{D} . The BT in the DM is given by

$$\mathcal{D}_{,x} = U(\tilde{u}(x, t), k)\mathcal{D} - \mathcal{D}U(u(x, t), k) \quad [28a]$$

$$\mathcal{D}_{,t} = V(\tilde{u}(x, t), k)\mathcal{D} - \mathcal{D}V(u(x, t), k) \quad [28b]$$

In the particular case of the Hilbert–Riemann problem with zeros, providing the soliton solutions, the matrix \mathcal{D} can be expressed as a function of Ψ . In this way, one derives the Moutard or Darboux transformation (DT) (Moutard 1878, Levi *et al.* 1984), the most efficient way to get soliton solutions of the nonlinear partial differential equation.

Given a linear ordinary differential equation for the unknown ψ , depending on a set of arbitrary functions $u(x)$ and parameters k , the DT provides a discrete transformation which leaves the equation invariant. In the particular case of the KdV equation associated with the stationary Schrödinger spectral problem [16a], we have

$$\tilde{u}(x, t) = u(x, t) - 2(\log F(x, t))_{,xx} \quad [29a]$$

$$\begin{aligned} \tilde{\psi}(x, t; k) = & -\frac{i}{k + ip} \psi_{,x}(x, t; k) \\ & - \frac{F_x(x, t)}{F(x, t)} \psi(x, t; k) \end{aligned} \quad [29b]$$

where the intermediate wave function

$$F(x, t) = \psi(x, t; k = ip) + a\psi(x, t; k = -ip)$$

is a linear combination of the Jost solution of the Schrödinger spectral problem with p a real parameter and a an arbitrary constant. If one looks for an equation involving only the potentials u and \tilde{u} , from eqns [29], one gets the BT for the KdV equation. Given a trivial solution of the KdV equation, together with the corresponding solution

of the spectral problem, eqn [29a] provides a new solution of the KdV, while eqn [29b] gives a new solution of the spectral problem. This procedure can be carried out recursively and gives a ladder of explicit solutions for the KdV equation.

The DM is a particularly simple setting in which one can derive DTs. In fact, expressing the matrix \mathcal{D} in terms of Ψ , eqn [28a] gives a relation between the potentials of the type given by eqn [29a], while eqn [26] gives eqn [29b]. Depending on the form of the matrix \mathcal{D} in terms of k , one can introduce more parameters in the DT. The classical DT [29] depends on just one parameter; however, in the case of the Schrödinger spectral problem [16a], one can also have DTs depending on two parameters, a TDT.

A more general DT, which can provide solutions even when the initial solution is not bounded asymptotically, can be obtained for many equations and, in particular, also for the KdV equation. This is obtained in a particular limit of the TDT when the parameters coincide (Levi 1988) and it is often referred to as binary DT (Matveev and Salle 1991). The binary DT for the KdV is given by

$$\tilde{u}(x, t) = u(x, t) - 2(\log F(x, t))_{,xxx} \quad [30a]$$

$$\begin{aligned} \tilde{\psi}(x, t; k) = & \frac{1}{k^2 - \mu^2} \left\{ \left(k^2 - \mu^2 - \frac{F(x, t)_{,xx}}{2F(x, t)} \right) \psi_{,x}(x, t; k) \right. \\ & \left. - \frac{F_x(x, t)}{F(x, t)} \psi(x, t; k) \right\} \end{aligned} \quad [30b]$$

where μ is a value of k for which the function $\psi(x, t; k)$ is asymptotically bounded at $+\infty$ and the function $F(x, t)$ is given by

$$F(x, t) = 1 + \rho \int_x^{+\infty} \psi(y, t; \mu)^2 dy \quad [31]$$

with ρ an arbitrary constant. The corresponding BT obtained eliminating the function F from eqns [30] reads

$$\begin{aligned} \tilde{q}_{,xx} - q_{,xx} = & -\frac{1}{8}(\tilde{q} - q)^3 \\ & - [\tilde{q}_x + q_x - 2g(x) + 2\mu](\tilde{q} - q) \\ & + \frac{1}{2} \frac{(\tilde{q}_x - q_x)^2}{\tilde{q} - q} \end{aligned} \quad [32]$$

where $q = \int_x^\infty u_0(y, t) dy$ with $u_0(x, t) = u(x, t) - g(x)$, the asymptotically bounded part of $u(x, t)$, and $g(x)$ its asymptotic behavior, and $\tilde{q} = \int_x^\infty \tilde{u}_0(y, t) dy$ with $\tilde{u}_0(x, t) = \tilde{u}(x, t) - g(x)$.

Once the Lax operator L is given, we can obtain in a constructive way the operators M which give the admissible nonlinear partial differential

equations and the operators D which give the admissible BT. A technique to do so is provided by the so-called Lax technique introduced by Bruschi and Ragnisco (1980a–c). Using the Lax technique, we can easily obtain the nonlinear partial differential equations and BT associated with the Lax operator [16a] both in the isospectral and non-isospectral case (when $k_t=0$ and when $k_t \neq 0$) and the corresponding evolution of the spectrum. We have

$$u_{,t} = f(\mathcal{L}, t)u_x + g(\mathcal{L}, t)[xu_x + 2u] \quad [33a]$$

$$k_t = kg(-4k^2, t) \quad [33b]$$

$$\frac{dR(k, t)}{dt} = 2ikf(-4k^2, t)R(k, t)$$

$$F(\Lambda)(\tilde{u} - u) + G(\Lambda)\Gamma 1 = 0 \quad [33c]$$

$$\tilde{R}(k, t) = \frac{F(-4k^2) - 2ikG(-4k^2)}{F(-4k^2) + 2ikG(-4k^2)}R(k, t) \quad [33d]$$

where the functions $f, g, F,$ and G are entire functions of their first argument and the recursive operators \mathcal{L} and Λ are given by

$$\mathcal{L}f(x) = f_{,xx}(x) - 4u(x, t)f(x) + 2u_{,x}(x, t) \int_x^{+\infty} f(y) dy \quad [34a]$$

$$\Lambda f(x) = f_{,xx}(x) - 2[\tilde{u}(x, t) + u(x, t)]f(x) + \Gamma \int_x^{+\infty} f(y) dy \quad [34b]$$

$$\Gamma f(x) = [\tilde{u}_{,x}(x, t) + u_{,x}(x, t)]f(x) + [\tilde{u}(x, t) - u(x, t)] \times \int_x^{+\infty} [\tilde{u}(y, t) - u(y, t)]f(y) dy \quad [34c]$$

In the limit when $\tilde{u} \rightarrow u$ the operator $\Lambda \rightarrow \mathcal{L}$. A BT is obtained by choosing the functions F and G in eqn [33c]. The simplest BT is obtained by setting $F = \sigma$ and $G = 1$:

$$\tilde{v}_x + v_x + (\tilde{v} - v)[\sigma - \frac{1}{2}(\tilde{v} - v)] = 0 \quad [35]$$

with $u(x, t) = -v_{,x}(x, t)$ and σ is the Bäcklund parameter. By combining together BT of the form [35] with different parameters as in eqn [5], we get the permutability theorem for the KdV BTs:

$$\tilde{v}' = v - \frac{(\sigma_1 + \sigma_2)[v' - \tilde{v}]}{\sigma_1 - \sigma_2 + (1/2)(v' - \tilde{v})} \quad [36]$$

Its proof is immediate from the point of view of the spectrum.

Bäcklund and Symmetries

A symmetry of the nonlinear equation [15] is given by a flow commuting with it, that is, by an equation

$$u_{,\epsilon} = f(u, u_x, u_t, \dots) \quad [37]$$

where ϵ is the group parameter, $u = u(x, t; \epsilon)$, and the ϵ derivative of [15] is zero on its set of solutions. A group transformation is obtained by integrating it. Usually this is possible only when eqn [37] is a quasilinear partial differential equation of the first order. Taking into account the evolution of the spectrum of the KdV equation [15], it is easy to prove that its symmetries are given by

$$u_{,\epsilon} = \left\{ \sum_{n=0}^{+\infty} \alpha_n \mathcal{L}^n - 3 \sum_{n=0}^{+\infty} \beta_n t \mathcal{L}^n \right\} u_{,x} + \left\{ \sum_{n=0}^{+\infty} \beta_n \mathcal{L}^n \right\} [xu_{,x} + 2u] \quad [38]$$

where α_n and β_n are a set of constant parameters. For each choice of the parameters α_n and β_n , one gets a symmetry of the KdV equation [15]. With eqn [38] one can associate the following evolution of the reflection coefficient $R(k, t; \epsilon)$:

$$\frac{dR}{d\epsilon} = 2ik \left\{ \sum_{n=0}^{+\infty} \alpha_n (-4k^2)^n - 3 \sum_{n=0}^{+\infty} \beta_n t (-4k^2)^{n+1} \right\} R \quad [39]$$

and of the spectral parameter k

$$k_{,\epsilon} = \sum_{n=0}^{+\infty} \beta_n (-4k^2)^n k \quad [40]$$

As $-(1/2)\mathcal{L} 1 = xu_{,x} + 2u$, one can add to the symmetries [38] the exceptional one (which has no spectral counterpart as u is not bounded asymptotically):

$$u_{,\epsilon} = 1 + 6tu_{,x} \quad [41]$$

By a proper natural choice of the constant parameters α_n and β_n , one can define two infinite series of symmetries. The first one is obtained by choosing $\beta_n = 0$ and $\alpha_n = \delta_{n,m}$ with $m = 1, 2, \dots, \infty$ and can be denoted as the isospectral series as $k_{,\epsilon} = 0$. This is formed by commuting symmetries. The second one is given by $\alpha_n = 0$ and $\beta_n = \delta_{n,m}$ with $m = 1, 2, \dots, \infty$ and can be denoted as the nonisospectral series as $k_{,\epsilon} \neq 0$. The nonisospectral symmetries have a nonzero commutation relation among themselves and with the isospectral ones.

Except for a few Lie point symmetries (given by eqn [41] and by choosing inside the series [38] those with different from zero only β_0 or α_0 or α_1) they are all generalized symmetries (Olver 1993). By analyzing their spectrum, it is easy to prove that the choice [38] is such that they are all independent. For the isospectral class, the evolution of the spectrum is simple and can be integrated to provide the group transformation of the spectrum

$$R(k, t; \epsilon) = R(k, t) \times \exp \left[2ik \left\{ \sum_{n=0}^{+\infty} \alpha_n (-4k^2)^n \right\} \epsilon \right] \quad [42]$$

Let us now consider the simplest BT obtained by choosing, in eqn [33c], $F(\Lambda) = \sigma$ and $G(\Lambda) = 1$, where σ is an arbitrary parameter. In the spectral space, this corresponds to the following change of the spectrum:

$$\tilde{R}(k, t) = \frac{\sigma - 2ik}{\sigma + 2ik} R(k, t) \quad [43]$$

Defining $\tilde{R}(k, t) = R(k, t; \epsilon)$, eqn [42] is equal to eqn [43] iff

$$\alpha_n = -\frac{2}{\epsilon \sigma^{2n+1} (2n+1)}, \quad n = 0, 1, \dots, \infty \quad [44]$$

So we need an infinite number of symmetries to be able to reconstruct the change of the spectrum given by the BT. This shows that the existence of a BT is strictly connected to the existence of an infinity of symmetries which is a condition for the exact integrability of the nonlinear partial differential equation (Fokas 1980, Ibragimov and Shabat 1980).

Discretization via Bäcklund

BTs, apart from providing classes of exact solutions to nonlinear equations, play a very important role in the discretization of partial differential equations. As noted earlier, an auto-BT is a differential relation between two different solutions of the same nonlinear partial differential equation. If it is assumed that the new solution \tilde{u} is just the old solution u computed in a different point of a lattice, then the BT becomes just a differential–difference equation (Chiu and Ladik 1977, Levi and Benguria 1980). This can be carried out also at the level of the associated compatibility condition and in such a way one is able to also obtain its Lax pair. This demonstrates the integrability of the differential–difference equation

$$v(n+1, t)_{,t} + v(n, t)_{,t} + [v(n+1, t) - v(n, t)] \times \left\{ \sigma - \frac{1}{2}[v(n+1, t) - v(n, t)] \right\} = 0 \quad [45]$$

which is an integrable differential–difference approximation to the KdV equation or

$$w(n+1, t)_{,t} = w(n, t)_{,t} + 2a \sin \left[\frac{w(n+1, t) + w(n, t)}{2} \right] \quad [46]$$

a discrete integrable differential–difference approximation to the sG equation (Hirota 1977, Orfanidis 1978).

As the nonlinear superposition formulas are purely algebraic relations involving potentials associated with integrable nonlinear partial differential equations, one can interpret them as difference–difference equations. In the case of the sG equation from eqn [7], we have

$$w_{n+1, m+1} - w_{n, m} = 4 \arctan^{-1} \left(\frac{a_1 + a_2}{a_1 - a_2} \tan \frac{w_{n, m+1} - w_{n+1, m}}{4} \right) \quad [47]$$

where $w(x, t) = w_{n, m}$, $\tilde{w}(x, t) = w_{n+1, m}$, $w'(x, t) = w_{n, m+1}$, and $\tilde{w}'(x, t) = w_{n+1, m+1}$. In a similar manner, from [36], one gets

$$v_{n+1, m+1} = v_{n, m} - \frac{(\sigma_1 + \sigma_2)[v_{n+1, m} - v_{n, m+1}]}{\sigma_1 - \sigma_2 + \frac{1}{2}[v_{n+1, m} - v_{n, m+1}]} \quad [48]$$

The continuous limit of eqn [47], obtained by setting $x = \epsilon_1 n$ and $y = \epsilon_2 m$ and choosing

$$\frac{a_1}{a_2} = \frac{\epsilon_1 \epsilon_2}{4}$$

gives back eqn [2] (Rogers and Schief 1997). It is worth mentioning that one can also use known nonlinear lattice equations to construct BT for nonlinear partial differential equations (Levi 1981).

See also: Integrable Systems and Discrete Geometry; Integrable Systems: Overview; Painlevé Equations; Solitons and Kac–Moody Lie Algebras; Toda Lattices.

Further Reading

- Bäcklund AV (1874) Einiges über Curven und Flächentransformationen. *Lund Universitëts Arsskrift* 10: 1–12.
- Bianchi L (1879) Ricerche sulle superficie a curvatura costante e sulle elicoidi. *Annali della R. Scuola normale superiore di Pisa* 2: 285.
- Bruschi M and Ragnisco O (1980a) Existence of a Lax pair for any member of the class of nonlinear evolution equations associated to the matrix Schrödinger spectral problem. *Lettere al Nuovo Cimento* 29: 321–326.
- Bruschi M and Ragnisco O (1980b) Extension of the Lax method to solve a class of nonlinear evolution equations with x -dependent coefficients associated to the matrix Schrödinger spectral problem. *Lettere al Nuovo Cimento* 29: 327–330.
- Bruschi M and Ragnisco O (1980c) Bäcklund transformations and Lax technique. *Lettere al Nuovo Cimento* 29: 331–334.
- Chiu S-C and Ladik JF (1977) Generating exactly soluble nonlinear discrete evolution equations by a generalized

- Wronskian technique. *Journal of Mathematical Physics* 18: 690–700.
- Clarín J (1903) Sur quelques équations aux dérivées partielles du second ordre. *Annales de la Faculté des Sciences de Toulouse pour les Sciences Mathématiques et les Sciences Physiques. Serie 2* 5: 437–458.
- Coley A, Levi D, Milson R, Rogers C, and Winternitz P (eds.) (2001) Bäcklund and Darboux transformations. The Geometry of solitons. *Proceedings of the AARMS-CRM Workshop, Halifax, NS, June 4–9, 1999*. CRM Proceedings and Lecture Notes, vol. 29. Providence, RI: American Mathematical Society.
- Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Berlin: Springer.
- Fokas AS (1980) A symmetry approach to exactly solvable evolution equations. *Journal of Mathematical Physics* 21: 1318–1325.
- Gardner CS, Greene JM, Kruskal MD, and Miura RM (1967) Method for solving the Korteweg–de Vries equation. *Physical Review Letters* 19: 1095–1097.
- Hirota R (1978) Nonlinear partial difference equations. III. Discrete sine-Gordon equation. *Journal of the Physical Society of Japan* 43: 2079–2086.
- Ibragimov NH and Shabat AB (1980) Infinite Lie–Bäcklund algebras (in Russian). *Funktsional. Anal. i Prilozhen* 14: 79–80.
- Lamb GL (1976) Bäcklund transformations at the turn of the century. In: Miura RM (ed.) *Bäcklund Transformations*, pp. 69–79. Berlin: Springer.
- Lax PD (1968) Integrals of nonlinear equations of evolution and solitary waves. *Communications in Pure and Applied Mathematics* 21: 647–690.
- Levi D (1981) Nonlinear differential difference equations as Bäcklund transformations. *Journal of Physics A: Mathematical and General* 14: 1083–1098.
- Levi D (1988) On a new Darboux transformation for the construction of exact solutions of the Schrödinger equation. *Inverse Problems* 4: 165–172.
- Levi D and Benguria R (1980) Bäcklund transformations and nonlinear differential difference equations. *Proceedings of the National Academy of Science USA* 77: 5025–5027.
- Levi D and Winternitz P (1989) Non-classical symmetry reduction: example of the Boussinesq equation. *Journal of Physics A: Mathematical and General* 22: 2915–2924.
- Levi D, Ragnisco O, and Sym A (1984) Dressing method vs. classical Darboux transformation. *Il Nuovo Cimento* 83B: 34–42.
- Lie S (1888, 1890, 1893) *Theorie der Transformationsgruppen*. Leipzig: B.G. Teubner.
- Matveev VB and Salle LA (1991) *Darboux Transformations and Solitons*. Berlin: Springer.
- Moutard Th-F (1878) Sur la construction des équations de la forme $(1/z)(d^2z/dxdy) = \lambda(x,y)$, qui admettent une integrale général explicite. *Journal de l'Ecole Polytechnique, Paris* 28: 1–11.
- Olver PJ (1993) *Applications of Lie Groups to Differential Equations*. New York: Springer.
- Orfanidis SJ (1978) Discrete sine-Gordon equations. *Physical Review D* 18: 3822–3827.
- Rogers C and Schief WK (1997) The classical Bäcklund transformation and integrable discretization of characteristic equations. *Physics Letters A* 232: 217–223.
- Rogers C and Shadwick WF (1982) *Bäcklund Transformations and Their Applications*. New York: Academic Press.
- Toda M (1989) *Theory of Nonlinear Lattices*. Berlin: Springer.
- Wojciechowski S (1982) The analogue of the Bäcklund transformation for integrable many-body systems. *Journal of Physics A: Mathematical and General* 15: L653–L657.
- Zaharov VE and Shabat AB (1979) Integration of the nonlinear equations of mathematical physics by the method of the inverse scattering problem. II (Russian). *Funktsional Analiz i ego Prilozheniya* 13: 13–22. (English translation: *Functional Analysis and Applications* 13: 166–173 (1980)).

Batalin–Vilkovisky Quantization

A C Hirshfeld, Universität Dortmund,
Dortmund, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Batalin–Vilkovisky formalism for quantizing gauge theories has a long history of development. It begins with the Faddeev–Popov procedure for quantizing Yang–Mills theory, involving the Faddeev–Popov ghost fields (Faddeev and Popov 1967). It continued with the discovery of BRST symmetry by Becchi *et al.* (1976). Then Zinn-Justin (1975) introduced sources for these transformations, and a symmetric structure in the space of fields and sources in his study of renormalizability of these theories. Finally, Batalin and Vilkovisky (1981) systematized and generalized these developments. A more detailed account of this history can be found in Gomis *et al.* (1994), where many worked

examples of the Batalin–Vilkovisky formalism are given. At the present time, it is the most general treatment available. Alexandrov, Kontsevich, Schwarz, and Zabarovsky (AKSZ 1997) have presented a geometric interpretation for the case in which the action is topologically invariant.

Structure of the Set of Gauge Transformations

Consider a system whose dynamics is governed by a classical action $S[\phi^i]$ which depends on the fields $\phi^i(x)$, $i = 1, \dots, n$. We employ a compact notation in which the multi-index i may denote the various fields involved, the discrete indices on which they depend, and the dependence on the spacetime variables as well. The generalized summation convention then means that a repeated index may denote not only a sum over discrete variables, but also integration over the spacetime variables. $\epsilon_i = \epsilon(\phi^i)$ denotes the

Grassmann parity of the fields. Fields with $\epsilon_i = 0$ are called bosonic, with $\epsilon_i = 1$ fermionic. The graded commutation rule is

$$\phi^i(x)\phi^j(y) = (-1)^{\epsilon_i\epsilon_j}\phi^j(y)\phi^i(x) \quad [1]$$

For a gauge theory the action is invariant under a set of gauge transformations with infinitesimal form

$$\delta\phi^i = R_\alpha^i \varepsilon^\alpha, \quad \alpha = 1 \text{ or } 2 \text{ or } \dots m \quad [2]$$

The ε^α are the infinitesimal gauge parameters and R_α^i the generators of the gauge transformations. When $\epsilon_\alpha = \epsilon(\varepsilon^\alpha) = 0$ we have an ordinary symmetry, when $\epsilon_\alpha = 1$ the equation is characteristic of a supersymmetry. The Grassmann parity of R_α^i is $\epsilon(R_\alpha^i) = \epsilon_i + \epsilon_\alpha \pmod{2}$.

A subscript after a comma denotes the right derivative with respect to the corresponding field, that is, the field is to be commuted to the far right and then dropped. The field equations may then be written as

$$S_{0,i} = 0 \quad [3]$$

where S_0 is the classical action. Let Σ denote the surface in the space of solutions where the field equations are satisfied:

$$S_{0,i}|_\Sigma = 0 \quad [4]$$

If the gauge transformations are “independent” on-shell, that is,

$$\text{rank } R_\alpha^i|_\Sigma = m \quad [5]$$

the gauge theory is said to be “irreducible.” We assume here that this is the case. When it is not, the theory is “reducible.” For details of the treatment in that case, see Gomis, Paris, and Samuel. The classical solutions are $\phi_0 \in \Sigma$.

The Noether identities are

$$S_{0,i}R_\alpha^i = 0 \quad [6]$$

The general solution to the Noether identity is

$$\lambda^i = R_\alpha^i T^\alpha + S_{0,j} E^{ji} \quad [7]$$

The commutator of two gauge transformations is

$$[\delta_1, \delta_2]\phi^i = \left(R_{\alpha,j}^i R_\beta^j - (-1)^{\epsilon_\alpha\epsilon_\beta} R_{\beta,j}^i R_\alpha^j \right) \varepsilon_1^\beta \varepsilon_2^\alpha \quad [8]$$

Since this commutator is a symmetry of the action, it satisfies the Noether identity

$$S_{0,i} \left(R_{\alpha,j}^i R_\beta^j - (-1)^{\epsilon_\alpha\epsilon_\beta} R_{\beta,j}^i R_\alpha^j \right) = 0 \quad [9]$$

which by eqn [7] implies that

$$R_{\alpha,j}^i R_\beta^j - (-1)^{\epsilon_\alpha\epsilon_\beta} R_{\beta,j}^i R_\alpha^j = R_\gamma^i T_{\alpha\beta}^\gamma + S_{0,j} E_{\alpha\beta}^{ji} \quad [10]$$

Equations [8] and [10] lead to the following condition:

$$[\delta_1, \delta_2]\phi^i = \left(R_\gamma^i T_{\alpha\beta}^\gamma - S_{0,j} E_{\alpha\beta}^{ji} \right) \varepsilon_1^\beta \varepsilon_2^\alpha \quad [11]$$

The tensors $T_{\alpha\beta}^\gamma$ are called the structure constants of the gauge algebra, although they depend, in general, on the fields of the theory. When $E_{\alpha\beta}^{ij} = 0$, the gauge algebra is said to be “closed,” otherwise it is “open.” Equation [11] defines a Lie algebra if the algebra is closed and the $T_{\alpha\beta}^\gamma$ are independent of the fields.

The gauge tensors have the following graded symmetry properties:

$$\begin{aligned} T_{\alpha\beta}^\gamma &= -(-1)^{\epsilon_\alpha\epsilon_\beta} T_{\beta\alpha}^\gamma \\ E_{\alpha\beta}^{ij} &= -(-1)^{\epsilon_\alpha\epsilon_\beta} E_{\beta\alpha}^{ij} = -(-1)^{\epsilon_\alpha\epsilon_\beta} E_{\beta\alpha}^{ji} \end{aligned} \quad [12]$$

The Grassmann parities are

$$\epsilon(T_{\alpha\beta}^\gamma) = \epsilon_\alpha + \epsilon_\beta + \epsilon_\gamma \pmod{2} \quad [13]$$

and

$$\epsilon(E_{\alpha\beta}^{ij}) = \epsilon_i + \epsilon_j + \epsilon_\alpha + \epsilon_\beta \pmod{2} \quad [14]$$

Various restrictions are imposed by the Jacobi identity

$$\sum_{\text{cyclic}(123)} [\delta_1, [\delta_2, \delta_3]] = 0 \quad [15]$$

These restrictions are

$$\sum_{\text{cyclic}(123)} \left(R_\delta^i A_{\alpha\beta\gamma}^\delta - S_{0,j} B_{\alpha\beta\gamma}^{ji} \right) \varepsilon^\gamma \varepsilon^\beta \varepsilon^\alpha = 0 \quad [16]$$

where

$$\begin{aligned} 3A_{\alpha\beta\gamma}^\delta &\equiv \left(T_{\alpha\beta k}^\delta R_\gamma^k - T_{\alpha\eta}^\delta T_{\beta\gamma}^\eta \right) + (-1)^{\epsilon_\alpha(\epsilon_\beta + \epsilon_\gamma)} \\ &\quad \times \left(T_{\beta\gamma k}^\delta R_\alpha^k - T_{\beta\eta}^\delta T_{\gamma\alpha}^\eta \right) \\ &\quad + (-1)^{\epsilon_\gamma(\epsilon_\alpha + \epsilon_\beta)} \left(T_{\gamma\alpha k}^\delta R_\beta^k - T_{\gamma\eta}^\delta T_{\alpha\beta}^\eta \right) \end{aligned}$$

and

$$\begin{aligned} 3B_{\alpha\beta\gamma}^{ji} &\equiv \left(E_{\alpha\beta k}^{ji} R_\alpha^k - E_{\alpha\delta}^{ji} T_{\beta\gamma}^\delta - (-1)^{\epsilon_i\epsilon_\alpha} \right. \\ &\quad \times R_{\alpha,k}^j E_{\beta\gamma}^{ki} + (-1)^{\epsilon_j(\epsilon_i + \epsilon_\alpha)} R_{\alpha,k}^i E_{\beta\gamma}^{kj} \left. \right) \\ &\quad + (-1)^{\epsilon_\alpha(\epsilon_\beta + \epsilon_\gamma)} (\alpha \rightarrow \beta \rightarrow \gamma) + (-1)^{\epsilon_\gamma(\epsilon_\alpha + \epsilon_\beta)} \\ &\quad \times (\alpha \rightarrow \gamma \rightarrow \beta) \end{aligned}$$

As in the familiar Faddeev–Popov procedure, it is useful to introduce ghost fields C^α with opposite Grassmann parities to the gauge parameters ε^α :

$$\epsilon(C^\alpha) = \epsilon_\alpha + 1 \pmod{2} \quad [17]$$

and to replace the gauge parameters by ghost fields. One must then modify the graded symmetry properties of the gauge structure tensors according to

$$T_{\alpha_1\alpha_2\alpha_3\alpha_4\dots} \rightarrow (-1)^{\epsilon^{\alpha_2+\alpha_4+\dots}} T_{\alpha_1\alpha_2\alpha_3\alpha_4\dots} \quad [18]$$

The Noether identities then take the form

$$S_{0,j} R_\alpha^i C^\alpha = 0 \quad [19]$$

and the structure relations [10] become

$$(2R_{\alpha,j}^i R_\beta^j - R_\gamma^i T_{\alpha\beta}^\gamma + S_{0,j} E_{\alpha\beta}^{jj}) C^\beta C^\alpha = 0 \quad [20]$$

Introducing the Antifields

We incorporate the ghost fields into the field set $\Phi^A = \{\phi^i, C^\alpha\}$, where $i = 1, \dots, n$ and $\alpha = 1, \dots, m$. Clearly $A = 1, \dots, N$, where $N = n + m$. One then further increases the set by introducing an antifield Φ_A^* for each field Φ^A . The Grassmann parity of the antifields is

$$\epsilon(\Phi_A^*) = \epsilon(\Phi^A) + 1 \pmod{2} \quad [21]$$

Each field is assigned a ghost number, with

$$\begin{aligned} \text{gh}[\phi^i] &= 0 \\ \text{gh}[C^\alpha] &= 1 \\ \text{gh}[\Phi_A^*] &= -\text{gh}[\Phi^A] - 1 \end{aligned} \quad [22]$$

In the space of fields and antifields, the antibracket is defined by

$$(X, Y) = \frac{\partial_r X}{\partial \Phi^A} \frac{\partial_l Y}{\partial \Phi_A^*} - \frac{\partial_r X}{\partial \Phi_A^*} \frac{\partial_l Y}{\partial \Phi^A} \quad [23]$$

where ∂_r denotes the right, ∂_l the left derivative. The antibracket is graded antisymmetric:

$$(X, Y) = -(-1)^{(\epsilon_X+1)(\epsilon_Y+1)} (Y, X) \quad [24]$$

It satisfies a graded Jacobi identity

$$\begin{aligned} &((X, Y), Z) + (-1)^{(\epsilon_X+1)(\epsilon_Y+1)} \\ &\times ((Y, Z), X) + (-1)^{(\epsilon_Z+1)(\epsilon_X+\epsilon_Y)} ((Z, X), Y) = 0 \end{aligned} \quad [25]$$

It is a graded derivation

$$\begin{aligned} (X, YZ) &= (X, Y)Z + (-1)^{\epsilon_X \epsilon_Y} (X, Z)Y \\ (XY, Z) &= X(Y, Z) + (-1)^{\epsilon_X \epsilon_Y} Y(X, Z) \end{aligned} \quad [26]$$

It has ghost number

$$\text{gh}[(X, Y)] = \text{gh}[X] + \text{gh}[Y] + 1 \quad [27]$$

and Grassmann parity

$$\epsilon((X, Y)) = \epsilon(X) + \epsilon(Y) + 1 \pmod{2} \quad [28]$$

For bosonic fields

$$(B, B) = 2 \frac{\partial B}{\partial \Phi^A} \frac{\partial B}{\partial \Phi_A^*} \quad [29]$$

for fermionic fields

$$(F, F) = 0 \quad [30]$$

and for any X

$$((X, X), X) = 0 \quad [31]$$

If one groups the fields and the antifields together into the set

$$z^a = \{\Phi^A, \Phi_A^*\}, \quad a = 1, \dots, 2N \quad [32]$$

then the antibracket is seen to define a symplectic structure on the space of fields and antifields

$$(X, Y) = \frac{\partial_r X}{\partial z^a} \omega^{ab} \frac{\partial_l Y}{\partial z^b} \quad [33]$$

with

$$\omega^{ab} = \begin{pmatrix} 0 & \delta_B^A \\ -\delta_B^A & 0 \end{pmatrix} \quad [34]$$

The antifields can be thought of as conjugate variables to the fields, since

$$(\Phi^A, \Phi_B^*) = \delta_B^A \quad [35]$$

The Classical Master Equation

Let $S[\Phi^A, \Phi_A^*]$ be a functional of the fields and antifields with the dimension of an action, vanishing ghost number and even Grassmann parity. The equation

$$(S, S) = 2 \frac{\partial S}{\partial \Phi^A} \frac{\partial S}{\partial \Phi_A^*} = 0 \quad [36]$$

is the classical master equation. Solutions of the classical master equation with suitable boundary conditions turn out to be generating functionals for the gauge structure of the theory. S is also the starting point for the quantization. One denotes by Σ the subspace of stationary points of the action in the space of fields and antifields:

$$\Sigma = \left\{ z^a \left| \frac{\partial S}{\partial z^a} = 0 \right. \right\} \quad [37]$$

Given a classical solution ϕ_0 of S_0 one stationary point is

$$\phi^i = \phi_0^i, \quad C^\alpha = 0, \quad \Phi_A^* = 0 \quad [38]$$

An action which satisfies the classical master equation has its own set of invariances:

$$\frac{\partial S}{\partial z^a} R_b^a = 0 \quad [39]$$

with

$$R_b^a = \omega^{ac} \frac{\partial_l \partial_r S}{\partial z^c \partial z^b} \quad [40]$$

This equation implies

$$R_c^a R_b^a \Big|_{\Sigma} = 0 \quad [41]$$

One says that R_b^a is invariant on-shell. A nilpotent $2N \times 2N$ matrix has rank $\leq N$. Let r be the rank of the hessian of S at the stationary point:

$$r = \text{rank} \frac{\partial_l \partial_r S}{\partial z^a \partial z^b} \Big|_{\Sigma} \quad [42]$$

We then have $r \leq N$. The relevant solutions of the classical master equation are those for which $r = N$. In this case the number of independent gauge invariances of the type in eqn [39] equals the number of antifields. When at a later stage the gauge is fixed, the nonphysical antifields are eliminated.

To ensure the correct classical limit, the proper solution must contain the classical action S_0 in the sense that

$$S[\Phi^A, \Phi_A^*] \Big|_{\Phi_A^*=0} = S_0[\phi^i] \quad [43]$$

The action $S[\Phi^A, \Phi_A^*]$ can be expanded in a series in the antifields, while maintaining vanishing ghost number and even Grassmann parity:

$$S[\Phi, \Phi^*] = S_0 + \phi_i^* R_\alpha^i C^\alpha + C_a^* \frac{1}{2} T_{\beta\gamma}^\alpha (-1)^{\epsilon_\beta} C^\gamma C^\beta + \phi_i^* \phi_j^* (-1)^{\epsilon_i} \frac{1}{4} E_{\alpha\beta}^{ij} (-1)^{\epsilon_\alpha} C^\beta C^\alpha + \dots \quad [44]$$

When this is inserted into the classical master equation, one finds that this equation implies the gauge structure of the classical theory.

Gauge Fixing and Quantization

Equation [39] shows that the action S still possesses gauge invariances, and hence is not yet suitable for quantization via the path integral approach: a gauge-fixing procedure is necessary. In the Batalin–Vilkovisky approach the gauge is fixed, and the antifields eliminated, by use of a gauge-fixing fermion Ψ which has Grassmann parity $\epsilon(\Psi) = 1$ and $\text{gh}[\Psi] = -1$. It is a functional of the fields Φ^A only; its relation to the antifields is

$$\Phi_A^* = \frac{\partial \Psi}{\partial \Phi^A} \quad [45]$$

We define a surface in functional space

$$\Sigma_\Psi = \left\{ (\Psi^A, \Psi_A^*) \mid \Psi_A^* = \frac{\partial \Psi}{\partial \Phi^A} \right\} \quad [46]$$

so that for any functional $X[\Phi, \Phi^*]$

$$X|_{\Sigma_\Psi} = X \left[\Psi, \frac{\partial \Psi}{\partial \Phi} \right] \quad [47]$$

To construct a gauge-fixing fermion Ψ of ghost number -1 , one must again introduce additional fields. The simplest choice utilizes a trivial pair $\bar{C}_\alpha, \bar{\pi}_\alpha$ with

$$\begin{aligned} \epsilon(\bar{C}_\alpha) &= \epsilon_\alpha + 1, & \epsilon(\bar{\pi}_\alpha) &= \epsilon_\alpha \\ \text{gh}[\bar{C}_\alpha] &= -1, & \text{gh}[\bar{\pi}_\alpha] &= 0 \end{aligned} \quad [48]$$

The fields \bar{C}_α are the Faddeev–Popov antighosts. Along with these fields we include the corresponding antifields $\bar{C}^{*\alpha}, \bar{\pi}^{*\alpha}$. Adding the term $\bar{\pi}_\alpha \bar{C}^{*\alpha}$ to the action S does not spoil its properties as a proper solution to the classical master equation, and one gets the nonminimal action

$$S^{\text{non}} = S + \bar{\pi}_\alpha \bar{C}^{*\alpha} \quad [49]$$

The simplest possibility for Ψ is

$$\Psi = \bar{C}_\alpha \chi^\alpha(\phi) \quad [50]$$

where χ^α are the gauge-fixing conditions for the fields ψ . The gauge-fixed action is denoted by

$$S_\Psi = S^{\text{non}}|_{\Sigma_\Psi} \quad [51]$$

Quantization is performed using the path integral to calculate a correlation function X , with the constraint [45] implemented by a δ -function:

$$\begin{aligned} I_\Psi(X) &= \int D\Phi D\Phi^* \delta \left(\Phi_A^* - \frac{\partial \Psi}{\partial \Phi^A} \right) \\ &\quad \times \exp \left(\frac{i}{\hbar} W[\Phi, \Phi^*] \right) X[\Phi, \Phi^*] \end{aligned} \quad [52]$$

Here W is the quantum action, which reduces to S in the limit $\hbar \rightarrow 0$. An admissible Ψ leads to well-defined propagators when the path integral is expressed as a perturbation series expansion.

The results of a calculation should be independent of the gauge fixing. Consider the integrand in eqn [52],

$$I[\Phi, \Phi^*] = \exp \left(\frac{i}{\hbar} W[\Phi, \Phi^*] \right) X[\Phi, \Phi^*] \quad [53]$$

Under an infinitesimal change in Ψ

$$I_{\Psi+\delta\Psi}(X) - I_\Psi(X) \approx \int D\Phi \Delta I \delta\Psi \quad [54]$$

where the Laplacian Δ is

$$\Delta = (-1)^{\epsilon_A+1} \frac{\partial}{\partial \Phi^A} \frac{\partial}{\partial \Phi_A^*} \quad [55]$$

Obviously, the integral $I_\Psi(X)$ is independent of Ψ if $\Delta I = 0$. For $X=1$ one gets the requirement

$$\Delta \exp\left(\frac{i}{\hbar} W\right) = \exp\left(\frac{i}{\hbar} W\right) \times \left(\frac{i}{\hbar} \Delta W - \frac{1}{2\hbar^2} (W, W)\right) = 0 \quad [56]$$

The formula

$$\frac{1}{2} (W, W) = i\hbar \Delta W \quad [57]$$

is the quantum master equation. A gauge-invariant correlation function satisfies

$$(X, W) = i\hbar \Delta X \quad [58]$$

The terms of higher order in \hbar by which the quantum action W may differ from the solution of the classical master equation S correspond to the counter-terms of the renormalizable gauge theory if

$$\Delta S = 0 \quad [59]$$

One must, of course, use a regularization scheme which respects the symmetries of the theory. For $W = S + O(\hbar)$ the quantum master equation [57] reduces in this case to the classical master equation

$$(S, S) = 0 \quad [60]$$

Hence, up to possible counter-terms, one may simply choose $W = S$.

To implement the gauge fixing, one uses for the action $W = S^{\text{non}}$. For the path integral $Z = I_\Psi(X=1)$, the integration over the antifields in eqn [52] is performed by using the δ -function. The result is

$$Z = \int D\Phi \exp\left(\frac{i}{\hbar} S_\Psi\right) \quad [61]$$

Geometrical Interpretation of Topological Field Theories

The Batalin–Vilkovisky formalism for topological field theories has been given a geometrical interpretation by AKSZ (1997).

A supermanifold equipped with an odd vector field satisfying $Q^2=0$ is called a Q -manifold. A Q -manifold provided with an odd symplectic structure ω (P-structure) is called a QP-manifold if the odd symplectic structure is Q -invariant, that is, $L_Q\omega=0$. Every solution to the classical master equation determines a QP-structure on M and vice

versa. The geometric object corresponding to a classical mechanical system in the Batalin–Vilkovisky formalism is a QP-manifold.

The nondegenerate closed 2-form ω is written as

$$\omega = dz^a \zeta_{ab} dz^b \quad [62]$$

where z^a are local coordinates in the supermanifold M . For functions on M , an (odd) Poisson bracket is defined as in eqn [33], where ω^{ab} stands for the inverse matrix of ω_{ab} . An even function S on M satisfies the classical master equation if $(S, S) = 0$. The correspondence between vector fields and functions on M is given by $K_F G = (G, F)$, where K_F is the vector field, F the given function, and G an arbitrary function. The function F is called the Hamiltonian of the vector field K_F .

Geometrically, equivalent QP-manifolds describe the same physics. In particular, one can consider an even Hamiltonian vector field K_F corresponding to an odd function F . This vector field determines an infinitesimal transformation preserving P-structure. It transforms a solution S to the classical master equation into the physically equivalent solution $S + \epsilon(S, F)$, where ϵ is an infinitesimally small parameter.

A submanifold L of a P-manifold M is called a Lagrangian submanifold if the restriction of the form ω to L vanishes. In the particular case when $M = \Pi T^*N$ (the cotangent bundle to N with reversed parity of fibres) with standard P-structure, one can construct many examples of Lagrangian submanifolds in the following way. Fix an odd function Ψ on N , the gauge fermion. The submanifold $L_\Psi \in M$ determined by the equation

$$\xi^a = \frac{\partial \Psi}{\partial x^a} \quad [63]$$

where $\{x^a, \xi_a\}$ are coordinates corresponding to the identification of M , will be a Lagrangian submanifold of M .

The P-manifold M in the neighborhood of L can be identified with ΠT^*L . In other words, one can find such a neighborhood U of L in M and a neighborhood V of L in ΠT^*L that there exists an isomorphism of P-manifolds U and V leaving L intact. Using this isomorphism a function Ψ defined on a Lagrangian submanifold $L \subset M$ determines another Lagrangian submanifold $L_\Psi \subset M$.

Consider a solution S to the classical master equation on M . In the Batalin–Vilkovisky formalism we have to restrict S to a Lagrangian submanifold $L \in M$, then the quantization of S can be performed by integration of $\exp(iS/\hbar)$ over L . One may construct an odd vector field Q on L in such a

way that the functional S restricted to L is \mathcal{Q} -invariant. This invariance is BRST invariance.

AKSZ apply these geometric constructions to obtain in a natural way the action functionals of two-dimensional sigma-models (Witten 1998) and to show that the Chern–Simons theory (Axelrod and Singer 1991) in Batalin–Vilkovisky formalism arises as a sigma-model with target space $\Pi\mathcal{G}$, where \mathcal{G} stands for a Lie algebra and Π denotes parity inversion.

The Poisson-Sigma Model

The quantization of the Poisson-sigma model was performed by Hirshfeld and Schwarzweiler (2000) and by Cattaneo and Felder (2001). The Poisson-sigma model is the simplest topological field theory in two dimensions. It is a field theory on a two-dimensional world sheet without boundary (Schaller and Strobl 1994). It involves a set of bosonic scalar fields, which can be seen as a set of maps $X^i: M \rightarrow N$, where N is a Poisson manifold. In addition, one has a 1-form A on the world sheet M which takes values in $T^*(N)$, for x coordinates on M we have $A = A_{\mu i} dx^i \wedge dx^\mu$. Its action is

$$S_0[X, A] = \int_M \mu (\epsilon^{\mu\nu} (A_{\mu i} \partial_\nu X^i + P^{ij}(X) A_{\mu i} A_{\nu j})) \quad [64]$$

where $\epsilon^{\mu\nu}$ is the antisymmetric tensor and μ is the volume form on M . The gauge transformations of the model are

$$\delta X^i = P^{ij}(X) \varepsilon_j, \quad \delta A_{\mu i} = D_{\mu i}^j \varepsilon_j \quad [65]$$

where $D_{\mu i}^j = \partial_\mu \delta_i^j + P^{kj}{}_{,i} A_{\mu k}$. The equations of motion are

$$\epsilon^{\mu\nu} D_{\mu i}^j A_{\nu j} = 0 \quad [66]$$

and

$$\epsilon^{\mu\nu} (\partial_\nu X^i + P^{ij} A_{\nu j}) = \epsilon^{\mu\nu} D_\nu X^i = 0 \quad [67]$$

The gauge algebra is given by

$$\begin{aligned} [\delta(\varepsilon_1), \delta(\varepsilon_2)] X^i &= P^{ij} (P^{mn}{}_{,j} \varepsilon_{1n} \varepsilon_{2m}) \\ [\delta(\varepsilon_1), \delta(\varepsilon_2)] A_{\mu i} &= D_{\mu i}^j (P^{mn}{}_{,j} \varepsilon_{1n} \varepsilon_{2m}) \\ &\quad - (\epsilon^{\nu\rho} D_\rho X^j) \epsilon_{\nu\mu} P^{mn}{}_{,ji} \varepsilon_{1n} \varepsilon_{2m} \end{aligned} \quad [68]$$

In our general notation the generators of the gauge transformations R are here P^{ij} and $D_{\mu i}^j$. The gauge tensors T and E are $P^{ij}{}_{,k}$ and $\epsilon_{\nu\mu} P^{mn}{}_{,ji}$. The higher-order gauge tensors A and B vanish.

The ghost fields are again denoted by C^i . The Noether identities are then

$$\int_M \mu \left(\epsilon^{\mu\nu} D_{\mu i}^j A_{\nu j} P^{ki} + (\epsilon^{\mu\nu} D_\nu X^i) D_{\mu i}^k \right) C_k = 0 \quad [69]$$

Considering the commutator of two gauge transformations leads to (see eqns [8]–[11])

$$\begin{aligned} \int_M \mu (2P^{mi}{}_{,j} P^{nj} - P^{ij} P^{mn}{}_{,j}) C_m C_n &= 0 \\ \int_M \mu \left(2(P^{jk}{}_{,i} D_{\mu j}^l + P^{mk}{}_{,ij} A_{\mu m} P^{jl}) \right. \\ &\quad \left. - D_{\mu i}^m P^{kl}{}_{,m} + (\epsilon^{\rho\nu} D_\rho X^j) \epsilon_{\nu\mu} P^{kl}{}_{,ji} \right) C_l C_k = 0 \end{aligned} \quad [70]$$

The Jacobi identity is

$$P^{ij}{}_{,m} P^{mk} C_i C_j C_k = 0 \quad [71]$$

The fields and antifields of the model are

$$\Phi^A = \{A^{\mu i}, X^i, C_i\} \quad \text{and} \quad \Phi_A^* = \{A^{\mu i*}, X_i^*, C_i^*\} \quad [72]$$

The extended action is

$$\begin{aligned} \mathcal{S} = \int_M \mu \left(\epsilon^{\mu\nu} (A_{\mu i} \partial_\nu X^i + P^{ij}(X) A_{\mu i} A_{\nu j}) \right. \\ + A^{\mu i*} D_{\mu i}^j C_j + X_i^* P^{ij}(X) C_j + \frac{1}{2} C_i^* P^{jk}{}_{,i}(X) C_j C_k \\ \left. + \frac{1}{4} A^{\mu i*} A^{\nu j*} \epsilon_{\mu\nu} P^{kl}{}_{,ij}(X) C_k C_l \right) \end{aligned} \quad [73]$$

The gauge-fixing conditions are taken to be of the form $\chi_i(A, X)$, so that the gauge fermion [50] becomes $\Psi = \bar{C}^i \chi_i(A, X)$. The antifields are then fixed to be

$$\begin{aligned} A_{\mu i}^* &= \bar{C}_j \frac{\partial \chi_j(A, X)}{\partial A_{\mu i}} \\ X_i^* &= \bar{C}_j \frac{\partial \chi_j(A, X)}{\partial X^i} \\ C_i^* &= 0 \\ \bar{C}_i^* &= \chi_i(A, X) \end{aligned} \quad [74]$$

The gauge-fixed action is

$$\begin{aligned} \mathcal{S}_\Psi = \int_M \mu \left(\epsilon^{\mu\nu} (A_{\mu i} \partial_\nu X^i + P^{ij}(X) A_{\mu i} A_{\nu j}) \right. \\ + \bar{C}^k \frac{\partial \chi_k(A, X)}{\partial A_{\mu i}} D_{\mu i}^j C_j + \bar{C}^k \frac{\partial \chi_k(A, X)}{\partial X^i} P^{ij} C_j \\ + \frac{1}{4} \bar{C}^m \frac{\partial \chi_m(A, X)}{\partial A_{\mu i}} \bar{C}^n \frac{\partial \chi_n(A, X)}{\partial A_{\nu j}} \epsilon_{\mu\nu} P^{kl}{}_{,ij}(X) \\ \left. \times C_k C_l + \bar{\pi}^i \chi_i(A, X) \right) \end{aligned} \quad [75]$$

Now consider different gauge conditions:

1. First, the Landau gauge for the gauge potential $\chi_i = \partial^\mu A_{\mu i}$, so that the gauge fermion becomes $\Psi = \bar{C}^i \partial^\mu A_{\mu i}$. The antifields are fixed to be

$$\begin{aligned} A^{\mu i*} &= \partial^\mu \bar{C}^i \\ X_i^* &= C^i = 0 \\ \bar{C}_i^* &= \partial^\mu A_{\mu i} \end{aligned} \quad [76]$$

for this gauge choice the gauge-fixed action is

$$\begin{aligned}
 S_{\Psi} = & \int_M \mu \left(\epsilon^{\mu\nu} (A_{\mu i} \partial_{\nu} X^i + P^{ij}(X) A_{\mu i} A_{\nu j}) + \bar{C}^i \partial^{\mu} D_{\mu}^j C_j \right. \\
 & + \frac{1}{4} (\partial^{\mu} \bar{C}^i) (\partial^{\nu} \bar{C}^j) \epsilon_{\mu\nu} P^{kl}{}_{,ij}(X) \\
 & \left. \times C_k C_l - \bar{\pi}^i (\partial^{\mu} A_{\mu i}) \right) \quad [77]
 \end{aligned}$$

Translating this action into the notation of Cattaneo and Felder, one sees that it is exactly the expression they use to derive the perturbation series.

2. Now consider the temporal gauge $\chi_i = A_{0i}$. The gauge fermion is given by $\Psi = \bar{C}^i A_{0i}$. The anti-fields are fixed to

$$\begin{aligned}
 A^{*0i} &= \bar{C}^i \\
 A^{*1i} &= 0 \\
 X_i^* &= C^{*i} = 0 \\
 \bar{C}_i^* &= A_{0i}
 \end{aligned} \quad [78]$$

The gauge-fixed action is

$$\begin{aligned}
 S_{\Psi} = & \int_M \mu \left(\epsilon^{\mu\nu} (A_{\mu i} \partial_{\nu} X^i + P^{ij}(X) A_{\mu i} A_{\nu j}) \right. \\
 & \left. + \bar{C}^i D_{0i}^j C_j - \bar{\pi}^i (A_{0i}) \right) \quad [79]
 \end{aligned}$$

3. Finally consider the Schwinger–Fock gauge $\chi_i = x^{\mu} A_{\mu i}$. Then the antifields are fixed to be

$$\begin{aligned}
 A^{*\mu i} &= x^{\mu} \bar{C}^i \\
 X_i^* &= C^{*i} = 0 \\
 \bar{C}_i^* &= x^{\mu} A_{\mu i}
 \end{aligned} \quad [80]$$

for this gauge choice the gauge-fixed action is

$$\begin{aligned}
 S_{\Psi} = & \int_M \mu \left(\epsilon^{\mu\nu} (A_{\mu i} \partial_{\nu} X^i + P^{ij}(X) A_{\mu i} A_{\nu j}) \right. \\
 & \left. + \bar{C}^i x^{\mu} D_{\mu}^j C_j - \bar{\pi}^i (\partial^{\mu} A_{\mu i}) \right) \quad [81]
 \end{aligned}$$

Notice that in the noncovariant gauges 2 and 3 the action simplifies, in that the term which arose because of the nonclosed nature of the gauge algebra vanishes.

See also: BF Theories; BRST Quantization; Constrained Systems; Graded Poisson Algebras; Operads; Perturbative Renormalization Theory and BRST; Supermanifolds; Topological Sigma Models.

Further Reading

- Alexandrov M, Kontsevich M, Schwarz A, and Zaboronsky O (1997) Geometry of the Master Equation. *International Journal of Modern Physics A*12: 1405–1430.
- Axelrod S and Singer IM (1991) *Chern–Simons Perturbation Theory*, Proceedings of the XXth Conference on Differential Geometric Methods in Physics, Baruch College/CUNY, NY. (hep-th/9110056).
- Batalin IA and Vilkovisky GA (1977) Gauge algebra and quantization. *Physics Letters* 69B: 309–312.
- Becchi C, Rouet A, and Stora R (1976) Renormalization of gauge theories. *Annals of Physics (NY)* 98: 287–321.
- Cattaneo AS and Felder G (2001) On the AKSZ formulation of the Poisson–Sigma model. *Letters of Mathematical Physics* 56: 163–179.
- Faddeev LD and Popov VN (1967) Feynman diagrams for the Yang–Mills field. *Physics Letters* 25B: 29–30.
- Gomis J, Paris J, and Samuel S (1994) Antibracket Antifields and gauge-theory quantization. *Physics Reports* 269: 1–145.
- Hirshfeld AC and Schwarzweller T (2000) Path integral quantization of the Poisson–Sigma model. *Annals of Physics (Leipzig)* 9: 83–101.
- Schaller P and Strobl T (1994) Poisson structure induced (topological) field theories. *Modern Physics Letters A*9: 3129–3136.
- Witten E (1988) Topological sigma models. *Communications in Mathematical Physics* 118: 411–449.
- Zinn-Justin J (1975) Renormalization of gauge theories. In: Rollnik H and Dietz K (eds.) *Trends in Elementary Particle Physics*, Lecture Notes in Physics, vol. 37. Berlin: Springer.

Bethe Ansatz

M T Batchelor, Australian National University, Canberra, ACT, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Bethe ansatz is a particular form of wave function introduced in the diagonalization of the Heisenberg spin chain. It underpins the majority of exactly solved models in statistical mechanics and quantum field

theory. At the heart of the Bethe ansatz is the way in which multibody interactions factor into two-body interactions. The Bethe ansatz is thus intimately entwined with the theory of integrability.

The way in which the Bethe ansatz works is best understood by working through an explicit hands-on example. The canonical example is the isotropic antiferromagnetic Heisenberg Hamiltonian

$$H = \sum_{i=1}^{L-1} h_{i,i+1} + h_{L,1}, \quad h_{ij} = \frac{1}{2} (\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j + 1) \quad [1]$$

where $\sigma = (\sigma^x, \sigma^y, \sigma^z)$ are Pauli matrices and L is the length of the chain. Periodic boundary conditions are imposed. However, open boundary conditions may also be treated, along with the addition of magnetic bulk and boundary fields. The z -components of each of the spins are either up or down. Since the z -component of the total spin commutes with the Hamiltonian, the total number n of up spins serves as a good quantum number. A state of the system can therefore be conveniently described in terms of the coordinates of all the up spins. Denote these coordinates by x_i , with $1 \leq x_i \leq L$. The quantum number n ensures that the Hamiltonian decomposes into $L + 1$ sectors, each of size L choose n . The antiferromagnetic ground state occurs in the largest sector.

The normalization of the Hamiltonian [1] is such that its action is that of the permutation operator:

$$\begin{aligned} b|--\rangle &= |--\rangle, & b|++\rangle &= |++\rangle \\ b|+-\rangle &= |-+\rangle, & b| -+\rangle &= |+-\rangle \end{aligned} \quad [2]$$

Diagonalization of Sectors

One can address the diagonalization of the sectors for various cases.

Case 1: $n=0$

Consider the case with all spins down. The eigenstate is $\Psi = |-\dots-\rangle$, with $H\Psi = L\Psi$ and, thus, $E=L$ is the trivial solution.

Case 2: $n=1$

There are L states, with

$$\Psi = \sum_{x=1}^L a(x) |\psi(x)\rangle \quad [3]$$

where $|\psi(x)\rangle$ is the state with an up spin at site x . The aim is to find the amplitudes $a(x)$. It is clear that

$$\begin{aligned} H|\psi(x)\rangle &= (L-2)|\psi(x)\rangle + |\psi(x-1)\rangle \\ &+ |\psi(x+1)\rangle \end{aligned} \quad [4]$$

in the bulk (away from either boundary). Insertion of [3] into $H\Psi = E\Psi$ gives

$$Ea(x) = (L-2)a(x) + a(x-1) + a(x+1) \quad [5]$$

Substitution of spin waves $a(x) = e^{ikx}$ gives

$$E = L - 2 + 2 \cos k \quad [6]$$

The boundary conditions are such that $a(0) = a(L)$ and $a(L+1) = a(1)$; either gives $e^{ikL} = 1$, from which the L values of k follow.

Case 3: $n=2$

Here the wave function can be written in terms of the two flipped spins as

$$\Psi = \sum_{x < y} a(x, y) |\psi(x, y)\rangle \quad [7]$$

It is to be emphasized that one is working in the region with $x < y$. There are two cases to consider: (1) $y > x + 1$ and (2) $y = x + 1$. Consider the interactions in the bulk. For (1) the action of the Hamiltonian implies

$$\begin{aligned} Ea(x, y) &= (L-4)a(x, y) + a(x-1, y) + a(x+1, y) \\ &+ a(x, y-1) + a(x, y+1) \end{aligned} \quad [8]$$

and for (2)

$$\begin{aligned} Ea(x, x+1) &= (L-2)a(x, x+1) \\ &+ a(x-1, x+1) + a(x, x+2) \end{aligned} \quad [9]$$

The compatibility of these two equations requires that

$$2a(x, x+1) = a(x, x) + a(x+1, x+1) \quad [10]$$

which is known as the ‘‘collision’’ or ‘‘meeting’’ condition.

Some adjustments need to be made for spins which get flipped at the boundaries. Looking at [8] and [9] with $x=1$ and $x=L$, it is evident that one can take

$$a(y, x+L) = a(x, y) \quad [11]$$

to restore the original ordering. The terms which arise involve up spins at sites 0 and $L+1$. This illustrates the periodic boundary condition.

We now assume (the Bethe ansatz) that

$$a(x, y) = A_{12} e^{ik_1 x} e^{ik_2 y} + A_{21} e^{ik_2 x} e^{ik_1 y} \quad [12]$$

Substitution of the ansatz [12] into [8] gives

$$E = L - 4 + 2 \cos k_1 + 2 \cos k_2 \quad [13]$$

Substitution of [12] into [10] gives

$$\frac{A_{12}}{A_{21}} = - \frac{1 - 2 e^{ik_1} + e^{i(k_1+k_2)}}{1 - 2 e^{ik_2} + e^{i(k_1+k_2)}} \quad [14]$$

The three relations [11], [12], and [14] give the Bethe equations

$$e^{ik_1 L} = \frac{A_{12}}{A_{21}} \quad \text{and} \quad e^{ik_2 L} = \frac{A_{21}}{A_{12}} \quad [15]$$

which are to be solved for k_1 and k_2 . Note that $e^{i(k_1+k_2)L} = 1$.

Case 4: $n=3$

The full power of the Bethe ansatz method becomes evident for three particles. Here

$$\Psi = \sum_{x < y < z} a(x, y, z) |\psi(x, y, z)\rangle \quad [16]$$

There are several cases to consider:

1. $y > x + 1$ and $z > y + 1$, where

$$Ea(x, y, z) = (L - 6)a(x, y, z) + a(x \pm 1, y, z) + a(x, y \pm 1, z) + a(x, y, z \pm 1) \quad [17]$$

By $a(x \pm 1, y, z)$, we mean $a(x + 1, y, z) + a(x - 1, y, z)$, etc.

2. $y = x + 1$ and $z > y + 1$, with

$$\begin{aligned} Ea(x, x + 1, z) &= (L - 4)a(x, x + 1, z) + a(x - 1, x + 1, z) \\ &\quad + a(x, x + 2, z) + a(x, x + 1, z \pm 1) \end{aligned} \quad [18]$$

3. $y > x + 1$ and $z = y + 1$, where

$$\begin{aligned} Ea(x, y, y + 1) &= (L - 4)a(x, y, y + 1) + a(x \pm 1, y, y + 1) \\ &\quad + a(x, y - 1, y + 1) + a(x, y, y + 2) \end{aligned} \quad [19]$$

4. $y = x + 1$ and $z = y + 1$, for which

$$\begin{aligned} Ea(x, x + 1, x + 2) &= (L - 2)a(x - 1, x + 1, x + 2) \\ &\quad + a(x, x + 1, x + 3) \end{aligned} \quad [20]$$

Again, we must ensure that these equations are compatible. This involves comparison of the last three equations with [17]. The three equations to be satisfied are

$$2a(x, x + 1, z) = a(x, x, z) + a(x + 1, x + 1, z) \quad [21]$$

$$2a(x, y, y + 1) = a(x, y, y) + a(x, y + 1, y + 1) \quad [22]$$

$$\begin{aligned} 4a(x, x + 1, x + 2) &= a(x, x, x + 2) + a(x, x + 1, x + 1) \\ &\quad + a(x, x + 2, x + 2) \\ &\quad + a(x + 1, x + 1, x + 2) \end{aligned} \quad [23]$$

But note that setting $z = x + 2$ in [21] and $y = x + 1$ in [22] leads to [23] being automatically satisfied. We are thus left with only two equations [21] and [22]. Note the similarity between these two equations and the meeting condition [10] for the $n=2$ case.

In this case the Bethe ansatz is

$$\begin{aligned} a(x, y, z) &= A_{123}z_1^x z_2^y z_3^z + A_{132}z_1^x z_3^y z_2^z \\ &\quad + A_{213}z_2^x z_1^y z_3^z + A_{231}z_2^x z_3^y z_1^z \\ &\quad + A_{321}z_3^x z_2^y z_1^z + A_{312}z_3^x z_1^y z_2^z \end{aligned} \quad [24]$$

in which $z_j = e^{ik_j}$. This is a sum over the 3! permutations of the integers 1, 2, 3. Inserting this ansatz into [17] gives

$$E = L - 6 + 2(\cos k_1 + \cos k_2 + \cos k_3) \quad [25]$$

To determine the k_j , it is convenient to define

$$s_{ij} = 1 - 2z_j + z_i z_j \quad [26]$$

Substitution of [24] into the meeting conditions [21] and [22] then gives

$$\begin{aligned} s_{12}A_{123} + s_{21}A_{213} + s_{13}A_{132} + s_{31}A_{312} \\ + s_{23}A_{231} + s_{32}A_{321} = 0 \end{aligned} \quad [27]$$

$$\begin{aligned} s_{23}A_{123} + s_{32}A_{132} + s_{13}A_{213} + s_{31}A_{231} \\ + s_{21}A_{321} + s_{12}A_{312} = 0 \end{aligned} \quad [28]$$

These equations are assumed to be satisfied in permutation pairs, that is,

$$\begin{aligned} s_{12}A_{123} + s_{21}A_{213} = 0 \\ s_{23}A_{123} + s_{32}A_{132} = 0, \text{ etc.} \end{aligned} \quad [29]$$

Up to an overall constant, the relations [27] and [28] are satisfied by

$$\begin{aligned} A_{123} = s_{21}s_{31}s_{32}, \quad A_{132} = -s_{31}s_{21}s_{23} \\ A_{312} = s_{13}s_{23}s_{21}, \quad A_{321} = -s_{23}s_{13}s_{12} \\ A_{231} = s_{32}s_{12}s_{13}, \quad A_{213} = -s_{12}s_{32}s_{31} \end{aligned} \quad [30]$$

The boundary condition, $a(y, z, x + L) = a(x, y, z)$, gives

$$\begin{aligned} (z_1^L A_{321} - A_{132})z_1^x z_3^y z_2^z + (z_2^L A_{312} - A_{231})z_2^x z_3^y z_1^z \\ + (z_1^L A_{231} - A_{123})z_1^x z_2^y z_3^z + (z_3^L A_{213} - A_{321})z_3^x z_2^y z_1^z \\ + (z_2^L A_{132} - A_{213})z_2^x z_1^y z_3^z + (z_3^L A_{123} - A_{312})z_3^x z_1^y z_2^z \\ = 0 \end{aligned} \quad [31]$$

This leads to the equations

$$\begin{aligned} z_1^L = \frac{A_{123}}{A_{231}} = \frac{A_{132}}{A_{321}} = \frac{s_{21}s_{31}}{s_{12}s_{13}} \\ z_2^L = \frac{A_{213}}{A_{132}} = \frac{A_{231}}{A_{312}} = \frac{s_{12}s_{32}}{s_{21}s_{23}} \\ z_3^L = \frac{A_{321}}{A_{213}} = \frac{A_{312}}{A_{123}} = \frac{s_{13}s_{23}}{s_{31}s_{32}} \end{aligned} \quad [32]$$

which can be solved for the Bethe roots k_j .

General n

The general Bethe ansatz is

$$a(x_1, \dots, x_n) = \sum_P A_{p_1, \dots, p_n} z_{p_1}^{x_1} \dots z_{p_n}^{x_n} \quad [33]$$

where the sum is over all $n!$ permutations $P = \{p_1, \dots, p_n\}$ of the integers $1, \dots, n$. The boundary condition is

$$a(x_2, x_3, \dots, x_n, x_1 + L) = a(x_1, x_2, \dots, x_n) \quad [34]$$

leading to the Bethe equations

$$z_{p_1}^L = \frac{A_{p_1, \dots, p_n}}{A_{p_2, \dots, p_n, p_1}} \quad [35]$$

for all permutations, with

$$A_{p_1, \dots, p_n} = \epsilon_P \prod_{1 \leq i < j \leq n} s_{p_i, p_j} \quad [36]$$

where ϵ_P is the signature of the permutation. Finally,

$$z_{p_1}^L = (-)^{n-1} \prod_{\ell=2}^n \frac{s_{p_\ell, p_1}}{s_{p_1, p_\ell}} \quad \text{or} \quad z_j^L = (-)^{n-1} \prod_{\substack{\ell=1 \\ \ell \neq j}}^n \frac{s_{\ell, j}}{s_{j, \ell}} \quad [37]$$

for $j = 1, \dots, n$. The eigenvalues are given by

$$E = L + \sum_{j=1}^n (2 \cos k_j - 2) \quad [38]$$

Another form of the Bethe equations is obtained by defining

$$e^{ik_j} = \frac{u_j - (1/2)i}{u_j + (1/2)i} \quad [39]$$

which gives

$$E = L - \sum_{j=1}^n \frac{1}{u_j^2 + 1/4} \quad [40]$$

with u_j satisfying

$$\left(\frac{u_j - (1/2)i}{u_j + (1/2)i} \right)^L = - \prod_{\ell=1}^n \frac{u_j - u_\ell - i}{u_j - u_\ell + i} \quad [41]$$

for $j = 1, \dots, n$.

All eigenvalues of the Heisenberg spin chain may be obtained in terms of the Bethe ansatz solution. For example, the distribution of roots u_j for the ground state are real and symmetric about the origin. Excitations may involve complex roots. Although obtained exactly in terms of the Bethe roots, the Bethe ansatz wave function is cumbersome.

We have thus seen how the Bethe ansatz works for the Heisenberg spin chain. The underlying mechanism is the way in which the collision or

meeting conditions can be handled in terms of two-body interactions. To see this more clearly, the six permutation pair equations [29] can be written in the general form $A_{abc} = Y_{ab} A_{bac}$ and $A_{abc} = Y_{bc} A_{acb}$, where $Y_{ab} = -s_{ba}/s_{ab}$. Now there are two possible paths to get from A_{abc} to A_{cba} , namely

$$\begin{aligned} A_{cba} &= Y_{ab} Y_{ac} Y_{bc} A_{abc} \\ A_{cba} &= Y_{bc} Y_{ac} Y_{ab} A_{abc} \end{aligned} \quad [42]$$

Both paths must be equivalent, with

$$Y_{ab} Y_{ba} = 1 \quad \text{and} \quad Y_{ab} Y_{ac} Y_{bc} = Y_{bc} Y_{ac} Y_{ab} \quad [43]$$

The latter is a condition of nondiffraction or equivalently a manifestation of the Yang–Baxter equation.

Historically, the next model to be exactly solved in terms of the Bethe ansatz was the one-dimensional model of N interacting bosons on a line of length L defined by the Hamiltonian

$$H = - \sum_{i=1}^N \frac{\partial^2}{\partial x_i^2} + 2c \sum_{1 \leq i < j \leq N} \delta(x_i - x_j) \quad [44]$$

where c is a measure of the interaction strength. For this model the Bethe ansatz wave function is of the same form as [33] with the two-body interaction term given by

$$s_{ab} = k_a - k_b + ic \quad [45]$$

The Bethe equations are given by

$$\begin{aligned} \exp(ik_j L) &= - \prod_{\ell=1}^N \frac{k_j - k_\ell + ic}{k_j - k_\ell - ic} \\ \text{for } j &= 1, \dots, N \end{aligned} \quad [46]$$

The energy eigenvalue is

$$E = \sum_{j=1}^N k_j^2 \quad [47]$$

For repulsive ($c > 0$) interactions, one can prove that all Bethe roots are real.

The Bethe ansatz has been applied to a number of other and more general models, both for discrete spins and in the continuum. These include the anisotropic Heisenberg (XXZ) spin chain, for which the above working readily generalizes to trigonometric functions. The underlying ansatz [33] remains the same. One key generalization is the nested Bethe ansatz, which arises, for example, in the solution of the general N -state permutator model, the Hubbard model, and the Gaudin–Yang model of interacting fermions. For such models the nested Bethe ansatz involves an additional level of work to determine the amplitudes appearing in the

wave function [33] due to higher symmetries. This results in Bethe equations involving different types or colors of roots.

The exactly solved one-dimensional quantum spin chains may also be obtained from their two-dimensional classical counterparts – the vertex models. For example, the six-vertex model shares the same Bethe ansatz wave function and Bethe equations as the XXZ spin chain. The more general permutator Hamiltonians are related to multistate vertex models. One may also consider other spin- S models.

The discussion in this article has centered on what is known as the coordinate Bethe ansatz. Another formulation is the algebraic Bethe ansatz, which was developed for the systematic treatment of the higher-spin models. In this formulation, operators create the Bethe states by acting on a vacuum. The algebraic Bethe ansatz goes hand-in-hand with the quantum inverse-scattering method. In all of the exactly solved Bethe ansatz models, it is possible to derive quantities like the ground-state energy per site via the root density method, which assumes that the Bethe roots form a uniform distribution in the infinite-size limit. The thermodynamics of the Bethe ansatz solvable models may also be calculated in a systematic fashion.

Despite Bethe's early optimism, the Bethe ansatz has not been extended to higher-dimensional systems.

See also: Affine Quantum Groups; Eight Vertex and Hard Hexagon Models; Integrability and Quantum Field

Theory; Integrable Systems: Overview; Quantum Spin Systems; Yang–Baxter Equations.

Further Reading

- Baxter RJ (1983) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.
- Baxter RJ (2003) Completeness of the Bethe ansatz for the six- and eight-vertex models. *Journal of Statistical Physics* 108: 1–48.
- Bethe HA (1931) Zur Theorie der Metalle I. Eigenwerte und Eigenfunktionen der linearen Atomkette. *Zeitschrift für Physik* 71: 205–226.
- Gaudin M (1967) Un Système à Une Dimension de Fermions en Interaction. *Physics Letters A* 24: 55–56.
- Gaudin M (1983) *la Fonction d'onde de Bethe*. Paris: Masson.
- Korepin VE, Izergin AG, and Bogoliubov NM (1993) *Quantum Inverse Scattering Method and Correlation Functions*. Cambridge: Cambridge University Press.
- Lieb EH and Liniger W (1963) Exact analysis of an interacting Bose gas I. The general solution and the ground state. *Physical Review* 130: 1605–1616.
- Mattis DC (1993) *The Many-Body Problem: An Encyclopaedia of Exactly Solved Models in One-Dimension*. Singapore: World Scientific.
- McGuire JB (1964) Study of exactly soluble one-dimensional N -body problems. *Journal of Mathematical Physics* 5: 622–636.
- Sutherland B (2004) *Beautiful Models: 70 Years of Exactly Solved Quantum Many-Body Problems*. Singapore: World Scientific.
- Takahashi M (1999) *Thermodynamics of One-Dimensional Solvable Models*. Cambridge: Cambridge University Press.
- Yang CN (1967) Some exact results for the many-body problem in one-dimension with repulsive Delta-function interaction. *Physical Review Letters* 19: 1312–1315.

BF Theories

M Blau, Université de Neuchâtel, Neuchâtel, Switzerland

© 2006 Elsevier Ltd. All rights reserved.

Introduction

BF theories are a class of gauge theories with a nontrivial metric-independent classical action. As such these theories are candidate topological field theories akin to the Chern–Simons theory in three dimensions, but in contrast to the Chern–Simons theory these exist and are well defined in arbitrary dimensions.

The name “BF theories” derives from the fact that, roughly (see [1] below and the subsequent discussion for a more precise description), the action of the BF theory takes the form $\int B \wedge F_A$ with F_A the curvature of a connection A and B a Lagrange multiplier. The classical equations of motion imply

that A is flat, $F_A = 0$, and thus BF theories are topological gauge theories of flat connections.

Abelian BF theories and their relation to topological invariants (the Ray–Singer torsion) were originally discussed by Schwarz (1978, 1979). In the context of the topological field theory, non-abelian BF theories were introduced in Horowitz (1989) and Blau and Thompson (1989, 1991).

Since then, BF theories have attracted a lot of attention as simple toy-models of (topological) gauge theories, and also because of their relationships with the Chern–Simons theory, the Yang–Mills theory, and gauge-theory formulations of gravity, as well as because of the rather rich and intricate structure of their quantum theories.

The purpose of this article is to provide an overview of these various features of BF theories. The standard reference for the basic classical and quantum properties of BF theories is Birmingham *et al.* (1991).

Basic Classical Properties of BF Theories

Nonabelian BF Theories

The classical action and equations of motion Typically, the classical action of the BF theory takes the form

$$S_{\text{BF}}(A, B) = \int_M \text{tr}_G B \wedge F_A \quad [1]$$

where F_A is the curvature of a connection A on a principal G -bundle $P \rightarrow M$ over an n -dimensional manifold M , B is an ad-equivariant horizontal $(n-2)$ -form on P , and tr_G (a trace) denotes an ad-invariant nondegenerate scalar product on the Lie algebra \mathfrak{g} of the Lie group G . Generalizations of this are possible, in particular, for G abelian or for $n=3$ and are mentioned below.

We consider F_A and B as forms on M taking values in the bundle of Lie algebras $\text{ad}P = P \times_{\text{ad}} \mathfrak{g}$ and refer to such objects as elements of $\Omega^*(M, \mathfrak{g})$. Then $\text{tr} B \wedge F_A \in \Omega^n(M, \mathbb{R})$ is a volume form on M . In order to simplify the exposition, in the following we will mostly assume that G is compact semisimple and that M is compact without a boundary (even though relaxing any one of these conditions is possible and also of interest in its own right).

Varying the action [1] with respect to A and B , one obtains the classical equations of motion

$$F_A = 0, \quad d_A B = 0 \quad [2]$$

where

$$d_A B = dB + [A, B] \quad [3]$$

is the covariant exterior derivative. In particular, therefore, the equations of motion imply that the connection A is flat.

Gauge invariance For any n , the action [1] is invariant under G gauge transformations (vertical automorphisms of P) acting on A and B as

$$A \rightarrow g^{-1}Ag + g^{-1}dg, \quad B \rightarrow g^{-1}Bg \quad [4]$$

(the latter is what is meant by the fact that B takes values in $\text{ad}P$), because F_A is also ad-equivariant, $F_A \rightarrow g^{-1}F_{Ag}$, and tr_G is ad-invariant. The infinitesimal version of this statement is that the action is invariant under the variations

$$\delta A = d_A \lambda, \quad \delta B = [B, \lambda] \quad [5]$$

where $\lambda \in \Omega^0(M, \mathfrak{g})$ can (formally) be thought of as an element of the Lie algebra of the group of gauge transformations.

Gauge-fixing this symmetry can proceed in the usual way (via the Faddeev–Popov or Becchi–Rouet–

Stora–Tyupkin procedure), a typical gauge choice being $d_{A_0} \star (A - A_0) = 0$ where A_0 is a reference connection, and \star is the Hodge duality operator corresponding to a choice of metric on M .

Local p -form symmetries For $n=2$, the only local symmetries of the BF action are the above G gauge transformations. For $n > 2$, however, there are other local symmetries associated with shifts of $B_p \in \Omega^p(M, \mathfrak{g})$ with $p = n - 2 > 0$. Indeed, integration by parts using Stokes’ theorem and $\partial M = 0$ shows that [1] is invariant under

$$A \rightarrow A, \quad B_p \rightarrow B_p + d_A \lambda_{p-1}, \quad \lambda_{p-1} \in \Omega^{p-1}(M, \mathfrak{g}) \quad [6]$$

For $p=1$, λ is a 0-form and the invariance follows. For $p > 1$, however, the gauge parameter has, in some sense, its own gauge invariance. Namely, under the shift

$$\lambda_{p-1} \rightarrow \lambda_{p-1} + d_A \lambda_{p-2} \quad [7]$$

one has

$$d_A \lambda_{p-1} \rightarrow d_A \lambda_{p-1} + [F_A, \lambda_{p-2}] \quad [8]$$

Thus for $F_A = 0$, the shift [7] has no effect on the local symmetry [6]. Likewise, for $p > 2$ the parameter λ_{p-2} itself has a similar invariance, etc. Since $F_A = 0$ is one of the classical equations of motion, the shift symmetry [6] is what is called an “on-shell reducible symmetry.” Gauge-fixing such symmetries is not straightforward, and one generally appeals to the Batalin–Vilkovisky formalism to accomplish this.

Diffeomorphisms and local symmetries One manifestation of the general covariance of the BF action [1] is the on-shell equivalence of (infinitesimal) diffeomorphisms and (infinitesimal) local symmetries. Diffeomorphisms are generated by the Lie derivative L_X along a vector field X . The action of L_X on differential forms is given by the Cartan formula $L_X = di_X + i_X d$, where $i_{(\cdot)}$ is the operation of contraction. The action of the Lie derivative on A and B can be written in gauge covariant form as

$$\begin{aligned} L_X A &= i_X F_A + d_A \lambda(X), \\ L_X B &= i_X d_A B + [B, \lambda(X)] + d_A \lambda'(X) \end{aligned} \quad [9]$$

where $\lambda(X) = i_X A$ and $\lambda'(X) = i_X B$. This shows that on-shell diffeomorphisms are equivalent to field-dependent gauge and p -form symmetries of the BF action.

The classical moduli space The classical moduli space $\mathcal{C} = \mathcal{C}(P, M, G)$ is the space of solutions to the classical equations of motion modulo the local symmetries of the action. Since the field content

and the nature of the local symmetries of the BF theory depend strongly on the dimension n of M , the structure and interpretation of the classical moduli space also depend on n .

For $n=2$, by [5] the equation of motion [2] for $B \in \Omega^0(M, \mathfrak{g})$ says that A is invariant under the infinitesimal gauge transformation generated by B . Thus if A is “irreducible,” there are no nontrivial solutions for B and, away from reducible flat connections, the classical moduli space is just the moduli space of flat connections on $P \rightarrow M$ over the surface M :

$$C_{n=2} = \mathcal{M}_{\text{flat}}(P, G) \tag{10}$$

This space may or may not be empty, depending on whether P admits flat connections or not.

For $n=3$, the equation of motion [2] for $B \in \Omega^1(M, \mathfrak{g})$ says that B is a tangent vector to the space of flat connections at the flat connection A , in the sense that under the variation $\delta A = B$, one has

$$\delta F_A = d_A B = 0 \tag{11}$$

The local G gauge symmetry and the 1-form symmetry [6] now imply that the moduli space of classical solutions can be identified with the (co-)tangent bundle of the moduli space of flat connections on $P \rightarrow M$ over the 3-manifold M :

$$C_{n=3} = T\mathcal{M}_{\text{flat}}(P, G) \tag{12}$$

In higher dimensions there appears to be less geometrical structure associated with BF theories, and all that can be said in general is that the tangent space to C_n at a solution (A, B) of the equations of motion [2] is the vector space:

$$T_{(A,B)}C_n = H_A^1(M, \mathfrak{g}) \oplus H_A^{n-2}(M, \mathfrak{g}) \tag{13}$$

where $H_A^k(M, \mathfrak{g})$ are the cohomology groups of the deformation complex

$$d_A : \Omega^*(M, \mathfrak{g}) \rightarrow \Omega^{*+1}(M, \mathfrak{g}) \tag{14}$$

associated with the flat connection A , $F_A = (d_A)^2 = 0$.

When M is topologically of the form $M = \Sigma \times \mathbb{R}$ (where one can think of \mathbb{R} as time), one has

$$T_{(A,B)}C_n = H_A^1(\Sigma, \mathfrak{g}) \oplus H_A^{n-2}(\Sigma, \mathfrak{g}) \tag{15}$$

This is naturally a symplectic vector space (necessary for a phase space), the nondegenerate antisymmetric pairing being given by Poincaré duality:

$$\omega([a_1], [b_1]; [a_2], [b_2]) = \int_{\Sigma} \text{tr}_G(a_1 \wedge b_2 - a_2 \wedge b_1) \tag{16}$$

Metric independence Perhaps the most important property of the action [1] is that, in contrast to,

for example, the usual Yang–Mills action for nonabelian gauge fields

$$S_{\text{YM}} = \frac{1}{4g^2} \int_M \text{tr}_G F_A \wedge \star F_A \tag{17}$$

it does not require a metric (or the corresponding Hodge duality operator \star) for its formulation. This makes it a candidate action for a “topological field theory,” this term loosely referring to field theories which, in a suitable sense, do not depend on additional structures imposed on the underlying space(-time) manifold M , in this case a Riemannian structure.

To establish that BF theories are “topological quantum field theories,” one needs to show that the partition function (and correlation functions) of the quantized BF theory are also metric independent. This is not completely automatic as typically the metric enters in the gauge fixing of the local symmetries of the action which is required to make the quantum theory well defined. The usual lore is that since the metric only enters through the gauge fixing and since the quantum theory should be independent of the choice of gauge, it should also be metric independent. In the case of nonabelian BF theories, the complexity of their local symmetries complicates the analysis somewhat, but it can nevertheless be shown that BF theories indeed define topological field theories also at the quantum level.

Special Features of Abelian BF Theories

All the features of nonabelian BF theories discussed above are, of course, also valid when G is abelian (with some obvious modifications and simplifications). However, when G is abelian, a more general action than [1] is possible. Indeed, although there is no obvious higher p -form analog of nonabelian gauge fields, in the abelian case $G = \text{U}(1)$ or $G = \mathbb{R}$, and the condition $F_A \in \Omega^2(M, \mathbb{R})$ can be relaxed. In particular, one can consider the actions

$$S(n, p) \equiv S(B_p, C_{n-p-1}) = \int_M B_p \wedge dC_{n-p-1} \tag{18}$$

with $B_p \in \Omega^p(M, \mathbb{R})$, $C_{n-p-1} \in \Omega^{n-p-1}(M, \mathbb{R})$, and $F_C = dC$, its $(n-p)$ -form field strength. More generally, one can also consider the hybrid action

$$S_A(n, p) = \int_M B_p \wedge d_A C_{n-p-1} \tag{19}$$

where A is a fixed (nondynamical) flat G -connection, $d_A^2 = 0$, and B and C take values in the corresponding adjoint bundle. This action can be considered as the linearization of the nonabelian BF action [1] around

the flat connection A , and it reduces to the abelian BF action [18] for $\mathfrak{g} = \mathbb{R}$.

The action is invariant under the (reducible) local symmetries

$$\begin{aligned} B_p &\rightarrow B_p + d_A \lambda_{p-1} \\ C_{n-p-1} &\rightarrow C_{n-p-1} + d_A \lambda'_{n-p-2} \end{aligned} \quad [20]$$

The space of solutions to the equations of motion $d_A C = d_A B = 0$ modulo gauge symmetries is (cf. [13]) the finite-dimensional vector space

$$C_{n,p} = H_A^p(M, \mathfrak{g}) \oplus H_A^{n-p-1}(M, \mathfrak{g}) \quad [21]$$

which is naturally symplectic for $M = \Sigma \times \mathbb{R}$.

Uses and Applications of Quantum Abelian BF Theories

Quantization of Abelian BF Theories and the Ray–Singer Torsion

We will now show that the partition function of the abelian BF theory (actually more generally that of the linearized nonabelian BF action [19]) is related to the Ray–Singer torsion of M . This requires some preparatory material on Gaussian path integrals, determinants, and gauge fixing that we present first.

In order to simplify the exposition, we assume that there are no harmonic modes, either because they have been gauged away or because the cohomology groups of d_A are trivial, $H_A^k(M, \mathfrak{g}) = 0$, that is, the deformation complex [14] is “acyclic.”

Laplacians, determinants, and the Ray–Singer torsion Choosing a Riemannian metric g (and Hodge duality operator \star) on M , the twisted Laplacian on p -forms is

$$\Delta_A^{(p)} = (d_A + d_A^*)^2 = d_A d_A^* + d_A^* d_A \quad [22]$$

where $d_A^* = \pm \star d_A \star$ is the adjoint of d with respect to the scalar product on p -forms defined by \star . This is an elliptic operator whose determinant can be defined, for example, by a ζ -function regularization. Denoting the (nonzero) eigenvalues of $\Delta_A^{(p)}$ by $\lambda_k^{(p)}$, its ζ -function is

$$\zeta^{(p)}(s) = \sum_k \left(\lambda_k^{(p)} \right)^{-s} \quad [23]$$

This converges for $\text{Re}(s)$ sufficiently large and can be analytically continued to a meromorphic function of s analytic at $s=0$, so that

$$\det \Delta_A^{(p)} := e^{-\zeta^{(p)'}(0)} \quad [24]$$

is well defined. The Ray–Singer torsion of (M, \mathfrak{g}) (with respect to the flat connection A) is then defined by

$$T_A(M) = \prod_{p=0}^n \left(\det \Delta_A^{(p)} \right)^{(-1)^p p/2} \quad [25]$$

Even though this definition depends strongly on the metric g on M , the Ray–Singer torsion has the remarkable property of being independent of g . The Ray–Singer torsion can be shown to be trivial (essentially $=1$ modulo zero-mode contributions) in even dimensions, but is a nontrivial topological invariant in odd dimensions. Henceforth, we will suppress the dependence on M and denote the n -dimensional Ray–Singer torsion by $T_A(n)$.

Gaussian path integrals and determinants The path integral for abelian BF theories is modeled on the usual formula for a δ -function

$$\delta^n(x) = \frac{1}{(\sqrt{2\pi})^n} \int_{\mathbb{R}^n} d^n \pi e^{i\pi x} \quad [26]$$

from which one deduces the Gaussian integral formula

$$\begin{aligned} &\frac{1}{(\sqrt{2\pi})^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} d^n \pi d^n x e^{i\pi D x + iK x + i\pi J} \\ &= \int_{\mathbb{R}^n} d^n x \delta^n(Dx + J) e^{iK x} \\ &= \frac{1}{\det D} e^{-iK \cdot D^{-1} J} \end{aligned} \quad [27]$$

Here, we have assumed that the operator (matrix) D is invertible. The model that one uses in the path integral is that

$$\int d[\phi] d[\chi] e^{i \int_M \phi \star D \chi} = (\det D)^{\mp 1} \quad [28]$$

where ϕ is a set of fields and the χ are a set of dual fields with D again a nondegenerate operator. The inverse determinant arises for Grassmann even fields (as in [27]), while it is the determinant that appears for Grassmann odd fields.

Gauge fixing – the Faddeev–Popov trick If the action [19], $S_A(n, p) = \int B_p d_A C_{n-p-1}$, were nondegenerate, its partition function could be defined directly by [28]. However, because of gauge invariance of the action, the kinetic term is degenerate and one needs to eliminate the gauge freedom to obtain an (at least formally) well-defined expression for the partition function. Concretely, this degeneracy can be seen by

recalling that, when there are no harmonic forms (as we have assumed), there is a unique orthogonal Hodge decomposition of a p -form $B_p \in \Omega^p(M, \mathfrak{g})$ into a sum of a d_A -exact and a d_A -coexact form:

$$B_p = d_A \lambda_{p-1} + d_A^* \tau_{p+1} \quad [29]$$

(and likewise for C). Evidently, the exact (longitudinal) parts $d_A \lambda$ of B and C do not appear in the action, and these are precisely the gauge-dependent parts of B and C under the gauge transformation [20]. Gauge fixing amounts to imposing a condition $\mathcal{F}(B_p) = 0$ on B_p that determines the longitudinal part uniquely in terms of the transversal part $d_A^* \tau$. A natural condition is

$$d_A \lambda_{p-1} = 0 \Leftrightarrow \mathcal{F}(B_p) = d_A^* B_p = 0 \quad [30]$$

A gauge-fixing condition independent of the partition function results from inserting “1” in the form of

$$1 = \int_{\mathcal{G}} d[g] \delta(\mathcal{F}(B^g)) \Delta_{\mathcal{F}}(B) \quad [31]$$

into the functional integral (the Faddeev–Popov trick), where \mathcal{G} is the gauge group. This defines the Faddeev–Popov determinant $\Delta_{\mathcal{F}}$, and the functional properties of the delta functional imply that $\Delta_{\mathcal{F}}$ is the determinant of the operator that one obtains upon gauge variation of $\mathcal{F}(B)$.

In the general case of reducible gauge symmetries, the nature of the gauge group is complicated and requires some more thought. In the irreducible case, however, that is, for $p=1$, the Lie algebra of the gauge group can be identified with $\Omega^0(M, \mathfrak{g})$, and $\Delta_{\mathcal{F}}$ is the determinant of the operator:

$$\frac{\delta \mathcal{F}}{\delta B} d_A : \Omega^0(M, \mathfrak{g}) \rightarrow \Omega^0(M, \mathfrak{g}) \quad [32]$$

For [30], this is simply the Laplacian on 0-forms, and thus

$$\Delta_{\mathcal{F}} = \det \Delta_A^{(0)} \quad [33]$$

The partition function Following the finite-dimensional model, both the δ -function implementing the gauge-fixing condition and the Faddeev–Popov determinant can be lifted into the exponential, the former by a Lagrange multiplier π [26], a Grassmann even 0-form, and the latter by a pair of Grassmann odd 0-forms c and \bar{c} [28], the ghost and antighost fields, respectively. The sum of the classical action and these gauge-fixing and ghost terms defines the (BRST-invariant) “quantum action” $S_A^q(n, p)$, and the partition function is

$$Z_A(n, p) = \int d[\phi] e^{iS_A^q(n, p)(\phi)} \quad [34]$$

where ϕ denotes collectively all the fields. Concretely, when $n=2$ and $p=0$ (or, equivalently, $p=1$), the quantum action is

$$S_A^q(2, 0) = \int B_0 d_A C_1 + \pi d_A \star C_1 + \bar{c} \star \Delta_A^{(0)} c \quad [35]$$

Likewise, for $n=3$ and $p=1$ (the only other case when the gauge symmetry is indeed irreducible), both B_1 and C_1 require separate gauge fixing, and the quantum action is

$$S_A^q(3, 1) = \int B_1 d_A C_1 + \pi d_A \star C_1 + \bar{c} \star \Delta_A^{(0)} c + \pi' d_A \star B_1 + \bar{c}' \star \Delta_A^{(0)} c' \quad [36]$$

Formally, therefore, the two-dimensional partition function is

$$Z_A(2, 0) = \frac{\det \Delta^{(0)}}{\det D_A} \quad [37]$$

where D_A is the operator:

$$D_A = \begin{pmatrix} \star d_A & \\ & \star d_A \star \end{pmatrix} : \Omega^1(M, \mathfrak{g}) \rightarrow \Omega^0(M, \mathfrak{g}) \oplus \Omega^0(M, \mathfrak{g}) \quad [38]$$

One can define the determinant of this operator as the square root of the determinant of the operator $D_A^* D_A = \Delta_A^{(1)}$, and therefore the partition function

$$Z_A(2, 0) = \det \Delta^{(0)} (\det \Delta^{(1)})^{-1/2} = T_A(2) \quad [39]$$

is equal to the two-dimensional Ray–Singer torsion [25]. In this case, it is easy to see directly that the even-dimensional Ray–Singer torsion is trivial, as one could have equally well defined the determinant of D_A as the square root of the operator $D_A D_A^* = \Delta_A^{(0)} \oplus \Delta_A^{(0)}$, which implies $Z_A(2, 0) = 1$.

In three dimensions, the two pairs of ghosts each contribute a $\det \Delta_A^{(0)}$, and thus

$$Z_A(3, 1) = \frac{(\det \Delta^{(0)})^2}{\det D_A} \quad [40]$$

where

$$D_A = \begin{pmatrix} \star d_A & d_A \\ d_A \star & 0 \end{pmatrix} : \Omega^0(M, \mathfrak{g}) \oplus \Omega^1(M, \mathfrak{g}) \rightarrow \Omega^0(M, \mathfrak{g}) \oplus \Omega^1(M, \mathfrak{g}) \quad [41]$$

is the operator acting on the fields (B_1, C_1, π, π') . As before, this operator can be diagonalized by squaring it, $D_A^* D_A = \Delta^{(0)} \oplus \Delta^{(1)}$, and thus

$$Z_A(3, 1) = (\det \Delta_A^{(0)})^{3/2} (\det \Delta_A^{(1)})^{-1/2} = T_A(3)^{-1} \quad [42]$$

is again related to the (this time genuinely nontrivial) Ray–Singer torsion.

In spite of the complications caused by reducible gauge symmetries, it can be shown that all of the above generalizes to arbitrary n and p , with the result that (for n odd)

$$Z_A(n, p) = T_A(n)^{(-1)^p} \quad [43]$$

confirming the topological nature of BF theories.

In the nonabelian case, the situation is significantly more complicated because of the complexity of the classical moduli space, the (higher cohomology) zero modes, and the on-shell reducibility of the gauge symmetries. Nevertheless, ignoring all the zero modes except those of A , that is, except the moduli m of flat connections $A(m)$, the result is similar to that in the abelian case, in that the partition function reduces to an integral over the moduli space of flat connections, with measure determined by the Ray–Singer torsion $T_{A(m)}$.

Linking Numbers as Observables of Abelian BF Theories

With the exception of $p=0$, there are no interesting “local” observables (gauge-invariant functionals of the fields C and B) in the abelian BF theory, since the gauge-invariant field strengths dC and dB vanish by the equations of motion. (For $p=0$, B is a gauge-invariant 0-form and hence $B(x)$ is a good local observable.) However, as in the Chern–Simons and Yang–Mills theories, certain (weakly) nonlocal observables such as Wilson loops are also of interest. In the case at hand (eqn [18]), we have abelian Wilson surface operators

$$W_S[B] = \int_S B, \quad W_{S'}[C] = \int_{S'} C \quad [44]$$

associated with p - and $(n-p-1)$ -dimensional submanifolds S and S' of M , respectively. These operators are gauge invariant, that is, invariant under the local symmetries [20] provided that $\partial S = \partial S' = 0$, so that S and S' represent homology cycles of M .

For $M = \mathbb{R}^n$, correlation functions of these operators are related to the topological linking number of S and S' . We choose $S = \partial\Sigma$ and $S' = \partial\Sigma'$ to be disjoint compact-oriented boundaries of oriented submanifolds Σ and Σ' of \mathbb{R}^n . We also introduce de Rham currents Δ_Σ and Δ_S (essentially distributional differential forms with δ -function support on Σ or S , respectively), characterized by the properties

$$\begin{aligned} \int_S \omega_p &= \int_M \Delta_S \wedge \omega_p \\ \int_\Sigma \omega_{p+1} &= \int_M \Delta_\Sigma \wedge \omega_{p+1} \end{aligned} \quad [45]$$

for all $\omega_k \in \Omega^k(M, \mathbb{R})$ (and likewise for S' and Σ').

Since the dimension of Σ is equal to the codimension of $S' = \partial\Sigma'$, Σ and S' will generically intersect transversally at isolated points, and we define the “linking number” of S and S' to be the intersection number of Σ and S' , expressed in terms of de Rham currents as

$$L(S, S') = \int_\Sigma \Delta_{S'} = \int_M \Delta_\Sigma \Delta_{S'} \quad [46]$$

In terms of de Rham currents, the Wilson surface operators can be written as $W_S[B] = \int_M \Delta_S \wedge B$, etc. Thus, the generating functional for correlation functions of Wilson surface operators

$$\begin{aligned} \langle e^{i\beta W_S[B]} e^{i\alpha W_{S'}[C]} \rangle \\ = \int D[C] D[B] e^{i \int_M (B dC + \alpha \Delta_{S'} C + \beta \Delta_S B)} \end{aligned} \quad [47]$$

is simply a Gaussian path integral. Using the defining properties of de Rham currents, this can be formally evaluated (using [27]) to give

$$\langle e^{i\beta W_S[B]} e^{i\alpha W_{S'}[C]} \rangle = e^{\pm i\alpha\beta L(S, S')} \quad [48]$$

As expected, correlation functions of these topological field theories encode topological information.

Uses and Applications of Classical Nonabelian BF Theories

Low-dimensional BF theories are closely related to other theories of interest, for example, the Yang–Mills theory, the Chern–Simons theory, and gravity. Here, we briefly review some of these relationships. In order to avoid the complexities of quantum nonabelian BF theories, we focus on their classical features. Brief suggestions for **further reading** are provided at the end of each subsection.

Relation with Yang–Mills Theory

In any dimension, the nonabelian BF action can be regarded as the zero-coupling limit $g^2 \rightarrow 0$ of the Yang–Mills theory since the Yang–Mills action [17] can be written in first-order form as

$$\begin{aligned} \frac{1}{4g^2} \int_M \text{tr}_G F_A \wedge \star F_A \\ \equiv \int_M \text{tr}_G [iB_{n-2} \wedge F_A + g^2 B_{n-2} \wedge \star B_{n-2}] \end{aligned} \quad [49]$$

However, whereas for $n \geq 3$ the B^2 -term breaks the p -form gauge invariance of the BF action (and thus liberates the physical Yang–Mills degrees of freedom), this limit is nonsingular in two dimensions where this p -form symmetry is absent and, indeed, both theories have zero physical degrees of freedom.

A nonsingular BF-like zero coupling limit of the Yang–Mills theory for $n \geq 3$ can be obtained by introducing an auxiliary (Stückelberg) field $\eta \in \Omega^{n-3}(M, \mathfrak{g})$ which restores the p -form gauge invariance. The resulting BF Yang–Mills action is

$$S_{\text{BFYM}} = \int_M \text{tr}_G \left[iB_{n-2} \wedge F_A + g^2 \left(B_{n-2} - \frac{1}{\sqrt{2}g} d_A \eta \right) \wedge * \left(B_{n-2} - \frac{1}{\sqrt{2}g} d_A \eta \right) \right] \quad [50]$$

This action is not only invariant under ordinary G gauge transformations, but also under the p -form gauge symmetry $B \rightarrow B + d_A \lambda$ [6] provided that η transforms as $\eta \rightarrow \eta + \sqrt{2}g\lambda$. Thus, this shift can be used to set η to zero, upon which one recovers the first-order form of the Yang–Mills action. Moreover, in the zero-coupling limit all that survives is a standard (and nontopological) minimal coupling of η to the BF action:

$$\lim_{g^2 \rightarrow 0} S_{\text{BFYM}} = \int_M \text{tr}_G [iB_{n-2} \wedge F_A + \frac{1}{2} d_A \eta \wedge * d_A \eta] \quad [51]$$

accounting for the correct number of degrees of freedom of the Yang–Mills theory (the $(n - 3)$ -form η being absent for $n = 2$).

Two-dimensional quantum BF and Yang–Mills theories have a variety of interesting topological properties. An account of some of them can be found in [Blau and Thompson \(1994\)](#) and [Witten \(1991\)](#). For a detailed discussion of the gauge symmetries and gauge fixing of the BFYM action, see [Cattaneo et al. \(1998\)](#).

Chern–Simons Theory, Gravity, and (Deformed) BF Theory

The Chern–Simons theory is a three-dimensional gauge theory. The Chern–Simons action for an H -connection C , H the gauge group, is

$$S_{\text{CS}}(C) = \int_M \text{tr}_H (C \wedge dC + \frac{2}{3} C \wedge C \wedge C) \quad [52]$$

It is invariant under the infinitesimal gauge transformations $\delta C = d_C \lambda$, $\lambda \in \Omega^0(M, \mathfrak{h})$, and the gauge-invariant equation of motion is the flatness condition $F_C = 0$. Now let $H = TG$ be the tangent bundle group $TG \sim G \times_s \mathfrak{g}$. This is a semidirect product group with G acting on \mathfrak{g} via the adjoint and \mathfrak{g} regarded as an abelian Lie algebra of translations. Thus, in terms of generators (J_a, P_a) , where the J_a are generators of G , the commutation relations are

$[J_a, J_b] = f_{ab}^c J_c$, $[J_a, P_b] = f_{ab}^c P_c$ and $[P_a, P_b] = 0$, and the curvature of the TG -connection $C = J_a A^a + P_a B^a$ is

$$F_C = J_a F_A^a + P_a d_A B^a \quad [53]$$

Thus, the equations of motion of the TG Chern–Simons theory are equivalent to the equations of motion [2] of the BF theory with gauge group G . This equivalence also holds at the level of the action:

$$\frac{1}{2} S_{\text{CS}}(C) = S_{\text{BF}}(A, B) \quad [54]$$

provided that one chooses the nondegenerate invariant scalar product to be

$$\begin{aligned} \text{tr}_{TG}(J_a P_b) &= \text{tr}_G(J_a J_b) \\ \text{tr}_{TG}(J_a J_b) &= \text{tr}_{TG}(P_a P_b) = 0 \end{aligned} \quad [55]$$

For $G = \text{SO}(3)$, TG is the Euclidean group of isometries of \mathbb{R}^3 and for $G = \text{SO}(2, 1)$, TG is the Poincaré group of isometries of the three-dimensional Minkowski space $\mathbb{R}^{2,1}$. For these gauge groups, the BF action takes the form of the three-dimensional (Euclidean or Lorentzian) Einstein–Hilbert action, with the interpretation of $B = e$ as the dreibein and $A = \omega$ as the spin connection. The equations of motion for e and ω express the vanishing of the torsion and the Riemann tensor (equivalent to the vanishing of the Ricci tensor for $n = 3$), respectively. This Chern–Simons interpretation of three-dimensional gravity extends to gravity with a cosmological constant, with H the appropriate de Sitter or anti-de Sitter isometry group ($\text{SO}(4)$, $\text{SO}(3, 1)$, or $\text{SO}(2, 2)$, depending on the signature and the sign of the cosmological constant). In terms of the BF interpretation, this corresponds to the simple topological deformation

$$S_{\mu\text{BF}}(A, B) = \int_M \text{tr}_G (B \wedge F_A + \frac{1}{3} \mu B \wedge B \wedge B) \quad [56]$$

of the BF action, which has the deformed local symmetries (cf. [5] and [6])

$$\delta A = d_A \lambda + \mu[B, \lambda'], \quad \delta B = [B, \lambda] + d_A \lambda' \quad [57]$$

A simple way to understand these symmetries is to note that the action can be written as the difference of two Chern–Simons actions:

$$\begin{aligned} S_{\text{CS}}(A + \sqrt{\mu}B) - S_{\text{CS}}(A - \sqrt{\mu}B) \\ = 4\sqrt{\mu} S_{\mu\text{BF}}(A, B) \end{aligned} \quad [58]$$

whose evident standard local gauge symmetries $\delta(A \pm \sqrt{\mu}B) = d_{A \pm \sqrt{\mu}B} \lambda^\pm$ are equivalent to [57] for $\lambda^\pm = \lambda \pm \sqrt{\mu} \lambda'$.

A detailed account of three-dimensional classical and quantum gravity can be found in [Carlip \(1998\)](#).

Relation with Gravity

Theories of two-dimensional gravity and topological gravity also have a BF formulation (Blau and Thompson 1991, Birmingham *et al.* 1991) which resembles the Chern–Simons BF formulation of three-dimensional gravity described above, the natural gauge group now being $SO(2, 1)$ or $SO(3)$ or one of its contractions.

In the first-order (Palatini) formulation, the Einstein–Hilbert action for four-dimensional gravity can be written as

$$S_{\text{EH}} = \int \text{tr}(e \wedge e \wedge F_\omega) \quad [59]$$

where e is the vierbein and ω is the spin connection. This action has the general form of a BF action with a constraint that $B = e \wedge e$ be a simple bi(co-)vector. Thus, four-dimensional general relativity can be regarded as a constrained BF theory. Although this constraint drastically changes the number of physical degrees of freedom (BF theory has zero degrees of freedom, while four-dimensional gravity has two), this is nevertheless a fruitful analogy which also lies at the heart of the spin-foam quantization approach to quantum gravity. This constrained BF description of gravity is also available for higher-dimensional gravity theories.

For further details, and references, see Freidel *et al.* (1999) and the review article (Baez 2000).

Knot and Generalized Knot Invariants

The known relationship between Wilson loop observables of the Chern–Simons theory with a compact gauge group and knot invariants (Witten 1989), and the interpretation of the three-dimensional BF theory as a Chern–Simons theory with a noncompact gauge group raise the question of the relation of observables of an $n = 3$ BF theory to knot invariants, and suggest the possibility of using an $n \geq 4$ BF theory to define higher-dimensional analogs of knot invariants. It turns out that an appropriate observable of $n = 3$ BF theory for $G = \text{SU}(2)$ is related to the Alexander–Conway polynomial. The analysis of higher-dimensional BF theories requires the full power of the Batalin–Vilkovisky (BV) formalism. BV observables generalizing Wilson loops have been shown to give rise to cohomology classes on the space of imbedded curves.

For a detailed discussion of these issues, see Cattaneo and Rossi (2001) and references therein. A relation between the algebra of generalized

Wilson loops and string topology has been investigated in Cattaneo *et al.* (2003).

See also: Batalin–Vilkovisky Quantization; BRST Quantization; Chern–Simons Models: Rigorous Results; Gauge Theories From Strings; Knot Invariants and Quantum Gravity; Loop Quantum Gravity; Moduli Spaces: An Introduction; Nonperturbative and Topological Aspects of Gauge Theory; Schwarz-Type Topological Quantum Field Theory; Spin Foams; Topological Quantum Field Theory: Overview.

Further Reading

- Baez J (2000) An introduction to spin foam models of quantum gravity and BF theory. *Lecture Notes in Physics* 543: 25–94.
- Birmingham D, Blau M, Rakowski M, and Thompson G (1991) Topological field theory. *Physics Reports* 209: 129–340.
- Blau M and Thompson G (1989) A New Class of Topological Field Theories and the Ray–Singer Torsion. *Physics Letters B* 228: 64–68.
- Blau M and Thompson G (1991) Topological gauge theories of antisymmetric tensor fields. *Annals of Physics* 205: 130–172.
- Blau M and Thompson G (1994) Lectures on 2d gauge theories: topological aspects and path integral techniques. In: Gava E, Masiero A, Narain KS, Randjbar–Daemi S, and Shafi Q (eds.) *Proceedings of the 1993 Trieste Summer School on High Energy Physics and Cosmology*, pp. 175–244. Singapore: World Scientific.
- Carlip S (1998) *Quantum Gravity in 2 + 1 Dimensions*. Cambridge: Cambridge University Press.
- Cattaneo A and Rossi C (2001) Higher-dimensional BF theories in the Batalin–Vilkovisky formalism: the BV action and generalized Wilson loops. *Communications in Mathematical Physics* 221: 591–657.
- Cattaneo A, Cotta-Ramusino P, Fucito F, Martellini M, and Rinaldi M, *et al.* (1998) Four-dimensional Yang–Mills theory as a deformation of topological BF theory. *Communications in Mathematical Physics* 197: 571–621.
- Cattaneo A, Pedrini P, and Fröhlich J (2003) Topological field theory interpretation of string topology. *Communications in Mathematical Physics* 240: 397–421.
- Freidel L, Krasnov K, and Puzio R (1999) BF description of higher-dimensional gravity theories. *Advances in Theoretical and Mathematical Physics* 3: 1289–1324.
- Horowitz GT (1989) Exactly soluble diffeomorphism invariant theories. *Communications in Mathematical Physics* 125: 417–437.
- Schwarz AS (1978) The partition function of a degenerate quadratic functional and Ray–Singer Invariants. *Letters in Mathematical Physics* 2: 247–252.
- Schwarz AS (1979) The partition function of a degenerate functional. *Communications in Mathematical Physics* 67: 1–16.
- Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 127: 351–399.
- Witten E (1991) On quantum gauge theories in two dimensions. *Communications in Mathematical Physics* 141: 153–209.

Bicrossproduct Hopf Algebras and Noncommutative Spacetime

S Majid, Queen Mary, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

One of the sources of quantum groups is a bicrossproduct construction coming in the case of Lie groups from considerations of Planck-scale physics in the 1980s. This article describes these objects and their currently known applications. See also the overview of Hopf algebras which provides the algebraic context (see Hopf Algebras and *q*-Deformation Quantum Groups).

The construction of quantum groups here is viewed as a microcosm of the problem of quantization in a manner compatible with geometry. Here quantization enters in the noncommutativity of the algebra of observables and “curvature” enters as a quantum nonabelian group structure on phase space. Among the main features of the resulting bicrossproduct models (Majid 1988) are

1. Compatibility takes the form of nonlinear “matched pair equations” generically leading to singular accumulation regions (event horizons or a maximum value of momentum depending on context).
2. The equations are solved in an “equal and opposite” form from local factorization of a larger object.
3. Different classical limits are related by observer-observed symmetry and Hopf algebra duality.
4. Nonabelian Born reciprocity re-emerges and is linked to *T*-duality.

It has also been argued that noncommutative geometry should emerge as an effective theory of the first corrections to geometry coming from any unknown theory of quantum gravity. Concrete models of noncommutative spacetime currently provide the first framework for the experimental verification of such effects. The most basic of these possible effects is curvature in momentum space or “cogravity.” We start with this.

Cogravity

We recall that curvature in space or spacetime means by definition noncommutativity among the covariant derivatives D_i . Here the natural momenta are $p_i = -i\hbar D_i$ and the situation is typified by the top line in Figure 1. There are also mixed relations between the D_i and position functions as indicated

	Position	Momentum
Gravity	Curved $\sum_{\mu} x_{\mu}^2 = \frac{1}{\gamma^2}$	Noncommutative $[p_i, p_j] = i\hbar\gamma\epsilon_{ijk}p_k$
Cogravity	Noncommutative $[x_i, x_j] = 2i\lambda\epsilon_{ijk}x_k$	Curved $\sum_{\mu} p_{\mu}^2 = \frac{1}{\lambda^2}$
Quantum mechanics	$[x_i, p_j] = i\hbar\delta_{ij}$	

Figure 1 Noncommutative spacetime means curvature in momentum space. The equations are for illustration.

for flat space in the bottom line, which is quantum mechanics (there is a similar story for quantum mechanics on a curved space). We see however a third and dual possibility – noncommutativity in position space which should be interpreted as curvature in momentum space, that is, the dual of gravity. This is an independent physical effect and comes therefore with its own length scale which we denote λ . These ideas were made precise in the mid 1990s using the quantum group Fourier transform; see Majid (2000). Here we show what is involved on three illustrative examples.

1. We consider the “spin space” algebra

$$\mathbb{R}_{\lambda}^3 : [x_i, x_j] = i2\lambda\epsilon_{ij}^k x_k$$

where $\epsilon_{12}^3 = 1$ and where it is convenient to insert a factor 2. This is the enveloping algebra $U(su_2)$, that is, just angular momentum space but now regarded “upside down” as a coordinate algebra (see Hopf Algebras and *q*-Deformation Quantum Groups). Then a plane wave is of the form

$$\psi_p = e^{ip \cdot x}, \quad p \in \mathbb{R}^3$$

where we set $\hbar = 1$ for this discussion. The momenta p_i are nothing but local coordinates for the corresponding point $e^{i\lambda p \cdot \sigma} \in SU_2$ where $\lambda\sigma$ is the representation by Pauli matrices. It is really elements of this curved space SU_2 where momenta live. Here $\mathbb{R}_{\lambda}^3 = U(su_2)$ has dual $C[SU_2]$ and Hopf algebra Fourier transform (after suitable completion) takes one between these spaces. Thus, in one direction

$$\mathcal{F}(f) = \int_{SU_2} duf(u)u \approx \int d^3p J(p)f(p) e^{ip \cdot x}$$

for f a function on SU_2 . We use the Haar measure on SU_2 . The local result on the right has J the Jacobian for the change to the local p coordinates and f is written in terms of these. Note that the coproduct in

$\mathbb{C}[SU_2]$ in terms of the p^i generators is an infinite series given by the Campbell–Baker–Hausdorff series, and not the usual linear one (this is why the measure is not the Lebesgue one). The physical content here is in the plane waves themselves, one can use any other momentum coordinates to parametrize them with the corresponding measure and coproduct. Differential operators on \mathbb{R}_λ^3 are given by the action of elements of $\mathbb{C}[SU_2]$ and are diagonal on these plane waves,

$$f \cdot \psi_p = f(\mathbf{p})\psi_p$$

which corresponds under Fourier transform simply to pointwise multiplication in $\mathbb{C}[SU_2]$. For example, the function $\lambda^{-2}(\text{tr} - 2)$ as a function on SU_2 will give a rotationally invariant wave operator which is also invariant under inversion in the group. Its value on plane waves is

$$\frac{1}{\lambda^2} \text{tr}(e^{i\lambda p \cdot \sigma} - 1) = \frac{2}{\lambda^2} (\cos(\lambda|\mathbf{p}|) - 1)$$

In the limit $\lambda \rightarrow 0$ this gives the usual wave operator on \mathbb{R}^3 .

It is also possible to put a differential graded algebra (DGA) structure of differential forms on this algebra, the natural one being

$$\begin{aligned} dx_i &= \lambda \sigma_i, & x_i \theta - \theta x_i &= i \frac{\lambda^2}{\mu} dx_i \\ (dx_i)x_j - x_j dx_i &= i \lambda \epsilon_{ij}^k dx_k + i \mu \delta_{ij} \theta \end{aligned}$$

where θ is the 2×2 identity matrix which, together with the Pauli matrices σ_i , completes the basis of left-invariant 1-forms. The 1-form θ provides a natural time direction, even though there is no time coordinate, and the new parameter $\mu \neq 0$ appears as the freedom to change its normalization. The partial derivatives ∂^i are defined by

$$d\psi(x) = (\partial^i \psi) dx_i + (\partial^0 \psi) \theta$$

and act diagonally on plane waves as

$$\partial^i = \frac{i}{2\lambda} \text{tr}(\sigma_i(\cdot)) = i \frac{p^i}{\lambda|\mathbf{p}|} \sin(\lambda|\mathbf{p}|)$$

while $\partial^0 = i\mu(\text{tr} - 2)/2\lambda^2$ is computed as above.

Note that μ cannot be taken to be zero due to an anomaly for translation invariance of the DGA. It is in fact a typical feature of noncommutative differential geometry that there is a 1-form θ generating d by commutator which can be required as an extra cotangent direction with its associated partial derivative an induced Hamiltonian. In the present model we have

$$\partial^0 \psi = i \frac{\mu}{2} \sum_i (\partial^i)^2 \psi + O(\lambda^2)$$

which is of the form of Schrödinger’s equation with respect to an auxiliary time variable and for a particle with mass $1/\mu$.

The reader may ask what happens to the Euclidean group of translations and rotations in this context. From the above we find that $U_\lambda(\text{poinc}_3) = \mathbb{C}[SU_2] \bowtie U(\mathfrak{su}_2)$, the semidirect product generated by translations ∂^i and usual rotations. This in turn is the quantum double $D(U(\mathfrak{su}_2))$ of the classical enveloping algebra, and as such a quantum group with braiding etc. (see Hopf Algebras and q -Deformation Quantum Groups). This quantum double has been identified as part of an effective theory in $2 + 1$ quantum gravity in a Euclidean version based on Chern–Simons theory with Lie algebra poinc_3 and the spin space algebra proposed as an effective theory for this. The quotient of \mathbb{R}_λ^3 by an allowed value of the quadratic Casimir x^2 (which then makes it a matrix algebra) is called a “fuzzy sphere” and appears as a “world-volume algebra” in certain string theories and reduced matrix models. The noncommutative differential geometry that we have described is due to Batista and the author.

2. We take the same type of construction to obtain the “bicrossproduct model” spacetime algebra

$$\mathbb{R}_\lambda^{1,3} : \quad [t, x_i] = i\lambda x_i, \quad [x_i, x_j] = 0$$

These are the relations of a Lie algebra b_+ (say) but again regarded as coordinates on a noncommutative spacetime. Here λ is a timescale which can be written as a mass scale $\kappa = 1/\lambda$ instead. We parametrize the plane waves as

$$\psi_{p,p^0} = e^{ip \cdot x} e^{ip^0 t}, \quad \psi_{p,p^0} \psi_{p',p'^0} = \psi_{p+e^{-\lambda p^0} p', p^0+p'^0}$$

which identifies the p^μ as the coordinates of the nonabelian group $B_+ = \mathbb{R} \bowtie_\lambda \mathbb{R}^3$ with Lie algebra b_+ . The group law in these coordinates is read off as usual from the product of plane waves, which also gives the coproduct of $\mathbb{C}[B_+]$ on the p^μ . We have parametrized plane waves in this way (rather than the canonical way by the Lie algebra as before) in order to have a more manageable form for this. We do pay a price that in these coordinates group inversion is not simply $-p^\mu$, but

$$(p, p^0)^{-1} = (-e^{\lambda p^0} p, -p^0)$$

which is also the action of the antipode S on the abstract p^μ generators.

In particular, the right-invariant Haar measure on B_+ in these coordinates is the usual $d^4 p$ so the

quantum group Fourier transform reduces to the usual one but normal ordered,

$$\mathcal{F}(f) = \int_{\mathbb{R}^4} d^4p f(p)e^{ip \cdot x} e^{ip^0 t}$$

(one can also Fourier transform with respect to the left-invariant measure $d^4p e^{3\lambda p^0}$ on B_+). The inverse is again given in terms of the usual inverse transform if we specify general fields ψ in $\mathbb{R}_\lambda^{1,3}$ by normal ordering of usual functions, which we shall do. As before, the action of elements of $\mathbb{C}[B_+]$ defines differential operators on $\mathbb{R}_\lambda^{1,3}$ and these act diagonally on plane waves.

We also have a natural DGA with

$$(dx_j)x_\mu = x_\mu dx_j, \quad (dt)x_\mu - x_\mu dt = i\lambda dx_\mu$$

which leads to the partial derivatives

$$\begin{aligned} \partial^j \psi &:= \frac{\partial}{\partial x_j} \psi(x, t) := ip^j \cdot \psi \\ \partial^0 \psi &:= \frac{\psi(x, t + i\lambda) - \psi(x, t)}{i\lambda} := \frac{i}{\lambda} (1 - e^{-\lambda p^0}) \cdot \psi \end{aligned}$$

for normal-ordered polynomial functions ψ or in terms of the action of the coordinates p^μ in $\mathbb{C}[B_+]$. These ∂^μ do respect our implicit $*$ -structure (unitarity) on $\mathbb{R}_\lambda^{1,3}$ but in a Hopf algebra sense which is not the usual sense, since the action of the antipode S is not just $-p^\mu$. This can be remedied by using adjusted derivatives $L^{-(1/2)}\partial^\mu$ where

$$L\psi := \psi(x, t + i\lambda) := e^{-\lambda p^0} \cdot \psi$$

In this case the natural 4D Laplacian is $L^{-1}((\partial^0)^2 - \sum_i (\partial^i)^2)$, which acts on plane waves as

$$-\frac{2}{\lambda^2} (\cosh(\lambda p^0) - 1) + p^2 e^{\lambda p^0}$$

where

$$p^2 = \sum_{i=1}^3 p_i^2$$

This deforms the usual Laplacian in such a way as to remain invariant under the Lorentz group (which now acts nonlinearly on B_+ in this model) and under group inversion.

This model may provide the first experimental test for noncommutative spacetime and cogravity. For the analysis of an experiment, we assume the identification of noncommutative waves in the above normal-ordered form with classical ones that a detector might register. In that case one may argue (Amelino-Camelia and Majid 2000) that the dispersion relation for such waves has the classical derivation as $\partial p^0 / \partial p^i$ which now computes as propagation speed for a massless particle:

$$\left| \frac{\partial p^0}{\partial p} \right| = e^{\lambda p^0}$$

in units where 1 is the usual speed of light. So the prediction is that the speed of light depends on energy. What is remarkable is that even if $\lambda \sim 10^{-44}$ s (the Planck timescale), this prediction could in principle be tested, for example using γ -ray bursts. These are known in some cases to travel cosmological distances before arriving on Earth, and have a spread of energies from 0.1–100 MeV. According to the above, the relative time delay Δ_t on traveling distance L for frequencies corresponding to $p^0, p^0 + \Delta p^0$ is

$$\Delta_t \sim \lambda \Delta p^0 \frac{L}{c} \sim 10^{-44} \text{s} \times 100 \text{ MeV} \times 10^{10} \text{y} \sim 1 \text{ms}$$

which is in principle observable by statistical analysis of a large number of bursts correlated with distance (determined, e.g., by using the Hubble telescope to lock in on the host galaxy of each burst). Although the above is only one of a class of predictions, it is striking that even Planck-scale effects are now in principle within experimental reach.

We now explain what happens to the full Poincaré symmetry here. The nonlinear action of the Lorentz group on B_+ Fourier transforms to an action on the generators of $\mathbb{R}_\lambda^{1,3}$, which combines with the above action of the p^μ to generate an entire Poincaré quantum group $U(so_{1,3}) \bowtie \mathbb{C}[B_+]$. We will say more about its “bicrossproduct” structure in a later section. The above wave operator in momentum space is the natural Casimir in these momentum coordinates. A common mistake in the literature for this model is to suppose that the Casimir relation alone amounts to a physical prediction, whereas in fact the momentum coordinates are arbitrary and have meaning only in conjunction with the plane waves that they parametrize. The deformed Poincaré as an algebra alone is actually isomorphic to the undeformed one by a different choice of generators, so by itself has no physical content; one needs rather the noncommutative spacetime as well. Prior work on the relevant deformed Poincaré algebra either did not consider it acting on spacetime or took it acting on classical (commutative) Minkowski spacetime with inconsistent results (there is no such action as a quantum group).

The above model was introduced by Majid and Ruegg (1994) and later tied up with a dual approach of Woronowicz. There is also a previous “ κ -Poincaré” version of the Hopf algebra alone obtained (Lukierski *et al.* 1991) in another context (by contraction of $U_q(so_{2,3})$) but with fundamentally different generators and relations and hence different physical content (e.g., the Lorentz

generators there do not close among themselves but mix with momentum).

3. The usual Heisenberg algebra of quantum mechanics is another possible noncommutative (phase) space; one may also take the same algebra and view it as a noncommutative spacetime, so:

$$\mathbb{R}_\theta^{1,3}: [x_\mu, x_\nu] = i\theta_{\mu\nu}$$

for any antisymmetric tensor $\theta_{\mu\nu}$. This is not a Hopf algebra but it turns out that this model can also be completely solved by Hopf algebra methods, namely the theory of covariant twists. Twist models also include versions of the noncommutative torus studied by Connes, and related θ -spaces, which are nontrivial at the level of C^* -algebras. However, at an algebraic level, all covariant structures are automatically provided by applying the twisting functor \mathcal{T} to the desired classical construction (see Hopf Algebras and q -Deformation Quantum Groups). This is not usually appreciated in the physics literature on such models, but see Oeckl (2000).

Thus, consider $H = U(\mathbb{R}^{1,3})$ with generators $p^\mu = -i\partial^\mu$ acting as usual on functions on Minkowski space. It has a cocycle

$$F = e^{(i/2)p^\mu \otimes p^\nu \theta_{\mu\nu}}$$

which induces a new product \bullet on functions by $\phi \bullet \psi = \cdot(F^{-1}(\phi \otimes \psi))$. This is just the standard Moyal product, in the present case on $\mathbb{R}^{1,3}$, viewed as a covariant twist using Hopf algebra methods. The Hopf algebra $U(\mathbb{R}^{1,3})$ in principle has a twisted coproduct given by $\Delta_F = F(\Delta(\))F^{-1}$ but this does not change as the algebra is commutative.

Next, H also acts covariantly on $\Omega(\mathbb{R}^{1,3})$, the usual algebra of differential forms, and twisting this in the same way gives

$$\psi(x) \bullet dx_\mu = \psi dx_\mu = (dx_\mu)\psi = (dx_\mu) \bullet \psi$$

unchanged. This is because no terms higher than $p^\mu \otimes p^\nu \theta_{\mu\nu}$ contribute and then $d(1) = 0$. The associated partial derivatives defined by d are likewise unchanged and act in the usual way as derivations with respect to both the \bullet product and the undeformed product. The result may look different when the same $\psi(x)$ is expressed as a function of the variables with the \bullet product. In other words, the only deformation comes from the Moyal product itself, with the rest being automatic. Moreover, the plane waves themselves are unchanged because $(x \cdot k)^\bullet = (x \cdot k)$ due to θ being antisymmetric. Hence,

$$\psi_k(x) = e^{\bullet ix \cdot k} = e^{ix \cdot k}, \quad p^\mu \psi_k(x) = k^\mu \psi_k(x)$$

where $p^\mu = -i\partial^\mu$. The wave operator $-\partial_\mu \partial^\mu$ is therefore given by the action of $p_\mu p^\mu$ and has value $k_\mu k^\mu$ as usual on plane waves. On the other hand,

$$\psi_k \bullet \psi_{k'} = e^{(i/2)k^\mu k'^\nu \theta_{\mu\nu}} \psi_{k+k'}$$

or in algebraic terms the twist functor \mathcal{T} applied to the Fourier transform implies also a twisted coproduct or coaddition law for the abstract k^μ generators, now different from the linear one for the covariance momentum operators p^μ . This leads to some of the more interesting features of the model.

One immediately also has a Poincaré quantum group here, $U_\theta(\text{poinc}_{1,3})$, obtained by similarly twisting the classical $U(\text{poinc}_{1,3})$. We just view F as living here rather than in the original H . The translation sector is unchanged as before but if $M^{\alpha\beta}$ are the usual Lorentz generators, then

$$\begin{aligned} \Delta_F M^{\alpha\beta} &= M^{\alpha\beta} \otimes 1 + 1 \otimes M^{\alpha\beta} \\ &+ \frac{1}{2}(p^\alpha \otimes \theta^\beta_{\ \mu} p^\mu - \theta^\beta_{\ \mu} p^\mu \otimes p^\alpha) \\ &- \frac{1}{2}(p^\beta \otimes \theta^\alpha_{\ \mu} p^\mu - \theta^\alpha_{\ \mu} p^\mu \otimes p^\beta) \end{aligned}$$

using the metric $\eta_{\mu\nu}$ to raise or lower indices. The antipode is also modified according to the theory in Majid (1995). The relations in the Poincaré algebra are not modified (so, e.g., $p_\mu p^\mu$ will remain central). Any construction originally Poincaré covariant becomes covariant under this twisted one after application of the twisting functor. As with the differentials above, the action on $\mathbb{R}_\theta^{1,3}$ is not actually modified but may appear so when functions are expressed in terms of the \bullet product.

The above model is popular at the time of writing in connection with string theory. Here, an effective description of the endpoints of open strings landing on a fixed 4-brane has been modeled conveniently in terms of the \bullet product above (Seiberg and Witten 1999). It should be borne in mind, however, that this fixed 4-brane lives in some of the higher dimensions of the string spacetime, so this is not necessarily a prediction of noncommutative spacetime $\mathbb{R}^{1,3}$.

In fact, a proposal superficially similar to $\mathbb{R}_\theta^{1,3}$ above was already proposed in Snyder (1947). Here

$$[x^\mu, x^\nu] = i\lambda^2 M^{\mu\nu}$$

where λ is our length scale and the $M^{\mu\nu}$ are now operators with the usual commutation rules for the Lorentz algebra with themselves and with x^μ and the momenta p^μ . The latter obey

$$[p^\mu, x^\nu] = i(\eta^{\mu\nu} - \lambda^2 p^\mu p^\nu), \quad [p^\mu, p^\nu] = 0$$

so the entire Poincaré algebra is undeformed but the phase-space relations are deformed. Snyder also constructed the orbital angular momentum realization $M^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu$. This model is not a proposal for a noncommutative spacetime because the algebra does not even close among the x^μ . Rather it is a proposal for “mixing” of position and Lorentz generators. On the other hand (which was the point of view in Snyder (1947)), in any representation of the Poincaré algebra, the $M^{\mu\nu}$ become operators and in some sense numerical. The rotational sector has discrete eigenvalues as usual, so to this extent the spacetime has been discretized. Although not fitting into the methods in this article, it is also of interest that the relations above were motivated by considering p^μ as coordinates projected from a 5D flat space to de Sitter space and x^μ as the 5-component of orbital angular momentum in the flat space.

To conclude this section, let us note that there are further models that we have not included for lack of space. One of them is a much-studied $\mathbb{R}_q^{1,3}$ in which t is central but the x_i enjoy complicated q -relations best understood as q -deformed Hermitian matrices. One of the motivations in the theory was the result in Majid (1990) that q -deformation could be used to regularize infinities in quantum field theory as poles at $q=1$. Another entire class is to use noncommutative geometry and quantum group methods on finite or discrete spaces. Unlike lattice theory where a finite lattice is viewed as approximation, these models are not approximations but exact noncommutative geometries valid even on a few points. The noncommutativity enters into the fact that finite differences are bilocal and hence naturally have different left and right multiplications by functions. Both aspects are mentioned briefly in the overview article (see Hopf Algebras and q -Deformation Quantum Groups). Also, on the experimental front, another large area that we have not had room to cover is the prediction of modified uncertainty relations both in spacetime and phase space (Kempf et al. 1995).

Moreover, for all of the models above, once one has a noncommutative differential calculus one may proceed to gauge theory etc., on noncommutative spacetimes, at least at the level where a connection is a noncommutative (anti-Hermitian) 1-form α . Gauge transformations are invertible (unitary) elements u of the noncommutative “coordinate algebra” and the connection and curvature transform as

$$\alpha \rightarrow u^{-1}\alpha u + u^{-1} du$$

$$F(\alpha) = d\alpha + \alpha \wedge \alpha \rightarrow u^{-1}F(\alpha)u$$

The full extent of quantum bundles and gravity (see Quantum Group Differentials, Bundles and Gauge Theory) and quantum field theory is not always possible, although both have been done for covariant twist examples (for functorial reasons) and for small finite sets. For the first two models above, for example, it is not clear at the time of writing how to interpret scattering when the addition of momenta is nonabelian.

Matched Pair Equations

Although we have presented noncommutative spacetime first, the first actual application of quantum group methods to Planck-scale physics was the Planck-scale Hopf algebra obtained by a theory of bicrossproducts. Like the Snyder model, the intention here was to deform phase space itself, but since then bicrossproducts have had many further applications. The main ingredient here is the notion of a pair of groups (G, M) , say, acting on each other as we explain now. The mathematics here goes back to the early 1910s in group theory, but also arose in mathematical physics as a toy version of Einstein’s equation in the sense of compatibility between quantization and curvature (see the next section).

By definition, (G, M) are a matched pair of groups if there are left and right actions

$$M \xleftarrow{\triangleleft} M \times G \xrightarrow{\triangleright} G$$

of each group on the set of the other, such that

$$s \triangleleft e = s, \quad e \triangleright u = u, \quad s \triangleright e = e, \quad e \triangleleft u = e$$

$$(s \triangleleft u) \triangleleft v = s \triangleleft (uv), \quad s \triangleright (t \triangleright u) = (st) \triangleright u$$

$$s \triangleright (uv) = (s \triangleright u)((s \triangleleft u) \triangleright v)$$

$$(st) \triangleleft u = (s \triangleleft (t \triangleright u))(t \triangleleft u)$$

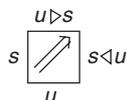
for all $u, v \in G, s, t \in M$. Here e denotes the relevant group unit element. As a first application of such data, one may make a “double cross product group” $G \bowtie M$ with product

$$(u, s) \cdot (v, t) = (u(s \triangleright v), (s \triangleleft v)t)$$

and with G, M as subgroups. Since it is built on the direct product space, the bigger group factorizes into these subgroups. Conversely, if X is a group factorization such that the product $G \times M \rightarrow X$ is bijective, each group acts on the other by actions $\triangleright, \triangleleft$ defined by $su = (s \triangleright u)(s \triangleleft u)$ for $u \in G$ and $s \in M$, where s, u are multiplied in X and the product is factorized as something in G and something in M . So finite group matched pairs are equivalent to group factorizations. In the Lie group context, the

corresponding system of differential equations is equivalent to a local factorization.

There is a nice graphical representation of the matched pair conditions which relates to “surface integration.” Thus, consider squares



labeled by elements of M on the left edge and elements of G on the bottom edge. We can fill in the other two edges by thinking of an edge transformed by the other edge as it goes through the square either horizontally or vertically, the two together is the surface transport \Rightarrow across the square. The matched pair equations have the meaning that a square can be subdivided either vertically or horizontally as shown in Figure 2, where the labeling on vertical edges is to be read from top down. The transport operation here is nothing other than normal ordering in the factorizing group. In the Lie setting, it means that the equations can be solved from infinitesimal solutions (a matched pair of Lie algebras) by a simultaneous double integration over the group (i.e., building up a large box from many small ones). If one considers solving the quantum Yang–Baxter equations on groups, they appear in this notation as an equality of surface transport going two ways around a cube, and the classical Yang–Baxter equations as curvature of the underlying higher-order connection.

Also in this notation there is a bicrossproduct quantum group defined in Figure 3, at least when M is finite. The expressions are considered zero unless the juxtaposed edges have the same group labels. In that case, the product is a semidirect product algebra $C(M) \bowtie CG$ of functions on M by the group algebra of G . The coproduct is the adjoint of

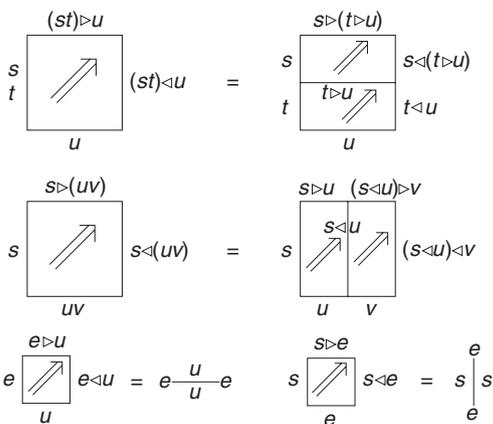


Figure 2 Matched pair condition as a subdivision property.

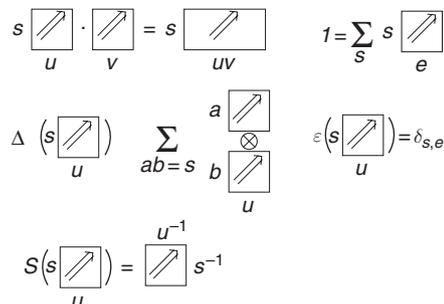


Figure 3 Bicrossproduct Hopf algebra showing horizontal product and vertical coproduct as an “unproduct.”

this, so is a semidirect coalgebra $C(M) \bowtie CG$. Hence the two together are denoted $C(M) \bowtie CG$. The dual needs G finite and has the same form but with vertical and horizontal compositions interchanged, that is, a bicrossproduct $CM \bowtie C(G)$. Both Hopf algebras have the above labeled squares as basis.

It is possible to generalize both bicrossproducts and double cross products associated to matched pairs to general Hopf algebras $H_1 \bowtie H_2$ and $H_1 \bowtie H_2$, respectively, where H_1, H_2 are Hopf algebras (see Majid 1990) and to relate the two in general by dualization of one factor. Another general result (Majid 1995) is that $H_1 \bowtie H_2$ acts covariantly on the algebra H_1^* from the right, or $H_1 \bowtie H_2$ acts covariantly on H_2^* from the left. A third general result is that bicrossproducts solve the extension problem

$$H_1 \rightarrow H \rightarrow H_2$$

meaning that such a Hopf algebra H subject to some technical requirements (such as an algebra splitting map $H_2 \rightarrow H$) is of the form $H \cong H_1 \bowtie H_2$. The theory was also extended to include cocycle bicrossproducts at the end of the 1980s (by the author). The finite group case, however, was first found by Kac and Paljutkin (1966) in the Russian literature and later rediscovered independently in Takeuchi (1981) and in the course of Majid (1988).

The Planck-Scale Hopf Algebra

We consider a quantum algebra of observables H and ask when it is a Hopf algebra extending some classical position coordinate algebra $C[M]$ and some possibly noncommutative momentum coordinate algebra $U(\mathfrak{g})$ in the form of a strict extension

$$C[M] \rightarrow H \rightarrow U(\mathfrak{g})$$

From the theory above this problem is governed by local solutions of the matched pair equations on (G, M) . It requires that $H \cong C[M] \bowtie U(\mathfrak{g})$ as an algebra, that is,

the quantization of a particle moving on orbits in M under some action of G (in an algebraic setting, or one can use von Neumann or C^* -algebras etc.). And it requires the classical phase space to be a nonabelian or “curved” group $M \bowtie \mathfrak{g}^*$. This extends to a coproduct on H which becomes the bicrossproduct Hopf algebra $C[M] \bowtie U(\mathfrak{g})$. In this way, the problem which was open at the start of the 1980s of finding true examples of Hopf algebras was given a physical interpretation as being equivalent to finding quantum-mechanical systems reconciled with curvature, and the equations that governed this were the matched pair ones (Majid 1988).

We still have to solve these equations. In the Lie case, they mean a pair of cross-coupled first-order equations on $G \times M$. These can be solved locally as a double-holonomy construction in line with the surface transport point of view, but are nonlinear typically with singularities in the non-compact case. The equations are also symmetric under interchange of G, M so Born reciprocity between position and momentum is extended to the quantum system with generally “curved” position and momentum spaces. Moreover, in so far as Einstein’s equation $G_{\mu\nu} = 8\pi T_{\mu\nu}$ is also a compatibility between a quantity in position space and a quantity originating (ultimately) in momentum space, the matched pair equations can be viewed as a toy version of these.

Let us note that the reason to look for H a Hopf algebra in the first place, aside from the reasons already given, is for observer-observed symmetry (this was put forward as a postulate for Planck-scale physics). Thus, H^* is also an algebra of observables of some dual system, in our case $U(\mathfrak{m}) \bowtie C[G]$ or particles in G moving on orbits under M . Thus, Born reciprocity is truly implemented in the quantum/curved system by Hopf algebra duality. Put another way, Hopf algebras are the simplest objects after abelian groups that admit Fourier transform (see Hopf Algebras and q -Deformation Quantum Groups) and we require this on phase space if Born reciprocity is to be extended to the quantum/curved system.

The Planck-scale Hopf algebra is the simplest example of these ideas (Majid 1988). Here $G = M = \mathbb{R}$ and the matched pair equations can be solved completely. The general solution is

$$\hat{p} = i\hbar(1 - e^{-\gamma x}) \frac{\partial}{\partial x}, \quad \hat{x} = \frac{i}{\hbar}(1 - e^{-\hbar\gamma p}) \frac{\partial}{\partial p}$$

for the action of one group with generator p on functions of x in the other group and vice-versa. It has two parameters which we have denoted as \hbar and

a background curvature scale γ , and the corresponding bicrossproduct $C[p] \bowtie C[x]$ is

$$\begin{aligned} [p, x] &= i\hbar(1 - e^{-\gamma x}), & \Delta x &= x \otimes 1 + 1 \otimes x \\ \Delta p &= p \otimes e^{-\gamma x} + 1 \otimes p, & \epsilon x &= \epsilon p = 0 \\ Sx &= -x, & Sp &= -pe^{\gamma x} \end{aligned}$$

where we should allow power series or take $e^{\gamma x}$ as an invertible generator.

It is important to note that the matched pair equations here have only this solution and it is necessarily singular at $p=0$ or $x=0$. The interpretation in position space is as follows. Consider an infalling particle of mass m with fixed momentum $p = mv_\infty$ (in terms of the velocity at infinity). By definition, p is the free-particle momentum and acts on \mathbb{R} as above. This corresponds to a free-particle Hamiltonian $\hat{p}^2/2m$ and induces

$$\begin{aligned} \dot{p} &= 0 \\ \dot{x} &= \frac{p}{m}(1 - e^{-\gamma x}) = v_\infty \left(1 - \frac{1}{1 + \gamma x + \dots}\right) \end{aligned}$$

at the classical level. We see that the particle takes an infinite time to reach the origin, which is an accumulation point. This can be compared with the formula in standard radial infalling coordinates

$$\dot{x} = v_\infty \left(1 - \frac{1}{1 + \frac{c^2 x}{2GM}}\right)$$

for distance x from the event horizon of a black hole of mass M (here G is Newton’s constant and c the speed of light). So $\gamma \sim c^2/GM$ and for the sake of further discussion we will use this value. With a little more work, one can then see that

$$\begin{array}{l} mM \ll m_p^2 \\ C[x] \bowtie C[p] \begin{array}{l} \rightarrow C[x] C[p] \text{ usual qu. mech.} \\ \leftarrow C(X) \text{ usual curved geometry} \end{array} \\ mM \gg m_p^2 \end{array}$$

where m_p is the Planck mass of the order of 10^{-5} g and $X = \mathbb{R} \bowtie \mathbb{R}$ is a nonabelian group. In the first limit, the particle motion is not detectably different from usual flat space quantum mechanics outside the Compton wavelength from the origin. In the second limit, the estimate is such that noncommutativity would not show up for length scales much larger than the background curvature scale.

This Hopf algebra is also the simplest way to extend classical position $C[x]$ and momentum $C[p]$ in the sense above. In other words, requiring to maintain observer-observed symmetry or Born reciprocity throws up both quantum mechanics (in the form of \hbar) and something with the flavor of

gravity (in the form of γ) and both are required for a nontrivial Hopf algebra. Moreover, the construction necessarily has a self-dual form and indeed the dually paired Hopf algebra is $C[p] \bowtie C[x]$ with new parameters $\hbar' = 1/\hbar$ and $\gamma' = \hbar\gamma$ if we take the standard pairing x, p across the two algebras. Hopf algebra duality realized by the quantum group Fourier transform \mathcal{F} takes one between the two models.

Bicrossproduct Poincaré Quantum Groups

Another example from the 1980s in the same family as the Planck-scale Hopf algebra is $G = SU_2$ and $M = B_+$, a nonabelian version of \mathbb{R}^3 with Lie algebra b_+ of the form

$$[x_3, x_i] = i\lambda x_i, \quad [x_i, x_j] = 0$$

for $i = 1, 2$. The required solution of the matched pair equations was found in Majid (1990) and has a nonlinear action of rotations on B_+ . The interpretation of $C[B_+] \bowtie U(su_2)$ is of particles moving along orbits which are deformed spheres in B_+ , and there is a dual model where particles move instead on orbits in SU_2 under the action of b_+ . Moreover, from the general theory of bicrossproducts, we automatically have a covariant action of $C[B_+] \bowtie U(su_2)$ on the auxiliary noncommutative space $\mathbb{R}_\lambda^3 = U(b_+)$ with relations as above.

The quantum group here was actually obtained as a Hopf-von Neumann algebra but we limit ourselves to the underlying algebraic version. Also, there is of course nothing stopping one considering this Hopf algebra equally well as $U_\lambda(\text{poinc}_3)$, that is, a deformation of the group of motions on \mathbb{R}^3 , rather than as an algebra of observables. The only difference is to denote the generators of $C[B_+]$ by the symbols p^i , reserving x_i instead for the auxiliary noncommutative space. We lower i, j, k indices using the Euclidean metric. Then the bicrossproduct has the form

$$[p_i, p_j] = 0, \quad [M_i, M_j] = i\epsilon_{ij}^k M_k$$

$$[M_3, p_j] = i\epsilon_{3j}^k p_k, \quad [M_i, p_3] = i\epsilon_{i3}^k p_k$$

as usual, for $i, j = 1, 2, 3$, and the modified relations

$$[M_i, p_j] = \frac{i}{2} \epsilon_{ij}^3 \left(\frac{1 - e^{-2\lambda p_3}}{\lambda} - \lambda p^2 \right) + i\lambda \epsilon_i^{k3} p_j p_k$$

for $i, j = 1, 2$ and $p^2 = p_1^2 + p_2^2$. The coproducts are

$$\Delta M_i = M_i \otimes e^{-\lambda p_3} + \lambda M_3 \otimes p_i + 1 \otimes M_i$$

$$\Delta p_i = p_i \otimes e^{-\lambda p_3} + 1 \otimes p_i$$

for $i = 1, 2$ and the usual additive ones for p_3, M_3 . There is also an appropriate counit and antipode. The deformed spheres under the nonlinear rotation in Majid (1990) are constant values of the Casimir for the above algebra. This is

$$\frac{2}{\lambda^2} (\cosh(\lambda p_3) - 1) + p^2 e^{\lambda p_3}$$

which from the group of motions point of view generates the noncommutative Laplacian when acting on \mathbb{R}_λ^3 . The model here is a Euclidean inhomogeneous one.

The four-dimensional (4D) version $U(so_{1,3}) \bowtie C[B_+]$ of this construction (Majid and Ruegg 1994) is again linked to Planck-scale predictions, this time as a generalized symmetry. In terms of translation generators p^μ , rotations M_i and boosts N_i we have

$$[p^\mu, p^\nu] = 0, \quad [M_i, M_j] = i\epsilon_{ij}^k M_k$$

$$[N_i, N_j] = -i\epsilon_{ij}^k M_k, \quad [M_i, N_j] = i\epsilon_{ij}^k N_k$$

$$[p^0, M_i] = 0, \quad [p^i, M_j] = i\epsilon^i_{jk} p^k, \quad [p^0, N_i] = -ip_i$$

as usual, and the modified relations and coproduct

$$[p^i, N_j] = -\frac{i}{2} \delta_j^i \left(\frac{1 - e^{-2\lambda p^0}}{\lambda} + \lambda p^2 \right) + i\lambda p^i p_j$$

$$\Delta N_i = N_i \otimes 1 + e^{-\lambda p^0} \otimes N_i + \lambda \epsilon_{ijk} p^j \otimes M_k$$

$$\Delta p^i = p^i \otimes 1 + e^{-\lambda p^0} \otimes p^i$$

and the usual additive coproducts on p^0, M_i . This time the Lorentz group orbits in B_+ are deformed hyperboloids rather than deformed spheres, and the Casimir that controls this has the same form as above but with $-$ in the cosh term, that is, the model is a Lorentzian one. We know from the general theory of bicrossproducts that this Hopf algebra acts on $U(b_+) = \mathbb{R}_\lambda^{1,3}$ the spacetime in the section ‘‘Cogravity,’’ and the Casimir induces the wave operator as we have seen there.

Let us look a bit more closely at the deformed hyperboloids. Because neither group here is compact, one expects from the general theory of bicrossproducts to have limiting accumulation regions. This is visible in the contour plot of p^0 against $|p|$ in Figure 4, where the $p^0 > 0$ mass shells are now cups with almost vertical walls, compressed into the vertical tube

$$|p| < \lambda^{-1}$$

In other words, the 3-momentum is bounded above by the Planck momentum scale (if λ is the Planck time). Indeed, the light-cone equation (setting the Casimir to zero) reads $\lambda|p| = 1 - e^{-\lambda p_3}$ so this is

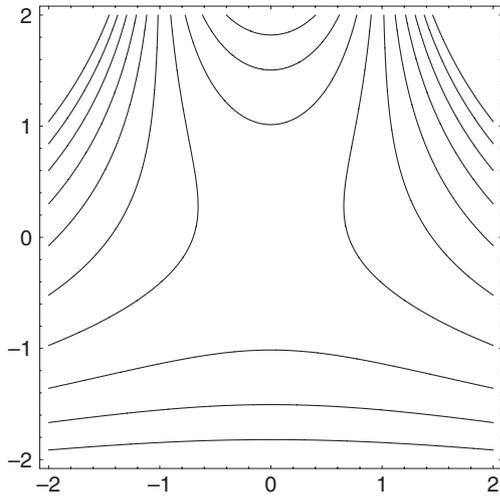


Figure 4 Deformed mass-shell orbits in the bicrossproduct curved momentum space for $\lambda = 1$.

immediate. Nevertheless, this observation is so striking that the bicrossproduct model has been dubbed “doubly special” and spawned the search for other such models. Such accumulation regions are a main discovery of the noncompact bicrossproduct theory visible already in the Planck-scale Hopf algebra. The model further confirms the role of the matched pair equations as a toy version of Einstein’s.

Poisson–Lie T-Duality

We have explained in Section 3 that the matched pair equations are equivalent to a local factorization of Lie groups, with the action and back-reaction created “equally and oppositely” from this. For the two models in the last section, these are $SL_2(\mathbb{C})$ factorizing as SU_2 and a 3D B_+ , and $SO_{2,3}$ locally as $SO_{1,3}$ and a 4D B_+ . The first of these examples is in fact one of a general family based on the Iwasawa decomposition $G_{\mathbb{C}} = G \bowtie G^*$ where G is a compact Lie group with complexification $G_{\mathbb{C}}$ and G^* a certain solvable group. From this, one may construct a solution (G, G^*) of the matched pair equations and bicrossproduct quantum group

$$\mathbb{C}[G^*] \bowtie U(\mathfrak{g})$$

associated to all complex simple Lie algebras. This is again part of the bicrossproduct theory from the 1980s. On the other hand, the Lie algebra \mathfrak{g}^* here can be identified with the dual of \mathfrak{g} in which case its Lie algebra corresponds to a Lie coproduct $\delta: \mathfrak{g} \rightarrow \mathfrak{g} \otimes \mathfrak{g}$ and makes (\mathfrak{g}, δ) into a Lie bialgebra in the sense of Drinfeld. This δ exponentiates to a Poisson bracket on G making it a “Poisson–Lie

group” and the quantization of this is provided by the quantum group coordinate algebras $\mathbb{C}_q[G]$ (see Hopf Algebras and q -Deformation Quantum Groups and Classical r -matrices, Lie Bialgebras, and Poisson Lie Groups). The bicrossproduct quantum groups are nevertheless unrelated to the latter even though they spring from related classical data.

As already discussed, one interpretation here is of quantized particles in G^* moving on orbits under G and in vice versa in the dual model. The dual model is equivalent in the sense that the states of one (in the sense of positive-linear functionals) lie in the algebra of observables of the other and we also saw in the Planck-scale example inversion of structure constants reminiscent of T -duality in string theory. Motivated in part by this duality Klimcik (1996) along with Severa in the mid 1990s showed that indeed a σ -model on G could be constructed in such a way that there was a matching dual σ -model on G^* in some sense equivalent in terms of solutions to the equations of motion. The Lagrangians here have the usual form

$$\begin{aligned} \mathcal{L} &= E_u(u^{-1}\partial_+u, u^{-1}\partial_-u), \\ \hat{\mathcal{L}} &= \hat{E}_s(s^{-1}\partial_+s, s^{-1}\partial_-s) \end{aligned}$$

where $u: \mathbb{R}^{1,1} \rightarrow G$ and $s: \mathbb{R}^{1,1} \rightarrow G^*$ are the dynamical fields, except that the inner products E, \hat{E} are not constant. Rather they are obtained by solving nonlinear differential equations on the groups defined through the structure constants of $\mathfrak{g}, \mathfrak{g}^*$ and the Drinfeld double $D(\mathfrak{g})$. At the time, T -duality here was well understood in the case of abelian groups while these Poisson–Lie T -duality models provided the first convincing nonabelian models.

This construction was extended by Beggs and Majid (2001) to a general matched pair (G, M) , that is, a σ -model on G dual to one on M . The Poisson–Lie case is the special case where the actions are coadjoint actions and the Lie algebra of $G \bowtie M$ is $D(\mathfrak{g})$. The solutions of the equations of motion for the two systems are created “equally and oppositely” from one on the factorizing group. It could be expected that T -duality ideas again play a role in Planck-scale physics.

Other Bicrossproducts

There are also infinite-dimensional factorizations such as the Riemann–Hilbert problem (see Riemann–Hilbert Problem) in the theory of integrable systems and hence infinite-dimensional matched pairs and bicrossproducts linked to

them. Here we mention just one partly infinite example of current interest.

Thus, the diffeomorphisms on the line \mathbb{R} may be factorized into transformations of the form $ax + b$ and diffeomorphisms that fix the origin and have unit differential there. After a (logarithmic) change of generators to arrive at an algebraic picture, one has a bicrossproduct

$$H(1) = U(b_+) \bowtie H_\infty$$

where b_+ is now the two-dimensional (2D) Lie algebra with relations $[x, y] = x$ and H_∞ is the algebra of polynomials in generators δ_n and a certain coalgebra as a model of the coordinate algebra of the group of diffeomorphisms that fix the origin with unit differential. The Hopf algebra $H(1)$ was introduced by [Connes and Moscovici \(1998\)](#) although not actually as a bicrossproduct (but motivated by the bicrossproduct theory) as part of a family $H(n)$ useful in cyclic cohomology computations. It has cross relations and coproduct determined by

$$\begin{aligned} [\delta_n, x] &= \delta_{n+1}, & [\delta_n, y] &= n\delta_n, \\ \Delta\delta_1 &= \delta_1 \otimes 1 + 1 \otimes \delta_1 \\ \Delta x &= x \otimes 1 + 1 \otimes x + \delta_1 \otimes y, \\ \Delta y &= y \otimes 1 + 1 \otimes y \end{aligned}$$

which we see has a semidirect product form where $\delta_n \triangleleft x = \delta_{n+1}$, $\delta_n \triangleleft y = n\delta_n$. The coalgebra is also a semidirect coproduct by means of a back-reaction of H_∞ in B_+ (expressed as a coaction). From the bicrossproduct theory, we also have a dual model

$$C[B_+] \bowtie U(\text{diff}_0)$$

where diff_0 is the Lie algebra of the group of diffeomorphisms fixing the origin. As such it could be viewed as in the family of examples in the section “Bicrossproduct Poincaré quantum groups” but now with a 2D B_+ . We also conclude from the bicrossproduct theory that this acts covariantly on $R_\lambda^2 = U(b_+)$ after introducing the scaling parameter λ .

Finally, the Hopf algebra $H(1)$ is also part of a family of bicrossproduct Hopf algebras built on rooted trees and related to bookkeeping of overlapping divergences in renormalizable quantum field theories (see Hopf Algebra Structure of Renormalizable Quantum Field Theory). While we have not had room to cover all bicrossproduct quantum groups of interest, it would appear that bicrossproducts are indeed intimately tied up with actual quantum physics.

See also: Classical r -Matrices, Lie Bialgebras, and Poisson Lie Groups; Hopf Algebra Structure of Renormalizable Quantum Field Theory; Hopf Algebras and q -Deformation Quantum Field Groups; Quantum Group Differentials, Bundles and Gauge Theory; Riemann–Hilbert Problem; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

Further Reading

- Amelino-Camelia G and Majid S (2000) Waves on noncommutative spacetime and gamma-ray bursts. *International Journal of Modern Physics A* 15: 4301–4323.
- Beggs E and Majid S (2001) Poisson–Lie T -duality for quasi-triangular Lie bialgebras. *Communications in Mathematical Physics* 220: 455–488.
- Connes A and Moscovici H (1998) Hopf algebras, cyclic cohomology and the transverse index theory. *Communications in Mathematical Physics* 198: 199–246.
- Kac GI and Paljutkin VG (1966) Finite ring groups. *Transactions of the American Mathematical Society* 15: 251–294.
- Kempf A, Mangano G, and Mann RB (1995) Hilbert space representation of the minimal length uncertainty relation. *Physical Review D* 52: 1108–1118.
- Klimcik C (1996) Poisson–Lie T -duality. *Nuclear Physics B (Proc. Suppl.)* 46: 116–121.
- Lukierski J, Nowicki A, Ruegg H, and Tolstoy VN (1991) q -Deformation of Poincaré algebra. *Physics Letters B* 268: 331–338.
- Majid S (1988) Hopf algebras for physics at the Planck scale. *Journal of Classical and Quantum Gravity* 5: 1587–1606.
- Majid S (1990) Physics for algebraists: non-commutative and non-cocommutative Hopf algebras by a bicrossproduct construction. *Journal of Algebra* 130: 17–64.
- Majid S (1990) Matched pairs of Lie groups associated to solutions of the Yang–Baxter equations. *Pacific Journal of Mathematics* 141: 311–332.
- Majid S (1990) On q -regularization. *International Journal of Modern Physics A* 5: 4689–4696.
- Majid S (1995) *Foundations of Quantum Group Theory*. Cambridge: Cambridge University Press.
- Majid S (2000) Meaning of noncommutative geometry and the Planck-scale quantum group. *Springer Lecture Notes in Physics* 541: 227–276.
- Majid S and Ruegg H (1994) Bicrossproduct structure of the κ -Poincaré group and non-commutative geometry. *Physics Letters B* 334: 348–354.
- Oeckl R (2000) Untwisting noncommutative R^d and the equivalence of quantum field theories. *Nuclear Physics B* 581: 559–574.
- Seiberg N and Witten E (1999) String theory and noncommutative geometry. *Journal of High Energy Physics* 9909: 032.
- Snyder HS (1947) Quantized space-time. *Physical Review D* 67: 38–41.
- Takeuchi M (1981) Matched pairs of groups and bismash products of Hopf algebras. *Communications in Algebra* 9: 841.

Bifurcation Theory

M Haragus, Université de Franche-Comté, Besançon, France

G Iooss, Institut Non Linéaire de Nice, Valbonne, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Consider the following equation:

$$F(X, \mu) = 0 \tag{1}$$

where X is the variable, μ is a parameter, and X, μ, F belong to appropriate (finite- or infinite-dimensional) spaces. The problem of *bifurcation theory* is to describe the singularities of the set of solutions

$$S_\mu = \{X; (X, \mu) \text{ satisfies } F(X, \mu) = 0\}$$

The word “bifurcation” was introduced by H Poincaré (1885) in his study of equilibria of rotating liquid masses.

The simplest example is the study of the real roots x of a quadratic polynomial

$$x^2 + bx + c = 0 \tag{2}$$

where μ is represented by the pair of parameters $(b, c) \in \mathbb{R}^2$. As it is well known, real roots are determined by the sign of

$$\Delta \stackrel{\text{def}}{=} b^2 - 4c$$

For $\Delta < 0$, there is no real solution of [2], while there are two solutions x_\pm in the region $\Delta > 0$, which merge when the distance between the point (b, c) and the parabola $\Delta = 0$ tends towards 0. It is then clear that a *singularity occurs in the structure of the set of solutions of [2] at the crossing of the parabola $\Delta = 0$* or, in other words, a *bifurcation occurs in the parameter space (b, c) on the parabola $\Delta = 0$* . A point $(\mu_0, x_0) \in \mathbb{R}^3$ is then called a *bifurcation point* if $\mu_0 = (b, c)$ satisfies $\Delta = 0$, and $x_0 = -b/2$.

In the theory of differential equations, $F(X, \mu)$ often represents a vector field. This study is then concerned with the existence of equilibrium solutions to the differential equation

$$\frac{dX}{dt} = F(X, \mu) \tag{3}$$

and is therefore referred to as *static bifurcation theory*. In addition, *dynamic bifurcation theory* is concerned here with “changes” in the dynamic properties of the solutions of the differential

equation as μ varies. A widely used way to characterize these “changes” is to say that the vector field $F(\cdot, \mu_0)$ is *structurally stable* if the sets of orbits of the differential equation are homeomorphic for μ close to μ_0 , with homeomorphisms which preserve the orientation of the orbits in time t . Then a bifurcation occurs at $\mu = \mu_0$ if $F(\cdot, \mu_0)$ is not structurally stable. It turns out that there is a close link between the stability properties of equilibrium solutions of the differential equation and the type of the bifurcation in static theory.

The tools developed in bifurcation theory are extensively used to solve concrete problems arising in physics and natural sciences. These problems may be modeled by ordinary or partial differential equations, integral equations, but also delay equations or iteration maps, and in all these cases the presence of parameters naturally leads to bifurcation phenomena. They can be regarded as problems of the form [1] or [3], in suitable function spaces, and bifurcation theory allows to detect solutions and to describe their qualitative properties. During the last decades, a class of problems in which the use of bifurcation theory led to significant progress is concerned with nonlinear waves in partial differential equations, including hydrodynamic problems, nonlinear water waves, elasticity, but also pattern formation, front propagation, or spiral waves in reaction–diffusion type systems.

Examples in One and Two Dimensions

The most complete results in bifurcation theory are available in one and two dimensions. The study of static bifurcations in one dimension is concerned with scalar equations

$$f(x, \mu) = 0 \tag{4}$$

where $x \in \mathbb{R}, \mu \in \mathbb{R}$, and the function f is supposed to be regular enough with respect to (x, μ) . When $f(x_0, \mu_0) = 0$ and the derivative of f with respect to x satisfies $\partial_x f(x_0, \mu_0) \neq 0$, the implicit function theorem gives a unique branch of solutions $x(\mu)$ for μ close to μ_0 , and shows the absence of bifurcation points near (μ_0, x_0) . Bifurcation theory intervenes when

$$\partial_x f(x_0, \mu_0) = 0 \tag{5}$$

and one cannot apply the implicit function theorem for solving with respect to x near x_0 . A complete description of the set of solutions near (x_0, μ_0) can be obtained by looking at the partial derivatives of f with respect to x and μ .

For example, if

$$\partial_\mu f(x_0, \mu_0) \neq 0,$$

it is possible to solve with respect to μ and obtain a regular solution $\mu(x)$ such that $\mu(x_0) = \mu_0$ and $f(x, \mu(x)) \equiv 0$. In addition, if the second order derivative

$$\partial_x^2 f(x_0, \mu_0) \neq 0$$

the picture of the solution set in the plane (μ, x) , also called *bifurcation diagram*, shows a turning point with a fold opened to the left or to the right depending upon the sign of the product $\partial_\mu f(x_0, \mu_0) \cdot \partial_x^2 f(x_0, \mu_0)$; see **Figure 1**. Notice that here the bifurcation point $(\mu_0, x_0) \in \mathbb{R}^2$ corresponds to the appearance of a pair of solutions of [4] “from nowhere”. This is the simplest example of a *one-sided bifurcation* in which the bifurcating solutions exist for either $\mu > \mu_0$ or $\mu < \mu_0$.

A particularly interesting situation arises when the equation possesses a symmetry. For example, assume that in [4] the function f is odd with respect to x . This implies that we always have the solution $x = 0$, for any value of the parameter μ . Assume now that f satisfies

$$\partial_x f(0, \mu_0) = 0 \tag{6}$$

and that

$$\partial_{x\mu}^2 f(0, \mu_0) \neq 0, \quad \partial_x^3 f(0, \mu_0) \neq 0 \tag{7}$$

Then the point $(\mu_0, 0)$ is a *pitchfork bifurcation point*, this denomination being related with the bifurcation diagram in the plane (μ, x) ; see **Figure 2**. Notice that here, the bifurcation point $(\mu_0, x_0) \in \mathbb{R}^2$ corresponds to the bifurcation from the origin of a pair of solutions exchanged by the symmetry $x \rightarrow -x$, in addition to the persistent “trivial” solution $x = 0$ which is invariant under the above symmetry. Such a bifurcation is also referred to as a *symmetry-breaking bifurcation*. Similar bifurcation diagrams are found when the equation [4] has a “known” branch of

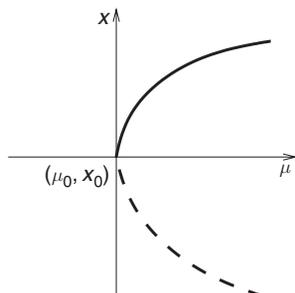


Figure 1 Turning point bifurcation in the case $\partial_\mu f(x_0, \mu_0) > 0$ and $\partial_x^2 f(x_0, \mu_0) < 0$. The solid (dashed) line indicates the branch of stable (unstable) solutions in the differential equation.

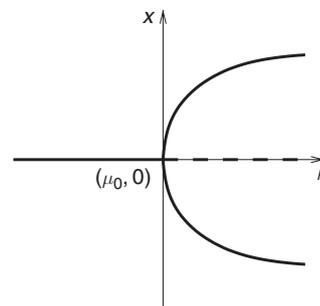


Figure 2 Supercritical pitchfork bifurcation in the case $\partial_{x\mu}^2 f(0, \mu_0) > 0$ and $\partial_x^3 f(0, \mu_0) < 0$. The solid (dashed) lines indicate the branch of stable (unstable) solutions in the differential equation.

solutions $x(\mu)$ for μ close to μ_0 . This situation arises often in applications where usually this branch consists of trivial solutions $x(\mu) = 0$. Then at a bifurcation point (μ_0, x_0) a second branch of solutions appears forming either a one-sided bifurcation, or a two-sided bifurcation; see **Figure 3**.

We can now view f as a vector field in the ordinary differential equation

$$\frac{dx}{dt} = f(x, \mu) \tag{8}$$

and the study above corresponds to looking for equilibrium solutions of [8]. The stability of such a solution is determined by the sign of the derivative $\partial_x f(x, \mu)$ of f at this equilibrium, and it is closely related to the type of the static bifurcation.

In the case of a *turning point bifurcation*, when $\partial_x^2 f(x_0, \mu_0) \neq 0$, the sign of $\partial_x f(x, \mu)$ is different for the two bifurcating solutions. This means that one solution is attracting (i.e., stable), the other one being repelling (i.e., unstable); see **Figure 1**. In the case of a *pitchfork bifurcation* as above, the stability of the trivial solution $x = 0$ changes when μ crosses μ_0 , and the stability of both bifurcating nonzero solutions is the opposite from the stability of the origin on the side of the bifurcation. The bifurcation

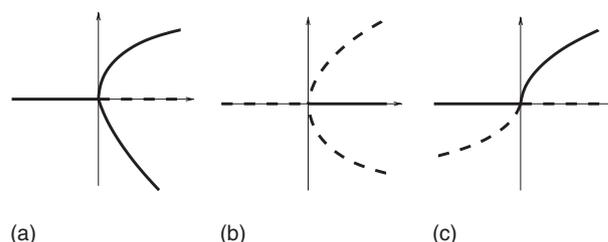


Figure 3 Typical bifurcation diagrams in the case of a branch of trivial solutions. One-sided bifurcations: (a) supercritical, (b) subcritical; two-sided bifurcation: (c) transcritical. The solid (dashed) lines indicate the branch of stable (unstable) solutions in the differential equation.

is called *supercritical* if the bifurcating solutions lie on the side of the bifurcation point where the basic solution $x = 0$ is unstable and *subcritical* otherwise; see Figure 2. The situation is the same in the case of one-sided bifurcations for an equation which has a “known” branch of solutions. In the case of a two-sided bifurcation, there is an *exchange of stability* at the bifurcation point (μ_0, x_0) , solutions on the two branches having opposite stability for $\mu > \mu_0$ and $\mu < \mu_0$, which changes at (μ_0, x_0) . Such a bifurcation is also referred to as *transcritical*; see Figure 3.

Notice that the study of fixed points or periodic points for maps enter in the above frame. Specifically, the period-doubling process occurring in successive bifurcations of one-dimensional maps is a common phenomenon in physics.

The analysis of bifurcations in two dimensions leads to more complicated scenarios. Consider the differential equation [8] in which now $x \in \mathbb{R}^2$ and $f(x, \mu) \in \mathbb{R}^2$, and assume that $f(x_0, \mu_0) = 0$. The behavior of solutions near (x_0, μ_0) is determined by the differential $D_x f(x_0, \mu_0) =: L$ of f with respect to x , which can be identified with a 2×2 matrix. For steady solutions, the implicit function theorem insures the existence of a unique branch of solutions $x(\mu)$ provided L is invertible or, in other words, zero does not belong to the spectrum of L . Consequently, the study of bifurcations of steady solutions is concerned with the case when zero belongs to the spectrum of L , and can be performed following the strategy described for one dimension, provided that the zero eigenvalue of L is simple. For example, assuming that the second eigenvalue is negative leads in general to a *saddle–node bifurcation*, where an additional dimension is added to the previous picture of a turning point bifurcation, in which one of the two bifurcating steady solutions is a stable node, while the other one is a saddle. If, in addition, there is a symmetry S commuting with f , that is, such that $f(Sx, \mu) = Sf(x, \mu)$, and if, for example, x_0 is invariant under S , $Sx_0 = x_0$, and the eigenvector ζ_0 associated to the zero eigenvalue of L is antisymmetric, $L\zeta_0 = -\zeta_0$, then there is again a *pitchfork bifurcation*. The equation possesses a branch of symmetric steady solutions the stability of which changes when crossing the value μ_0 of the parameter, node on one side and saddle on the other, and a pair of solutions is created in a one-sided bifurcation which are exchanged by the symmetry S and have stability opposite to the one of the symmetric solution, just as in the one-dimensional pitchfork bifurcation above.

A new type of bifurcation that arises for vector fields in two dimensions is the so-called *Hopf bifurcation*. This bifurcation was first understood

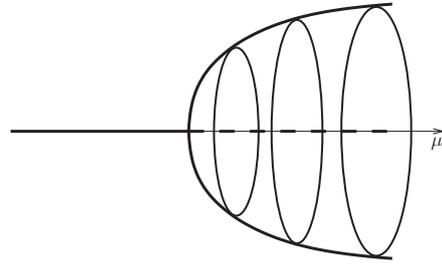


Figure 4 Supercritical Hopf bifurcation.

by Poincaré, and then proved in two dimensions by Andronov (1937) using a Poincaré map, and later in n dimensions by Hopf (1948) by means of a Liapunov–Schmidt-type method. For the differential equation, the absence of the zero eigenvalue in the spectrum of L is not enough to ensure that the vector field $f(\cdot, \mu_0)$ is structurally stable in a neighborhood of x_0 . This only holds when the spectrum of L does not contain purely imaginary eigenvalues, as asserted by the Hartman–Grobman theorem. We are then left with the case when L has a pair of purely imaginary eigenvalues $\pm i\omega$, $\omega \in \mathbb{R}^*$. Static bifurcation theory gives that the system has a unique branch of equilibria $(x(\mu), \mu)$ for μ close to μ_0 , and typically their stability changes as μ crosses μ_0 . For the differential equation a *Hopf bifurcation* occurs in which a branch of periodic orbits bifurcates on one side of μ_0 , and their stability is opposite to that of the steady solution on this side; see Figure 4. A convenient way to study this bifurcation is through “normal form theory,” which is briefly described below.

Local Bifurcation Theory

There are two aspects of bifurcation theory, *local* and *global* theory. As this designation suggests, local theory is concerned with (local) properties of the set of solutions in a neighborhood of a “known” solution, while global theory investigates solutions in the entire space.

An important class of tools in *local* bifurcation theory consists of *reduction methods*, among which the *Liapunov–Schmidt reduction* and the *center manifold reduction* are often used to investigate static and dynamic bifurcations, respectively. The basic idea is to replace the bifurcation problem by an equivalent problem in lower dimensions, for example, a one- or a two-dimensional problem as the ones above.

Consider again the equation [1] in which $F: \mathcal{X} \times \mathcal{M} \rightarrow \mathcal{Y}$ is sufficiently regular, and \mathcal{X}, \mathcal{Y} , and \mathcal{M} are Banach spaces. Assume, without loss of generality,

that $F(0, 0) = 0$, or, in other words, that one solution is known. The equation can be then written as

$$LX + G(X, \mu) = 0$$

in which $L = D_X F(0, 0)$ represents the differential of F with respect to X at $(0, 0)$, and is assumed to have a closed range. The implicit function theorem shows absence of bifurcation if L has a bounded inverse, so that bifurcations are related to the existence of a nontrivial kernel of L . The *Liapunov–Schmidt reduction* then goes as follows.

Let $N(L)$ and $R(L)$ denote the kernel and the range of L , respectively, and consider continuous projections $P: \mathcal{X} \rightarrow N(L)$ and $Q: \mathcal{Y} \rightarrow R(L)$. Then there exists a bounded linear operator $B: R(L) \rightarrow (\text{id} - P)\mathcal{X}$, the right inverse of L , satisfying $LB = \text{id}$ on $R(L)$ and $BL = \text{id} - P$ on \mathcal{X} . For $X \in \mathcal{X}$ one may write

$$X = X_0 + X_1, \quad X_0 = PX, X_1 = (\text{id} - P)X$$

and then by projecting with $\text{id} - Q$ and Q the equation becomes

$$\begin{aligned} (\text{id} - Q)G(X_0 + X_1, \mu) &= 0 \\ X_1 + BQG(X_0 + X_1, \mu) &= 0 \end{aligned}$$

The implicit function theorem allows to solve the second equation for $X_1 = \psi(X_0, \mu)$ in a neighborhood of the origin. Substitution into the first equation leads to the equation in $(\text{id} - Q)\mathcal{Y}$ for X_0 in $P\mathcal{X}$,

$$(\text{id} - Q)G(X_0 + \psi(X_0, \mu), \mu) = 0$$

also called *bifurcation equation*. This equation completely describes the set of solutions to [1] in a neighborhood of $(0, 0)$, and this problem is then posed in a space of dimension much smaller than the dimension of \mathcal{X} .

The basic principle of the Liapunov–Schmidt method has been discovered and used independently by different authors. E Schmidt (1908) used this method for integral equations, while Liapunov used it to study the stability of the zero solution of nonlinear partial differential equations when the linear part has zero eigenvalues (1947), and later in 1960 for the bifurcation problem studied by Poincaré (1885). In working in a Banach space of t -periodic functions, the Liapunov–Schmidt method may be used to solve the Hopf bifurcation problem, as did Hopf himself in 1948.

The analog of this reduction procedure for the differential equation [3] is the *center manifold reduction*. Assuming that $F(0, 0) = 0$, we obtain the differential equation

$$\frac{dX}{dt} = LX + G(X, \mu)$$

Since dynamic bifurcations are related to the existence of purely imaginary spectral values of L , the kernel of L alone is not enough to describe this situation. One has to consider the spectral space \mathcal{Y}_c of L associated to the purely imaginary spectrum of L . A spectral gap is needed between this part of the spectrum and the rest (always true in finite dimensions), so that the spectral projection P onto \mathcal{Y}_c is well defined. One writes

$$X = X_c + X_b, \quad X_c = PX, X_b = (\text{id} - P)X$$

and obtains the decomposed system

$$\begin{aligned} \frac{dX_c}{dt} &= LX_c + PG(X_c + X_b, \mu) \\ \frac{dX_b}{dt} &= LX_b + (\text{id} - P)G(X_c + X_b, \mu) \end{aligned}$$

The reduction procedure works provided the non-homogeneous linear equation

$$\frac{dX_b}{dt} = LX_b + f(t)$$

possesses a unique solution in suitably chosen function spaces with weak exponential growth, such that one can then solve the second equation for $X_b = \Psi(X_c)$ in a neighborhood of the origin in these function spaces. This property is always true in finite dimensions, but it has to be checked in infinite dimensions. Different results showing the solvability of this equation are available in both Banach and Hilbert spaces, relying upon additional conditions on the spectrum of L , decaying properties of the resolvent of L on the imaginary axis, and regularity properties of the nonlinearity G . The map Ψ is then used to construct a map $\psi: P\mathcal{X} \times \mathcal{M} \rightarrow (\text{id} - P)\mathcal{X}$, defined in a neighborhood of the origin, which parametrizes a *local center manifold* invariant under the flow of the equation. The flow on this center manifold is governed by the *reduced equation* in \mathcal{Y}_c ,

$$\frac{dX_c}{dt} = LX_c + PG(X_c + \psi(X_c, \mu), \mu)$$

which completely describes the bifurcation problem.

The first proofs of this result were given in finite dimensions by Pliss (1964) and Kelley (1967). Center manifolds in infinite dimensions have been studied in different settings determined by assumptions on the linear part L and the nonlinear part G . One typical assumption in infinite dimensions is that the spectrum of L contains only a finite number of purely imaginary eigenvalues, so that the reduced equation above is a differential equation in a finite-dimensional space.

These reduction methods work for a large class of problems and the advantage of such an approach is that one is left with a bifurcation problem in a lower-dimensional space. The methods involved in

solving this reduced bifurcation problem can be very different from one problem to another, and often make use of some additional structure in the problem, such as a gradient-like structure, Hamiltonian structure, or the presence of symmetries, which are preserved by the reduction procedure.

A powerful tool for the analysis of these reduced differential equations is provided by the *normal form theory*, which goes back to works of Poincaré (1885) and Birkhoff (1927). The idea is to use coordinate transformations to make the expression of the vector field as simple as possible. The transformed vector field is called *normal form*. There is an extensive literature on normal forms for vector fields in many different contexts, in both finite- and infinite-dimensional cases. Typically the classes of normal forms are characterized in terms of the linear part of the differential equation.

For differential equations of the form

$$\frac{dx}{dt} = Lx + g(x, \mu) \tag{9}$$

in which L is a matrix and g a sufficiently regular map such that $g(0, 0) = 0, D_x g(0, 0) = 0$, as encountered in bifurcation theory, one possible characterization of normal forms makes use of the adjoint matrix L^* . Fixing any order $k \geq 2$, there exist polynomials Φ and N of degree k in x with coefficients which are regular functions of μ , and $\Phi(0, 0) = N(0, 0) = 0, D_x \Phi(0, 0) = D_x N(0, 0) = 0$, such that by the change of variables

$$x = y + \Phi(y, \mu)$$

the equation [9] is transformed into the normal form

$$\frac{dy}{dt} = Ly + N(y, \mu) + o(\|y\|^k) \tag{10}$$

in which the polynomial N is characterized through

$$N(e^{tL^*} y, \mu) = e^{tL^*} N(y, \mu)$$

for all y, μ , and t , or, equivalently,

$$D_y N(y, \mu) L^* y = L^* N(y, \mu)$$

for all y and μ . This characterization allows to determine the classes of possible normal forms for a given matrix L , and also provides an efficient way to compute the normal form for a given vector field g . As for the reduction methods, normal form transformations can be made to preserve the additional structure of the problem, such as Hamiltonian structure or symmetries.

As an example, consider a differential equation of the form [9] with $x \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$, which supports a Hopf bifurcation so that L has simple eigenvalues $\pm i\omega, \omega > 0$, and no other eigenvalues with zero real

part. The center manifold reduction provides a two-dimensional reduced system with linear part having the simple eigenvalues $\pm i\omega$, for which it is convenient to write the normal form in complex variables

$$\frac{dA}{dt} = i\omega A + A\mathcal{Q}(|A|^2, \mu) + o(|A|^{2k+2})$$

for $A(t) \in \mathbb{C}$, where \mathcal{Q} is a complex polynomial of degree k in $|A|^2$ with $\mathcal{Q}(0, 0) = 0$, or, equivalently, in polar coordinates $A = re^{i\phi}$,

$$\begin{aligned} \frac{dr}{dt} &= rQ_r(r^2, \mu) + o(r^{2k+2}) \\ \frac{d\phi}{dt} &= \omega + Q_\phi(r^2, \mu) + o(r^{2k+1}) \end{aligned}$$

Q_r and Q_ϕ being the real and imaginary part of \mathcal{Q} , respectively. The radial equation for r truncated at order $2k + 1$ decouples and admits a pitchfork bifurcation. The bifurcating steady solutions of this equation then lead first to periodic solutions for the truncated system, which are then shown to persist for the full equation by a standard perturbation analysis.

A situation that occurs in a large class of problems is when the problem possesses a reversibility symmetry, which often comes from some reflection invariance in the physical space, that is, when the vector field $F(\cdot, \mu)$ anticommutes with a symmetry operator S . One of the simplest examples is the case of a differential equation [9] when the matrix L has a double eigenvalue in 0, no other eigenvalues with zero real part, and a one-dimensional kernel which is invariant by S . In this case, the center manifold reduction provides a two-dimensional reduced reversible system, which can be put in the normal form

$$\begin{aligned} \frac{da}{dt} &= b \\ \frac{db}{dt} &= \mu - a^2 + o((|a| + |b|)^3) \end{aligned}$$

which anticommutes with the symmetry $(a, b) \mapsto (a, -b)$. The above system undergoes a *reversible Takens–Bogdanov bifurcation* and has for $\mu > 0$ a phase portrait as in **Figure 5**. There are two equilibria, one a saddle, the other a center, and a family of periodic orbits with the zero-amplitude limit at the center equilibrium, and the infinite-period limit a homoclinic orbit, originating at the saddle point. In concrete problems the bounded orbits of such a reduced system determine the shape of physically interesting solutions of the full system of equations, such as, for example, in water-wave theory where to homoclinic and periodic orbits correspond solitary and periodic waves, respectively.

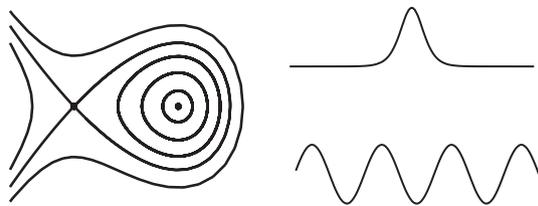


Figure 5 Phase portrait of the reduced system in a reversible Takens–Bogdanov bifurcation (left) and sketch of the a -component of solutions corresponding to homoclinic and periodic orbits (right).

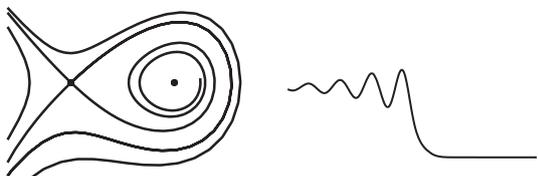


Figure 6 Phase portrait of the reduced system in absence of reversibility (left) and sketch of the a -component of the solution corresponding to the bounded orbit (right).

Notice that in the absence of the reversibility symmetry, the same type of bifurcation may lead to a completely different phase portrait for the reduced system as, for example, the one in [Figure 6](#) in which the homoclinic and the periodic orbits disappear. This situation often occurs in the presence of a small dissipation in nearly reversible systems.

Global Bifurcation Theory

Most of the existing results in global bifurcation theory concern the static problem [1]. The analysis of *global sets* of solutions often relies upon topological methods, degree theory, but also variational methods, or analytic function theory. Significant progress in understanding global branches of solutions has been made in the 1970s, in particular, for nonlinear eigenvalue problems and the Hopf bifurcation problem (see, e.g., works by Rabinowitz, Crandall, Dancer, and Alexander, Yorke, Ize, respectively).

A now-classical result in the topological theory of global bifurcations is the following theorem by Rabinowitz (1970), which gives a characterization of global sets of solutions for *eigenvalue problems* of the form

$$X = F(X, \mu) = \mu LX + H(X, \mu)$$

$H(X, \mu) = o(\|X\|)$, posed for $(X, \mu) \in \mathcal{X} \times \mathbb{R}$, \mathcal{X} being a Banach space. In contrast to local theory where the function F is usually k -times differentiable (with a suitable k), in the global theory a typical assumption is that $F: \mathcal{X} \times \mathbb{R} \rightarrow \mathcal{X}$ is *compact*. The equation above possesses a “trivial” branch of

solutions $(0, \mu)$ for any μ . The bifurcation result asserts that if for some real parameter value μ_0 zero is an eigenvalue of odd multiplicity of the operator $\text{id} - \mu_0 L$, then the set S of nontrivial solutions (X, μ) possesses a maximal subcontinuum which contains $(0, \mu_0)$ and meets either infinity in $\mathcal{X} \times \mathbb{R}$ or another trivial solution $(0, \mu_1)$, $\mu_1 \neq \mu_0$. In particular, $(\mu_0, 0)$ is a bifurcation point. A local version of this result is often referred to as Krasnoselski’s theorem.

Different versions and extensions of these theorems can be found in the literature, as, for example, in the case of a simple eigenvalue, or if the field F is real-analytic when the set of solutions is path-connected. More recent works address the question of lack of compactness, and a number of results are now available for problems with additional structure (gradient-like or Hamiltonian structure), but also for concrete problems, such as the water-wave problem.

See also: Bifurcations in Fluid Dynamics; Bifurcations of Periodic Orbits; Central Manifolds, Normal Forms; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Ginzburg–Landau Equation; Integrable Systems: Overview; Leray–Schauder Theory and Mapping Degree; Singularity and Bifurcation Theory; Stability Theory and KAM; Symmetry and Symmetry Breaking in Dynamical Systems.

Further Reading

- Arnold VI (1988) Geometrical Methods in the Theory of Ordinary Differential Equations. *Grundlehren der Mathematischen Wissenschaften*, vol. 250. New York: Springer.
- Buffoni B and Toland J (2003) *Analytic Theory of Global Bifurcation*. Princeton: Princeton University Press.
- Chossat P and Lauterbach R (2000) Methods in Equivariant Bifurcations and Dynamical Systems. *Advanced Series in Nonlinear Dynamics*, vol. 15. River Edge, NJ: World Scientific.
- Chow S-N and Hale JK (1982) Methods of Bifurcation Theory. *Grundlehren der Mathematischen Wissenschaften*, vol. 251. New York: Springer.
- Golubitsky M and Schaeffer DG (1985) Singularities and Groups in Bifurcation Theory, Vol. I. *Applied Mathematical Sciences*, vol. 51. New York: Springer.
- Golubitsky M, Stewart I, and Schaeffer DG (1988) Singularities and Groups in Bifurcation Theory, Vol. II. *Applied Mathematical Sciences*, vol. 69. New York: Springer.
- Guckenheimer J and Holmes P (1990) Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. *Applied Mathematical Sciences*, vol. 42. New York: Springer.
- Iooss G and Adelmeyer M (1998) Topics in Bifurcation Theory and Applications, *Advances Series in Nonlinear Dynamics*, 2nd edn., vol. 3, Singapore: World Scientific.
- Iooss G, Helleman RHG, and Stora R (eds.) (1983) *Chaotic behavior of deterministic systems*. Session XXXVI of the Summer School in Theoretical Physics held at Les Houches June 29–July 31, 1981. Amsterdam: North-Holland.

- Ize J and Vignoli A (2003) *Equivariant Degree Theory. de Gruyter Series in Nonlinear Analysis and Applications*, vol. 8. Berlin: de Gruyter and Co.
- Kielhöfer H (2004) *Bifurcation Theory. An Introduction with Applications to PDEs*, Applied Mathematical Sciences, vol. 156. New York: Springer.
- Kuznetsov YA (2004) *Elements of Applied Bifurcation Theory*, 3rd edn. *Applied Mathematical Sciences*, vol. 112. New York: Springer.

- Ruelle D (1989) *Elements of Differentiable Dynamics and Bifurcation Theory*. Boston MA: Academic Press.
- Vanderbauwhede A (1989) Centre Manifolds, Normal Forms and Elementary Bifurcations. *Dynamics Reported, Dynam. Report. Ser. Dynam. Systems Appl.*, vol. 2, pp. 89–169. Chichester: Wiley.
- Vanderbauwhede A and Iooss G (1992) Center Manifold Theory in Infinite Dimensions. *Dynamics Reported: Expositions in Dynamical Systems*, vol. 1, pp. 125–163. Berlin: Springer.

Bifurcations in Fluid Dynamics

G Schneider, Universität Karlsruhe, Karlsruhe, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Almost all classical hydrodynamical stability problems are experiments or gedankenexperiment which have been designed to understand and to extract special phenomena in more complicated situations. Examples are the Taylor–Couette problem, Bénard’s problem, Poiseuille flow, or Kolmogorov flow.

The Taylor–Couette problem consists in finding the flow of a viscous incompressible fluid contained in between two coaxial co- or counterrotating cylinders, cf. **Figure 1**. If the rotational velocity of the inner cylinder is below a certain threshold, the trivial solution, called the Couette flow, is asymptotically stable. At the threshold, this spatially homogenous solution becomes unstable and bifurcates via a pitchfork bifurcation or a Hopf bifurcation into different spatially periodic patterns, that is, depending on the rotational velocity of the outer cylinder the basic patterns are stationary (called the Taylor vortices) or

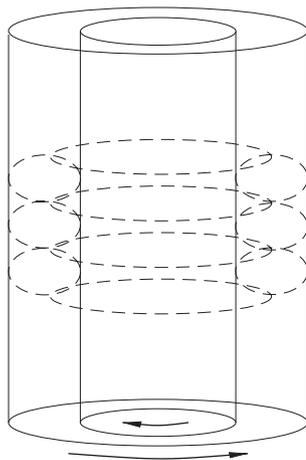


Figure 1 The Taylor–Couette problem with the Taylor vortices.

time-periodic. If the rotational velocity of the inner cylinder is increased further, more complicated patterns occur. The bifurcation scenario is well understood from experiments and analytic investigations.

Bénard’s problem consists in finding the flow of a viscous incompressible fluid contained in between two plates, where the lower plate is heated and the upper plate is kept at a constant temperature, cf. **Figure 2**. If the temperature difference between the two plates is below a certain threshold, the transport of energy from below to above is made by pure conduction. At this threshold, this spatially homogenous solution becomes unstable, convection sets in, and spatially periodic patterns as rolls or hexagons occur. Convection problems play a big role in geophysical applications, that is, in spherical domains, as the earth. The paradigm for an anisotropic pattern-forming system is electroconvection in nematic crystals.

Poiseuille flow consists in finding the flow of a viscous incompressible fluid flowing through a pipe driven by some pressure gradient, cf. **Figure 3**. In noncircular pipes, the trivial laminar flow becomes unstable at a critical pressure gradient. Experimentally, a direct transition to turbulent flow with large amplitudes is observed, according to the fact that in general at the instability point of the trivial solution a subcritical bifurcation occurs.

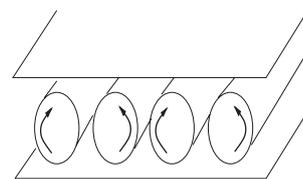


Figure 2 Bénard’s problem with rolls.

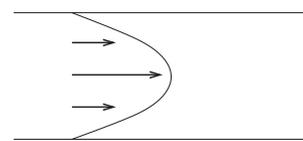


Figure 3 Poiseuille flow with the trivial solution.

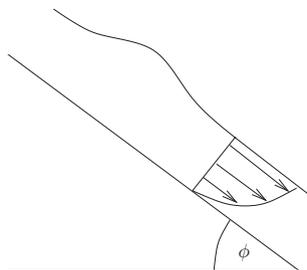


Figure 4 The inclined-plane problem. The trivial Nusselt solution possesses a flat top surface and a parabolic flow profile.

Kolmogorov flow consists in finding the flow of a viscous incompressible fluid under the action of an external force parallel to the flow direction x and varying periodically in the perpendicular y -direction. This gedankenexperiment has been designed by Kolmogorov in 1958 as a simplified model for the Poiseuille flow problem in order to study the nature of turbulence. The trivial solution which is called Kolmogorov flow can become unstable via a long-wave instability along the flow direction.

The inclined-plane problem consists in finding the flow of a viscous liquid running down an inclined plane, cf. **Figure 4**. The trivial solution, the so-called Nusselt solution, becomes sideband-unstable if the inclination angle ϕ is increased. Then the dynamics is dominated by traveling pulse trains, although the individual pulses are unstable due to the long-wave instability of the flat surface. Time series taken from the motion of the individual pulses indicates the occurrence of chaos directly at the onset of instability.

There are other famous hydrodynamical stability problems, with arbitrarily complicated bifurcation scenarios.

Spectral Analysis of the Trivial Solution

All classical hydrodynamical stability problems are described by the Navier–Stokes equations

$$\begin{aligned} \partial_t U &= \frac{1}{\nu} \Delta U - \nabla p - (U \cdot \nabla) U + f \\ 0 &= \nabla \cdot U \end{aligned} \quad [1]$$

where $U = U(x, t) \in \mathbb{R}^d$ with $d = 2, 3$ is the velocity field, $p = p(x, t) \in \mathbb{R}$ the pressure field, f some external forcing, and ν the dynamic viscosity. These equations are completed with boundary conditions. In case of Bénard’s problem, the Navier–Stokes equations are coupled to a nonlinear heat equation.

By projecting U onto the space of divergence-free vector fields and by taking the trivial solution as new origin all problems from the previous section can be written as evolutionary system

$$\partial_t U = \Lambda U + N(U)$$

where $U = 0$ corresponds to the trivial solution, where Λ is a linear and $N(U) = \mathcal{O}(U^2)$ for $U \rightarrow 0$ a nonlinear operator. Most of the examples from the previous section are semilinear, that is, from a functional analytic point of view, the nonlinear operator N can be controlled in terms of the linear operator Λ .

Since the form of the bifurcating pattern is only slightly influenced by far away boundaries, that is, for instance, the upper and lower end of the rotating cylinders in the Taylor–Couette problem, the problems are considered from a theoretical point of view in unbounded domains, $\Omega = \mathbb{R}^d \times \Sigma$, with $\Sigma \subseteq \mathbb{R}^m$ the bounded cross section that is, for instance, that the Taylor–Couette problem is considered with two cylinders of infinite length. Then the eigenfunctions of the linear operator Λ are given by Fourier modes, that is,

$$\Lambda(e^{ik \cdot x} \varphi_{k,n}(z)) = \lambda_n(k) e^{ik \cdot x} \varphi_{k,n}(z)$$

with $x \in \mathbb{R}^d, k \in \mathbb{R}^d, k \cdot x = \sum_{j=1}^d k_j x_j, z \in \Sigma, n \in \mathbb{N}$. If an external control parameter is changed, independent of the underlying physical problem, the trivial solution becomes unstable, then the surface $k \mapsto \text{Re} \lambda_1(k)$ intersects the plane $\{\text{Re} \lambda_1(k) = 0\}$. Generically, this happens first at a nonzero wave vector $k_c \neq 0$ (cf. **Figure 5**).

Examples for such an instability are the Taylor–Couette problem, Bénard’s problem, or Poiseuille flow. Very often, due to some conserved quantity in the problem we have $\text{Re} \lambda_1(0) = 0$ for all values of the bifurcation parameter. Then, a so-called sideband instability can occur, cf. **Figure 6**.

Examples for such an instability are the Kolmogorov flow problem or the inclined plane problem.

According to some symmetries in the problem, for instance, reflection along the cylinders in the Taylor–Couette problem or rotational symmetry in Bénard’s problem, the curves in **Figure 5** are double or rotational symmetric.

In case of Ω being spherical symmetric, we have

$$\Lambda(f_l(r) \varphi_{l,n}(z)) = \lambda_l f_l(r) \varphi_{l,n}(z)$$

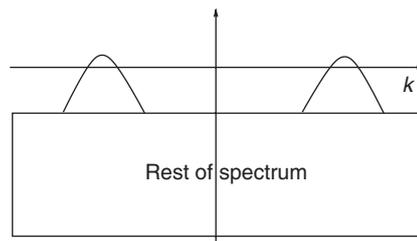


Figure 5 Real part of the spectrum in case of an instability at a wave number $k_c \neq 0$. Definition of the small bifurcation parameter ε .

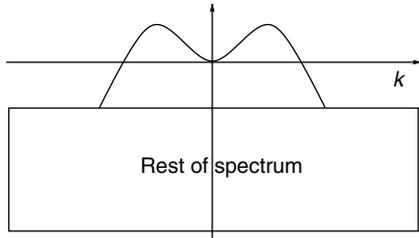


Figure 6 Real part of the spectrum in case of a sideband instability. Definition of the small bifurcation parameter ε .

with $r \geq 0, z \in S^d, \varphi_{l,n}$ for $l \in \mathbb{N}_0$ and $m = -l, l-1, \dots, l+1, l$ being a spherical harmonic, that is, if λ_{l_0} is the eigenvalue having first positive real part, then by symmetry, simultaneously $2l_0 + 1$ eigenvalues cross the imaginary axis.

Reduction of the Dimension

In order to understand the occurrence of the spatially periodic Taylor vortices in the Taylor–Couette problem and of the roll solutions and hexagons in Bénard’s problem, the problems are considered with periodic boundary conditions along the unbounded directions. Then the instability of the trivial solution occurs when at least one eigenvalue crosses the imaginary axis. Generically, this happens by a simple real eigenvalue or a pair of complex-conjugate eigenvalues crossing the imaginary axis (Figure 7). Center manifold theory and the Lyapunov–Schmidt reduction allow to reduce the *a priori* infinite-dimensional bifurcation problem to a finite-dimensional one.

In case of a real eigenvalue λ_1 crossing the imaginary axis, the solution u can be written as a sum of the weakly unstable mode and the stable modes, that is, $u = c_1\varphi_1 + u_r$, ($c_1 \in \mathbb{R}$), where u_r lives in the closure of the span of the stable eigenfunctions $\{\varphi_2, \varphi_3, \dots\}$. For the linearized system all solutions are attracted by the one-dimensional set $E_c = \{u \mid u_r = 0\}$, in which all solutions diverge to infinity.

For the nonlinear system and small bifurcation parameter this attracting structure survives, no longer as a linear space, but as a manifold

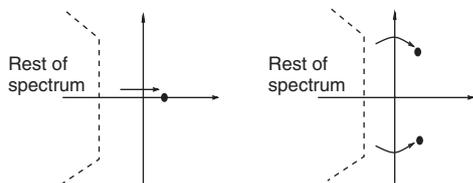


Figure 7 Generically, a simple real eigenvalue or a pair of complex-conjugate eigenvalues cross the imaginary axis.

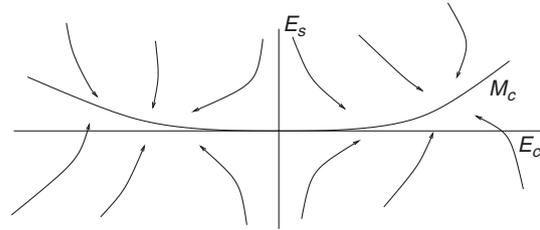


Figure 8 The center manifold is invariant under the flow, is tangential to the central subspace E_c , and attracts nearby solutions with some exponential rate.

$$M_c = \{u = c_1\varphi_1 + b(c_1) \mid b(c_1) \in \overline{\text{span}\{\varphi_2, \varphi_3, \dots\}}\}$$

the so-called center manifold which is tangential to E_c , that is, $\|b(c_1)\| \leq C\|c_1\|^2$ (Figure 8). The dynamics on M_c is no longer trivial due to the nonlinear terms.

Due to the fact that real problems are considered $\text{Re}\lambda_1(k_c) = 0$ implies $\text{Re}\lambda_1(-k_c) = 0$, that is, in case of $2\pi/k_c$ -periodic boundary conditions always two eigenvalues cross the imaginary axis simultaneously. For Bénard’s problem in a strip or for the Taylor–Couette problem in case of a bifurcation of fixed points, the reduced system on the center manifold is derived with the ansatz

$$U = \varepsilon A(\varepsilon^2 t)e^{ik_c x} + \text{c.c.} + \mathcal{O}(\varepsilon^2)$$

where $0 < \varepsilon \ll 1$ is the small bifurcation parameter, cf. Figure 5. Then due to $e^{ik_c x}e^{ik_c x}e^{-ik_c x} = e^{ik_c x}$ the complex-valued amplitude A satisfies the so-called Landau equation

$$\partial_T A = A - \gamma A|A|^2 + \mathcal{O}(\varepsilon^2)$$

where the Landau coefficient $\gamma \in \mathbb{R}$ is obtained by classical perturbation analysis (Figure 9). The reduced system is symmetric under the S^1 -symmetry

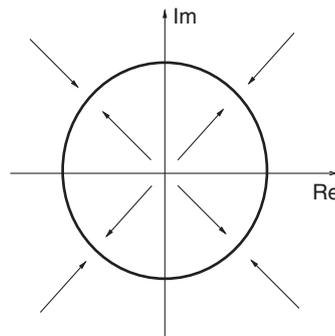


Figure 9 The dynamics of the Landau equation. Except of the origin which corresponds to the Couette flow, all solutions converge towards the circle of fixed points, which corresponds to the family of Taylor vortices. The translation invariance of the Taylor–Couette problem is reflected by the rotational symmetry of the reduced system.

$A \mapsto Ae^{i\phi}$ with $\phi \in \mathbb{R}$ which corresponds to the translation invariance of the original systems.

This so-called equivariant bifurcation theory has been applied successfully to convection problems in the plane and on the sphere.

The stability of time-periodic flows can be analyzed with Floquet multipliers. Bifurcations from a time-periodic solution can lead to quasiperiodic motion in time. Ruelle and Takens (1971) showed that already the next bifurcation leads to chaotic dynamics. Since this time many classical hydrodynamical stability problems have been analyzed with bifurcation theory up to turbulent flows.

It was observed that center manifold theory can also be applied successfully to elliptic PDE problems posed in spatially unbounded cylindrical domains. A famous example is the construction of capillary-gravity solitary waves for the so-called water-wave problem.

Modulation Equations

The analysis of the last section is of no use in case of a sideband instability occurring at the wave number $k_c = 0$, as it happens in the inclined-plane problem or in the Kolmogorov flow problem. Moreover, in case of an instability at a wave vector $k_c \neq 0$, based on the above analysis, front solutions cannot be described. In such situations, the method of modulation equations generalizes the role of the finite-dimensional amplitude equations from the last section.

The complex cubic Ginzburg–Landau equation in normal form is given by

$$\partial_T A = (1 + i\alpha)\partial_X^2 A + A - (1 + i\beta)A|A|^2$$

where the coefficients $\alpha, \beta \in \mathbb{R}$ are real, and we have $X \in \mathbb{R}, T \geq 0$, and $A(X, T) \in \mathbb{C}$. The Ginzburg–Landau equation is a universal amplitude equation that describes slowly varying modulations, in space and time, of the amplitude of bifurcating spatially periodic solutions in pattern-forming systems close to the threshold of the first instability. Whenever the instability drawn in Figure 5 occurs, that is, for the Taylor–Couette problem and Bénard’s problem in a strip, that is, $d = 1$, it can be derived by a multiple scaling ansatz

$$u(x, t) \sim \varepsilon A(\varepsilon(x - c_g t), \varepsilon^2 t) e^{i(k_c x - \omega_0 t)} + \text{c.c.}$$

For instance, in case of $\alpha = \beta = 0$, the Ginzburg–Landau equation possesses front solutions connecting the stable fixed point $A = 1$ with the unstable fixed point $A = 0$. Such solutions correspond in the Taylor–Couette problem to modulating fronts

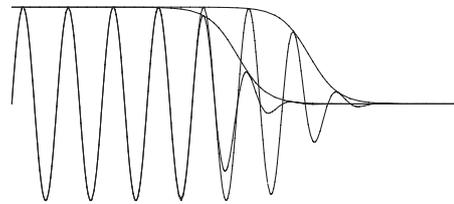


Figure 10 The front solution of the Ginzburg–Landau equation modulates the underlying pattern in the original system.

connecting the stable Taylor vortices with the unstable Couette flow, cf. Figure 10.

The diffusion operator in the Ginzburg–Landau equation reflects the parabolic shape of $\text{Re}\lambda_1$ close to $k = k_c$ in Figure 5. In case of the long-wave instability, as drawn in Figure 6, the second-order differential operator changes in a fourth-order differential operator.

For Kolmogorov flow with $T = \varepsilon^4 t$ and $X = \varepsilon x$ and the amplitude scaled with ε , we obtain that in lowest order A has to satisfy a Cahn–Hilliard equation

$$\partial_T A = -\sqrt{2}\partial_X^2 A - 3\partial_X^4 A + \gamma\partial_X^2(A^3)$$

where $A(X, T) \in \mathbb{R}$ and $\gamma \in \mathbb{R}$ a constant (cf. Figure 6).

The Kuramoto–Shivashinsky (KS)-perturbed KdV equation

$$\partial_T A = -\partial_X^3 u - \partial_X(A^2)/2 - \varepsilon(\partial_X^2 + \partial_X^4)u$$

with $A = A(X, T) \in \mathbb{R}, X \in \mathbb{R}, T \geq 0$, where $0 < \varepsilon \ll 1$ is still a small parameter, can be derived for the inclined problem with $T = \varepsilon^3 t$ and $X = \varepsilon x$ and the amplitude scaled with ε^2 .

The theory of modulation equations is nowadays a well-established mathematical tool which allows us to construct special solutions, global existence results for the solutions of pattern-forming systems, or allows to characterize the attractors in such systems. The method is based on approximation results, showing that solutions of the original systems can be approximated by the modulation equation and attractivity results showing that every solution of the original system develops in such a way that it can be described by the modulation equation.

This method can also be applied to secondary bifurcations describing instabilities of spatially periodic wave trains. Then the so-called phase-diffusion equations, conservation laws, Burgers equations, and again the KS equations occur.

However, this method cannot be applied successfully in all situations. There are counterexamples showing that not every formally derived modulation equation describes the original system in a correct way. Moreover, very often according to some symmetries in the original problem no consistent

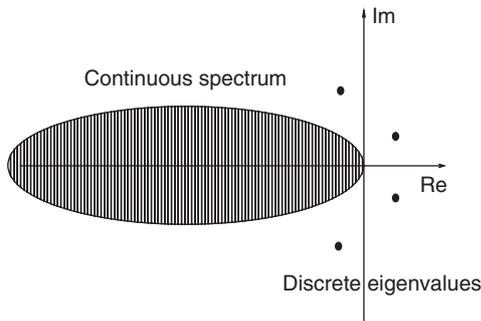


Figure 11 Spectrum for the flow around an obstacle.

multiple scaling analysis is possible, that is, that the modulation equations still depend on ε .

Discussion

There is no satisfactory bifurcation analysis for situations where boundary layers play a role. The most simple problem is the flow around some obstacle. The difficulties are according to the fact that due to the unbounded flow region there is always continuous spectrum up to the imaginary axis. From the localized obstacle discrete eigenvalues are created, (cf. Figure 11).

In such a situation, so far there is no mathematical bifurcation theory available.

See also: Bifurcation Theory; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves;

Leray–Schauder Theory and Mapping Degree; Multiscale Approaches; Newtonian Fluids and Thermohydraulics; Symmetry and Symmetry Breaking in Dynamical Systems; Turbulence Theories; Variational Methods in Turbulence.

Further Reading

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. Oxford: Clarendon.
 Chang H-C and Demekhin EA (2002) *Complex Wave Dynamics on Thin Films*, Studies in Interface Science, vol. 14. Amsterdam: Elsevier.
 Chossat P and Iooss G (1994) *The Taylor–Couette Problem*, Applied Mathematical Sciences, vol. 102. Springer.
 Chow S-N and Hale J (1982) *Methods of Bifurcation Theory*, Grundlehren der Mathematischen Wissenschaften, vol. 251. Berlin: Springer.
 Golubitsky M and Schaeffer DG (1985) *Singularities and Groups in Bifurcation Theory I*, Applied Mathematical Sciences, vol. 51. Berlin: Springer.
 Golubitsky M, Stewart I, and Schaeffer DG (1988) *Singularities and Groups in Bifurcation Theory II*, Applied Mathematical Sciences, vol. 69. Berlin: Springer.
 Haken H (1987) *Advanced Synergetics*. Berlin: Springer.
 Henry D (1981) *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Mathematics, vol. 840. Berlin: Springer.
 Mielke A (2002) The Ginzburg–Landau equation in its role as a modulation equation. In: Fiedler B (ed.) *Handbook of Dynamical Systems II*, pp. 759–834. Amsterdam: North-Holland.
 Ruelle D and Takens F (1971) On the nature of turbulence. *Communications in Mathematical Physics* 20: 167–192.
 Temam R (1988) *Infinite-Dimensional Systems in Mechanics and Physics*. Berlin: Springer.

Bifurcations of Periodic Orbits

J-P François, Université P.-M. Curie, Paris VI, Paris, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Bifurcation theory of periodic orbits relates to modeling of quite diverse subjects. It appeared classically in the field of celestial mechanics with the contributions of H Poincaré. Van der Pol (1926, 1927, 1928, 1931) observed the frequency-locking phenomenon in electrical circuits. More recently, Malkin’s theory (Malkin 1952, 1956, Roseau 1966) was used to justify synchronization of weakly coupled oscillators modeling the electrical activity of the cells of the sinus node in the heart. This article provides the essential mathematical background necessary for existence of frequency locking. Applications can be found, for instance, in Weakly Coupled Oscillators.

The Asymptotic Phase of a Stable Periodic Orbit

Let Γ be a periodic orbit of a vector field and let $S(\Gamma)$ denote the stable manifold of Γ (resp. $U(\Gamma)$ denotes the unstable manifold of Γ). The following theorem can be found, for instance, in Hartman (1964).

Theorem *There exist α and K such that $\text{Re}(\lambda_j) < -\alpha$, $j = 1, \dots, k$ and $\text{Re}(\lambda_j) > \alpha$, $j = k + 1, \dots$, and for all $x \in S(\Gamma)$, there is an asymptotic phase t_0 such that for all $t \geq 0$*

$$|\phi_t(x) - \gamma(t - t_0)| < K e^{-\alpha(t/T)}$$

Similarly, for any $x \in U(\Gamma)$, there is a t_0 such that $t \leq 0$,

$$|\phi_t(x) - \gamma(t - t_0)| < K e^{\alpha(t/T)}$$

If the periodic orbit is stable, the local stable manifold coincides with an open neighborhood of Γ . In such a case, there is a foliation of this open set

whose leaves are the points with a given asymptotic phase. The asymptotic phase can be considered as a coordinate function ϕ defined on the neighborhood $S(\Gamma)$.

If we consider now the particular case of a plane system, this function can be completed with the square of the distance function to the orbit into a coordinate system called the “amplitude–phase” system and denoted as (ρ, ϕ) .

Frequency Locking and Phase Locking

The term “oscillator” has two meanings. A conservative “oscillator” is a plane vector field which displays an open set of periodic orbits. It is said to be isochronous if all orbits have same period. A dissipative “oscillator” is a planar vector field which displays an attractive limit cycle (attractive periodic orbit).

We consider N dissipative oscillators:

$$\begin{aligned} \frac{dx_i}{dt} &= f(x_i, y_i) \\ \frac{dy_i}{dt} &= g(x_i, y_i) \end{aligned} \tag{1}$$

where $i = 1, \dots, m$.

The dynamical system obtained by considering the space of all the variables (x_i, y_i) , $i = 1, \dots, m$, displays an invariant torus full of periodic orbits that we denote by $T^m(0)$.

Assume now that the N oscillators are weakly coupled:

$$\begin{aligned} \frac{dx_i}{dt} &= f(x_i, y_i) + \epsilon F_i(x, y, \epsilon) \\ \frac{dy_i}{dt} &= g(x_i, y_i) + \epsilon G_i(x, y, \epsilon) \end{aligned} \tag{2}$$

where ϵ can be considered as small as we wish.

Definition The system [2] has a frequency locking if it displays a family of stable periodic orbits Γ_ϵ for all values of ϵ small enough which tends to (in the sense of Hausdorff’s topology) a periodic orbit of [1] contained in the periodic torus $T^m(0)$.

Assume now that [2] has a frequency locking associated with the periodic orbit $\Gamma(t)$. Consider the projections $\Gamma_i(t)$ of $\Gamma(t)$ on the coordinates plane (x_i, y_i) , $i = 1, \dots, m$. Assume that ϵ is small enough so that the projection belongs to the open set S_i on which are defined the “amplitude–phase” coordinates of the system [1]. We can write the system [2], restricted to the open set $S = \Pi_{i=1}^m S_i$, as

$$\begin{aligned} \frac{d\rho_i}{dt} &= f_i(\rho, \alpha, \epsilon) \\ \frac{d\alpha_i}{dt} &= \Phi_i(\rho, \alpha, \epsilon), \quad i = 1, \dots, m \end{aligned} \tag{3}$$

Definition The system [2] has a phase locking if the system induced by [3] on $\Gamma(t)$

$$\frac{d\alpha_i}{dt} = \Phi_i(0, \alpha, \epsilon) \tag{4}$$

has an attractive singular point.

As the attractive singular points are structurally stable, this is enough to assume that the system

$$\frac{d\alpha_i}{dt} = \Phi_i(0, \alpha, 0) \tag{5}$$

displays an attractive singular point.

Periodic Orbits of Linear Systems

Consider the linear system

$$\frac{dx}{dt} = P(t) \cdot x + q(t) \tag{6}$$

where P is a continuous T -periodic matrix function and q is a vector T -periodic continuous function, $x = (x_1, \dots, x_n)$. Consider also the two associated homogeneous equations:

$$\frac{dx}{dt} = P(t) \cdot x \tag{7a}$$

$$\frac{dx}{dt} = -P^*(t) \cdot x \tag{7b}$$

where P^* denotes the transposed of P .

The set of T -periodic solutions of [7b] is a vector space. m denotes its dimension. Let $U^j(t)$, $j = 1, \dots, m$, be a basis of this vector space. This basis is completed by adding $n - m$ solutions $U^j(t)$, $j = m + 1, \dots, n$, to obtain a basis of R^n . Let $U(t)$ be the matrix whose columns are these vectors; denote $U_{ij}(t)$ the elements of this matrix.

With the change of variable $x = U^*(0)^{-1}y$, system [6] gets transformed into

$$\frac{dy}{dt} = Q(t)y + r(t) \tag{8}$$

with $Q(t) = U^*(0)P(t)U^*(0)^{-1}$ and $r(t) = U^*(0)q(t)$.

Matrix $V(t) = U^{-1}(0)U(t)$ is such that

$$\frac{dV}{dt} + Q^*(t)V = 0, \quad V(0) = I$$

and the k first column vectors $V(t)$, denoted as $V^j(t), j = 1, \dots, m$, are T -periodic.

Let $X(t)$ be the fundamental solution defined by

$$\frac{dX}{dt} = Q(t) \cdot X, \quad X(0) = I$$

then,

$$X^{-1}(t) = V^*(t)$$

The solution of [8] can be written as

$$y(t) = X(t) \cdot y(0) + X(t) \cdot \int_0^t X^{-1}(u)r(u) du \quad [9]$$

This yields that T -periodic solutions of [8] have initial data $y(0)$ given by

$$(V^*(T) - I) \cdot y(0) = \int_0^T V^*(s)r(s) ds \quad [10]$$

Conversely, given a solution $y(0)$ of [10], T -periodicity of P and q and uniqueness of solutions of a differential equation imply that $y(0)$ represents the initial data of a T -periodic solution of [8]. Hence, the T -periodic solutions of [8] are in one-to-one correspondence with the affine space defined by the solutions of [10]. The m first rows of $V^*(T) - I$ are zero and its rank is exactly $n - m$. In the following, assume that the determinant Δ formed by the $(n - m)$ last rows and last columns of $(V^*(T) - I)$ is not zero.

A necessary and sufficient condition so that [8] displays a T -periodic solution is

$$\int_0^T \sum_{j=1}^n V_{jk}(u)r_j(u) du = 0, \quad k = 1, \dots, m \quad [11a]$$

$$\begin{aligned} &\sum_{j=m+1}^n (V_{jk}(T) - \delta_{jk})y_j(0) \\ &= \sum_{j=1}^n \int_0^T V_{jk}(s)r_j(s) ds, \quad m + 1 \leq s \leq n \quad [11b] \end{aligned}$$

This yields the Fredholm alternative, if the m conditions,

$$\sum_{j=1}^n \int_0^T U_{jk}(s)q_j(s) ds = 0, \quad k = 1, \dots, m \quad [12]$$

are satisfied, then [6] displays a family $x_\alpha(t)$ of T -periodic solutions depending of m parameters $(\alpha_1, \dots, \alpha_m)$:

$$x_\alpha(t) = \alpha_1\phi_1(t) + \dots + \alpha_m\phi_m(t) + \bar{x}(t) \quad [13]$$

where $\bar{x}(t)$ is a particular T -periodic solution and $\phi_j(t)$ denote T -periodic independent solutions of

[7a]. To be more specific, one can choose $\bar{x}(t)$ to be the unique solution of [6] such that $y(0)_k = 0, k = m + 1, \dots, n$, and $\phi_j(t)$ solutions of [7a], such that $y(0)_k = \delta_{jk}$. With these notations, $x_\alpha(t)$ is such that

$$y(0)_k = \alpha_k, \quad k = 1, \dots, m$$

and its other initial conditions $y(0)_k = \beta_k, k = m + 1, \dots, n$, are fixed:

$$\beta_k = \beta_k^0$$

Malkin's Theorem for Quasilinear Systems

Consider now nonlinear systems with the perturbation:

$$\frac{dx}{dt} = P(t) \cdot x + q(t) + \epsilon f(x, t, \epsilon) \quad [14]$$

where f is C^1 and T -periodic in t .

Assume that the solutions $y(t, y(0), \epsilon)$ of [14] exist for all values of $t, 0 \leq t \leq T$. The solutions define a differential function of their initial data $y(0)$. This is, for instance, true for perturbations of linear systems if ϵ is small enough.

Assume that q satisfies la condition [12] and that there is a solution

$$(\alpha_1^0, \dots, \alpha_m^0)$$

to the equations

$$\begin{aligned} \psi_k(\alpha) &= \sum_{j=1}^n \int_0^T U_{jk}(u)f_j(x_\alpha(u), u, 0) du = 0, \\ k &= 1, \dots, m \end{aligned} \quad [15a]$$

so that

$$\frac{\partial \psi_k(\alpha)}{\partial \alpha_j} \Big|_{\alpha=\alpha^0}; \quad k = 1, \dots, m, j = 1, \dots, m \quad [15b]$$

is invertible.

Proceed as in previous section with the coordinate change $x = U^*(0)^{-1}y$. Equation [14] gets transformed into

$$\frac{dy}{dt} = Q(t)y + r(t) + \epsilon F(y, t, \epsilon) \quad [16]$$

with $F = U^*(0)f(U^*(0)^{-1} \cdot y, t, \epsilon)$.

Solutions of [16] are uniquely determined by their initial data. We can understand the parameters (α, β) as coordinates on the space of solutions. With this viewpoint, for instance, the set of T -periodic solutions of [6] is an affine space of dimension m

given by the equations $\beta = \beta^0$ and is parametrized by the coordinates α . In this space, we pick up a point (which corresponds to a particular T -periodic solution of [6]): $(\alpha = \alpha^0)$. T -periodic solutions of [16] are in one-to-one correspondence with the solutions of

$$C_k(\alpha, \beta, \epsilon) = \sum_{j=1}^n \int_0^T V_{jk}(s) F_j(y(s, \epsilon, \alpha, \beta), s, \epsilon) ds = 0, \quad k = 1, \dots, m \tag{17a}$$

$$C_k(\alpha, \beta, \epsilon) = \sum_{j=m+1, \dots, n} (V_{jk}(T) - I) \beta_j - \sum_{j=1}^n \int_0^T V_{jk}(s) r_j(s) ds - \epsilon \sum_{j=1}^n \int_0^T V_{jk}(s) F_j(y(s, \epsilon, \alpha, \beta), s, \epsilon) ds = 0, \quad k = m + 1, \dots, n \tag{17b}$$

where $\alpha_k, k = 1, \dots, m$ and $\beta_k = y_k(0), k = m + 1, \dots, n$ parametrize the solutions $y(t, \epsilon, \alpha, \beta)$ of [14] in this way:

$$y(0) = U^*(0) \cdot x(0), \quad x(0) = \sum_{j=1}^m \alpha_j \phi_j(0) + \bar{x}(0) \tag{18}$$

Consider the determinant of the Jacobian matrix of the mapping

$$(\alpha, \beta) \mapsto C(\alpha, \beta, \epsilon) \tag{19}$$

for $\alpha = \alpha^0, \beta_k = \beta_k^0, k = m + 1, \dots, n, \epsilon = 0$. This is equal to the product of Δ and the determinant of

$$\frac{\partial \psi_k(\alpha)}{\partial \alpha_j} \Big|_{\alpha = \alpha^0} \tag{20}$$

which is nonzero.

The implicit-function theorem shows that the differential equation [14] (and thus [16] as well) has, for ϵ small enough, a unique T -periodic solution which tends to x_{α^0} when ϵ tends to 0.

Generalization of Malkin's Theorem

Finally, we consider the most general situation of the perturbation of a general system (not necessarily linear):

$$\frac{dx}{dt} = f(x, t) + \epsilon g(x, t, \epsilon) \tag{21}$$

where we assume that

$$\frac{dx}{dt} = f(x, t) \tag{22}$$

displays an m -parameter family $x_\alpha(t)$ of T -periodic orbits.

Assume that the solutions $y(t, y(0), \epsilon)$ exist for all $0 \leq t \leq T$ and define a differentiable mapping of the initial data $y(0)$. This is, for instance, the case if we assume that the nonperturbed equation defines a flow and if ϵ is small enough.

Assume also that the different solutions $x_\alpha(t)$ are independent in the sense that the mapping

$$\alpha \mapsto x_\alpha(t)$$

is an immersion for any t . In other words, the m vectors $dx_\alpha(t)/d\alpha_j$ are independent.

We linearize the solution along the family of periodic orbits:

$$x = x_\alpha(t) + \epsilon \xi \tag{23}$$

Equation [21] gets transformed into

$$\frac{d\xi}{dt} = Df_x(x_\alpha(t), t) \cdot \xi + g(x_\alpha(t), t, 0) + \epsilon F(\xi, t, \epsilon) \tag{24}$$

Set, furthermore,

$$P(t) = Df_x(x_\alpha(t), t), \quad r(t) = g(x_\alpha(t), t, 0)$$

and denote $U(t)$ the fundamental solution of [7b] described earlier.

Theorem Assume that there is a solution

$$(\alpha_1^0, \dots, \alpha_m^0)$$

of the m equations:

$$\sigma_k(\alpha) = \sum_{j=1}^n \int_0^T U_{jk}(u) g_j(x_\alpha(u), u, 0) du = 0, \quad k = 1, \dots, m \tag{25a}$$

such that

$$\frac{\partial \sigma_k(\alpha)}{\partial \alpha_j} \Big|_{\alpha = \alpha^0}; \quad k = 1, \dots, m, j = 1, \dots, m \tag{25b}$$

is invertible. Then, for all ϵ sufficiently small, eqn [21] has a unique T -periodic solution which tends to x_{α^0} when ϵ tends to 0.

We show that under the hypothesis of the theorem, we can apply the results proved in the preceding section. Note that one can prove the theorem for eqn [24] because it reduces to [21] with the change of variables [23].

Note first that the m conditions [25a] imply that the m equations,

$$\frac{d\xi}{dt} = Df_x(x_{\alpha^0}(t), t) \cdot \xi + g(x_{\alpha^0}(t), t, 0)$$

display a family of T -periodic solutions which depend on m parameters $\gamma = (\gamma_1, \dots, \gamma_m)$. From (13), one can write

$$\xi_\gamma(t) = \gamma_1 \phi_1(t) + \dots + \gamma_m \phi_m(t) + \bar{\xi}(t) \tag{26}$$

where $\bar{\xi}(t)$ is a particular T -periodic solution and the $\phi_j(t)$ are independent T -periodic solutions of (22a).

Lemma 1 *A possible choice for the solutions $\phi_j(t)$ is $\partial x_\alpha(t)/\partial \alpha_j|_{\alpha=\alpha^0}$.*

We have already assumed that these vectors are independent. They are obviously T -periodic solutions to (22a).

In the following, we will assume that all other periodic solutions of (22a) are linear combinations of these.

As a consequence of what was proved in the section on periodic orbits of linear systems, system [24] displays a periodic solution (for ϵ small enough) if there exists a solution

$$(\gamma_1^0, \dots, \gamma_m^0)$$

to equations

$$\nu_k(\gamma) = \sum_{j=1}^n \int_0^T U_{jk}(s) F_j(\xi_\gamma(s), s, 0) ds = 0,$$

$$k = 1, \dots, m$$

such that

$$\frac{\partial \nu_k(\gamma)}{\partial \gamma_j} \Big|_{\gamma=\gamma^0}; \quad k = 1, \dots, m, \quad j = 1, \dots, m$$

is invertible.

Lemma 2 *The quantities $\nu_k(\gamma)$ depend linearly in γ .*

Proof Observe first that the quantities $F_j(\xi, s, 0)$ depend quadratically of ξ :

$$\begin{aligned} F_j(\xi, s, 0) &= \frac{1}{2} \sum_{k,l} \frac{\partial^2 f_j}{\partial z_k \partial z_l} (x_{\alpha^0}(s), s) \xi_k \xi_l \\ &+ \sum_k \frac{\partial g_j}{\partial z_k} (x_{\alpha^0}(s), s, 0) \\ &+ \frac{\partial g_j}{\partial \epsilon} (x_{\alpha^0}(s), s, 0) \end{aligned} \tag{27}$$

Then, the solutions $\xi(t)$ depend linearly on γ . We thus obtain that *a priori* $\nu_p(\gamma)$ are quadratic functions of γ :

$$\begin{aligned} \nu_p(\gamma_1, \dots, \gamma_m) &= \frac{1}{2} \sum_{qrkl} \gamma_q \gamma_r \int_0^T U_{jp} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial \gamma_q} \cdot \frac{\partial z_l}{\partial \gamma_r} ds \\ &+ \sum_{qkl} \gamma_q \int_0^T U_{jp} \left[\frac{1}{2} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \left(\frac{\partial z_k}{\partial \gamma_q} \cdot \bar{\xi}_l + \frac{\partial z_l}{\partial \gamma_q} \bar{\xi}_k \right) \right. \\ &\left. + \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \gamma_q} \right] ds + \dots \end{aligned} \tag{28}$$

where the dots represent quantities independent of γ . We use then the expression

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial^2 z_j}{\partial \gamma_q \partial \gamma_r} \right) &= \sum_{kl} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial \gamma_q} \cdot \frac{\partial z_l}{\partial \gamma_r} + \sum_k \frac{\partial f_j}{\partial z_k} \frac{\partial^2 z_k}{\partial \gamma_q \partial \gamma_r} \end{aligned}$$

This allows one to find the homogeneous quadratic part as

$$\begin{aligned} \sum_{jkl} \int_0^T U_{jp} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial \gamma_q} \cdot \frac{\partial z_l}{\partial \gamma_r} ds \\ = \sum_j \int_0^T U_{jp}(s) \frac{d}{ds} \left(\frac{\partial^2 z_j}{\partial \gamma_q \partial \gamma_r} \right) ds \\ - \sum_{jk} \int_0^T U_{jp}(s) \frac{\partial f_j}{\partial z_k} \frac{\partial^2 z_k}{\partial \gamma_q \partial \gamma_r} ds \end{aligned}$$

Integration by parts yields

$$\begin{aligned} \sum_{jkl} \int_0^T U_{jp} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial \gamma_q} \cdot \frac{\partial z_l}{\partial \gamma_r} ds \\ = - \sum_j \int_0^T \left(\frac{dU_{jp}}{ds} + U_{jp}(s) \frac{\partial f_j}{\partial z_k} \right) \frac{\partial^2 z_k}{\partial \gamma_q \partial \gamma_r} ds = 0 \end{aligned}$$

because U^* is solution to [7a]. This shows that [28] is linear in γ . Suffices to show that the determinant of this system does not vanish to have existence and uniqueness of the solution such that

$$\frac{\partial \nu_1, \dots, \nu_m}{\partial \gamma_1, \dots, \gamma_m} \neq 0$$

Consider now the coefficient of the linear part:

$$\sum_{kl} \int_0^T U_{jp} \left[\frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \bar{\xi}_l + \frac{\partial g_j}{\partial z_k} \right] \cdot \frac{\partial z_k}{\partial \gamma_q} ds$$

and the coefficient

$$\sigma_p(\alpha) = \sum_{j=1}^n \int_0^T U_{jp}(u) g_j(x_\alpha(u), u, 0) du$$

We can write

$$\frac{d\sigma_p}{d\alpha_q} = \int_0^T \left(\frac{\partial U_{jp}}{\partial \alpha_q} \cdot g_j + U_{jp} \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \alpha_q} \right) ds$$

Note that

$$\frac{d\bar{\xi}_j}{dt} = \sum_r \frac{\partial f_j}{\partial z_r} \bar{\xi}_r + g_j(z(t), \alpha^0, 0)$$

and we obtain

$$\begin{aligned} \frac{d\sigma_p}{d\alpha_q} &= \int_0^T \left(\frac{\partial U_{jp}}{\partial \alpha_q} \cdot \left(\frac{d\bar{\xi}_j}{ds} - \sum_r \frac{\partial f_j}{\partial z_r} \bar{\xi}_r \right) \right. \\ &\quad \left. + U_{jp} \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \alpha_q} \right) ds \end{aligned}$$

Integration by parts yields

$$\begin{aligned} \frac{d\sigma_p}{d\alpha_q} \Big|_{\alpha=\alpha^0} &= - \int_0^T \left(\frac{d}{ds} \left(\frac{\partial U_{jp}}{\partial \alpha_q} \right) \cdot \bar{\xi}_j + \sum_r \frac{\partial f_j}{\partial z_r} \bar{\xi}_r \right) \\ &\quad + \int_0^T U_{jp} \left(\frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \alpha_q} \right) ds \end{aligned}$$

From the equation

$$\frac{dU_{jp}}{dt} + \sum_k \frac{\partial f_k}{\partial z_j} U_{kp} = 0$$

we deduce that

$$\frac{d}{dt} \left(\frac{\partial U_{jp}}{\partial \alpha_q} \right) = - \sum_k \frac{\partial f_k}{\partial z_j} \frac{\partial U_{jp}}{\partial \alpha_q} + \sum_k \frac{\partial^2 f_k}{\partial z_j \partial z_r} U_{kp} \frac{\partial z_r}{\partial \alpha_q}$$

and thus this shows that

$$\frac{d\sigma_p}{d\alpha_q} \Big|_{\alpha=\alpha^0} = \sum_{kl} \int_0^T U_{jp} \left[\frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \bar{\xi}_l + \frac{\partial g_j}{\partial z_k} \right] \cdot \frac{\partial z_k}{\partial \alpha_q} ds$$

This achieves the proof of the theorem. In the special case of Hamiltonian systems, in the case of the perturbations of an isochronous system, the method explained is equivalent to Moser’s averaging theory.

The reader is referred to other articles in this encyclopedia for a discussion of other aspects of synchronization, frequency locking, and phase locking.

See also: Bifurcation Theory; Fractal Dimensions in Dynamics; Integrable Systems: Overview; Isochronous Systems; Leray–Schauder Theory and Mapping Degree; Ljusternik–Schnirelman Theory; Singularity and Bifurcation Theory; Symmetry and Symmetry Breaking in Dynamical Systems; Synchronization of Chaos; Weakly Coupled Oscillators.

Further Reading

Hartman P (1964) *Ordinary Differential Equations*. New York: Wiley.
 Malkin I (1952) *Stability Theory of the Motion*. Moscow–Leningrad: Izdat. Gos.
 Malkin I (1956) *Some Problems in the Theory of Nonlinear Oscillations*. Gostekhizdat.
 Moser J (1970) Regularization of Kepler’s problem and the averaging method on a manifold. *Communication of Pure and Applied Mathematics* 23: 609–636.
 Roseau M (1966) *Vibrations non linéaires et théorie de la stabilité*, Springer Tracts in Natural Philosophy, vol. 8. Berlin: Springer.
 Van der Pol B (1926) On relaxation-oscillations. *Philosophical Magazine* 3(7): 978–992.
 Van der Pol B (1931) Oscillations sinusoidales et de relaxation. *L’onde électrique* 245–256.
 Van der Pol B and Van der Mark J (1927) Frequency demultiplication. *Nature* 120: 363–364.
 Van der Pol B and Van der Mark J (1928) The heart beat considered as a relaxation oscillation, and an electrical model of the heart. *Philosophical Magazine* 6(7): 763–775.

Bi-Hamiltonian Methods in Soliton Theory

M Pedroni, Università di Bergamo, Dalmine (BG), Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

At the end of the 1960s, the theory of integrable systems received a great boost by the discovery (made by Gardner, Green, Kruskal, and Miura) of the inverse-scattering method (see Integrable Systems: Overview). It allows one to reduce the

solution of the (nonlinear) Korteweg–de Vries equation (henceforth simply the KdV equation)

$$u_t = \frac{1}{4}(u_{xxx} - 6uu_x) \quad [1]$$

to the solution of linear equations. After the KdV equation, a lot of other nonlinear partial differential equations, solvable by means of the inverse-scattering method, were found out. A common feature of such equations is the existence of soliton solutions, that is, solutions in the shape of a solitary wave (with additional interaction properties). For this reason they are called “soliton equations.”

It was soon observed that the KdV equation can be seen as an infinite-dimensional Hamiltonian system with an infinite sequence of constants of motion in involution; the corresponding (commuting) vector fields are symmetries for the KdV equation, and form the so-called KdV hierarchy. In particular, Zakharov and Faddeev constructed action-angle variables for the KdV equation. These facts pointed out that the KdV equation is an infinite-dimensional analog of a classical integrable Hamiltonian system (Dubrovin *et al.* 2001), whose theory has been developed during the nineteenth century by Liouville, Jacobi, and many others. Moreover, the infinite-dimensional case suggested methods (such as the existence of a Lax pair) which were applied successfully also to finite-dimensional cases such as the Toda lattices and the Calogero systems. More recently, after the discovery by Witten and Kontsevich of remarkable relations between the KdV hierarchy and matrix models of two-dimensional (2D) quantum gravity, there has been a renewed interest in the study of soliton equations in the community of theoretical physicists. We also mention that the classical versions of the extended \mathcal{W}_n -algebras of 2D conformal field theory are the (second) Poisson structures of the Gelfand–Dickey hierarchies.

In this article we describe the so-called bi-Hamiltonian formulation of soliton equations. This approach to integrable systems springs from the observation, made by Magri at the end of the 1970s, that the KdV equation can be seen as a Hamiltonian system in two different ways. In the same circle of ideas, there were important works by Adler, Dorfman, Gelfand, Kupershmidt, Wilson, and many others. Thus, the concept of bi-Hamiltonian manifold, which constitutes the geometric setting for the study of bi-Hamiltonian systems, emerged. This notion and its applications to the theory of finite-dimensional integrable systems is discussed in Multi-Hamiltonian Systems.

In the first section of this article, we discuss the Hamiltonian form of soliton equations and, more generally, we present an important class of infinite-dimensional Poisson (also called Hamiltonian) structures, namely those of hydrodynamic type. Then we show how to use the bi-Hamiltonian properties of the KdV equation in order to construct its conserved quantities. We also recall that the KdV equation can be seen as an Euler equation on the dual of the Virasoro algebra. In the third section, we deal with other examples of integrable evolution equations admitting a bi-Hamiltonian representation, that is, the Boussinesq and the Camassa–Holm equations, and we consider the bi-Hamiltonian structures of hydrodynamic type.

Hamiltonian Methods in Soliton Theory

The most famous example of soliton equation is the KdV equation [1], where u is usually a periodic or rapidly decreasing real function. The choice of the coefficients in the equation has no special meaning, since they can be changed arbitrarily by rescaling x , t , and u . Right after the discovery of the inverse-scattering method for solving the Cauchy problem for the KdV equation, it was realized that this equation can be seen as an infinite-dimensional Hamiltonian system. Indeed, from a geometrical point of view, eqn [1] defines a vector field $X(u) = (1/4)(u_{xxx} - 6uu_x)$ on \mathcal{M} , the infinite-dimensional vector space of C^∞ functions from the unit circle S^1 to \mathbb{R} . (For the sake of simplicity, we consider only the periodic case; the integrals in this article are therefore understood to be taken on S^1 .) The vector field X associated with the KdV equation is Hamiltonian, that is, it can be factorized as

$$X(u) = [-2\partial_x] \left[\frac{1}{8}(-u_{xx} + 3u^2) \right]$$

where $dH = (1/8)(-u_{xx} + 3u^2)$ is the differential of the functional

$$H(u) = \frac{1}{8} \int \left(u^3 + \frac{1}{2} u_x^2 \right) dx$$

that is, the variational derivative $\delta h / \delta u$ of the density $h = (1/8)(u^3 + (1/2)u_x^2)$, and $P = -2\partial_x$ is a Poisson (or Hamiltonian) operator. This means that the corresponding composition law

$$\{F, G\} = \int dF P(dG) dx = -2 \int dF (dG)_x dx \quad [2]$$

between functionals of u has the usual properties of the Poisson bracket, that is, it is \mathbb{R} -bilinear and skew-symmetric, and it fulfills the Leibniz rule and the Jacobi identity. In other words, (\mathcal{M}, P) is an infinite-dimensional Poisson manifold. Using the Poisson bracket [2], eqn [1] can be written as

$$u_t = \{u, H\} \quad [3]$$

corresponding to the usual Hamilton equation in \mathbb{R}^{2n}

$$\dot{z}^i = \{z^i, H\}, \quad i = 1, \dots, 2n \quad [4]$$

up to the replacement of z with u , and of the discrete index i with the continuous index x . More precisely, in the expression $u_t = \{u, H\}$ the symbol u should be replaced by u^x (in analogy with z^i), the functional assigning to the generic function $v \in \mathcal{M}$ its value at a fixed point x , that is, $u^x : v \mapsto v(x)$. In

these notations, the Poisson bracket [2] takes the form

$$\{u^x, u^y\} = -2\delta'(x - y)$$

where the δ -function is as usual defined as

$$\int f(y)\delta(x - y) dx = f(x)$$

so that its derivatives are given by

$$\int f(y)\delta^{(k)}(x - y) dx = f^{(k)}(x)$$

Another important example is given by the Boussinesq equation

$$u_{tt} = \frac{1}{3}(-u_{xxxx} + 4u_x^2 + 4uu_{xx}) \quad [5]$$

describing, like KdV, shallow water (soliton) waves in a nonlinear approximation. It can be obtained by the first-order (in time) system

$$u_t^1 = \frac{2}{3}u^2u_x^2 + u_{xx}^1 - \frac{2}{3}u_{xxx}^2, \quad u_t^2 = 2u_x^1 - u_{xx}^2 \quad [6]$$

by taking the derivative of its second equation with respect to t , plugging the result in the first one, and setting $u = u^2$. The system [6] is Hamiltonian, since it can be written as

$$u_t^1 = \left(\frac{\delta h}{\delta u^2}\right)_x, \quad u_t^2 = \left(\frac{\delta h}{\delta u^1}\right)_x$$

with $h = (u^1)^2 + (1/9)(u^2)^3 - u^1u_x^2 + (1/3)(u_x^2)^2$, and

$$\begin{pmatrix} 0 & \partial_x \\ \partial_x & 0 \end{pmatrix} \quad [7]$$

is easily seen to be a Poisson operator. Thus, the Poisson manifold associated with the Boussinesq equation is the space of periodic C^∞ functions with values in \mathbb{R}^2 . More generally, one can consider the space \mathcal{M}^n of C^∞ functions from the unit circle S^1 to \mathbb{R}^n . If P^{ij} , for $i, j = 1, \dots, n$, are the entries of a constant skew-symmetric matrix and $u^{i,x}$ assigns to the generic function $v \in \mathcal{M}^n$ the value of its i th components at a fixed point x , then

$$\{u^{i,x}, u^{j,y}\} = P^{ij}\delta(x - y)$$

defines a Poisson bracket on \mathcal{M}^n . One can also let the P^{ij} depend on the u^k in such a way that they form the components of a Poisson tensor on \mathbb{R}^n . If $H = \int h dx$ is a functional on \mathcal{M}^n with density h , the associated Hamiltonian vector field gives rise to the following system of partial differential equations:

$$u_t^i = \sum_{j=1}^n P^{ij} \frac{\delta h}{\delta u^j}, \quad i = 1, \dots, n$$

In particular, if $n = 2N$ and

$$[P^{ij}] = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

then we have the Hamiltonian formulation of the field equations,

$$q_t^i = \frac{\delta h}{\delta p^i}, \quad p_t^i = -\frac{\delta h}{\delta q^i}, \quad i = 1, \dots, N$$

Another important example of Poisson bracket on \mathcal{M}^n is given by

$$\{u^{i,x}, u^{j,y}\} = g^{ij}\delta'(x - y) \quad [8]$$

where g^{ij} are the entries of a constant symmetric matrix. In this case, the Hamiltonian vector field associated with $H = \int h dx$ is given by

$$u_t^i = \sum_{j=1}^n g^{ij} \partial_x \left(\frac{\delta h}{\delta u^j}\right), \quad i = 1, \dots, n \quad [9]$$

Notice that this vector field is zero if $H = \int u^k dx$, with $k = 1, \dots, n$. This amounts to saying that such an H is a Casimir function of the Poisson bracket [8], that is, that $\{H, F\} = 0$ for all functionals F . A simple example of this class (with $n = 2$) is given by the Poisson structure of the Boussinesq equation, corresponding to the choice $g^{11} = g^{22} = 0$ and $g^{12} = g^{21} = 1$. Suppose now that the matrix with entries g^{ij} is invertible. Then they can be interpreted as the contravariant components of a flat pseudo-Riemannian metric in \mathbb{R}^n . A change of coordinates $(u^1, \dots, u^n) \mapsto (\bar{u}^1, \dots, \bar{u}^n)$ in \mathbb{R}^n transforms the Poisson bracket [9] in

$$\{\bar{u}^{i,x}, \bar{u}^{j,y}\} = g^{ij}(\bar{u})\delta'(x - y) + \Gamma_k^{ij}(\bar{u})\bar{u}_x^k\delta(x - y) \quad [10]$$

where $g^{ij}(\bar{u})$ are the components of the metric in the new coordinates and the Γ_k^{ij} are the contravariant Christoffel symbols related to the usual Christoffel symbols by

$$\Gamma_k^{ij} = -g^{il}\Gamma_{lk}^j \quad [11]$$

Conversely, the expression [10] gives a Poisson bracket if the metric defined by g^{ij} is flat and its Christoffel symbols are related to the Γ_k^{ij} by [11]. These are the Poisson structures of hydrodynamic type introduced by Dubrovin and Novikov. We will consider them again later.

Bi-Hamiltonian Formulation of the KdV Equation

The KdV equation [1] has a lot of remarkable properties, such as the Lax representation and the existence of a τ -function. In this section, we recall a geometrical feature of KdV, namely, the fact that it

has a second Hamiltonian structure, and we show that the integrability of KdV can be seen as a natural consequence of its double Hamiltonian representation. We have already seen that the KdV vector field $X(u) = (1/4)(u_{xxx} - 6uu_x)$ can be written as

$$X(u) = P_0 dH_2$$

where $P_0 = -2\partial_x$ and

$$H_2 = \frac{1}{8} \int \left(u^3 + \frac{1}{2} u_x^2 \right) dx$$

But X admits another Hamiltonian representation:

$$X(u) = P_1 dH_1$$

where $P_1 = -(1/2)\partial_{xxx} + 2u\partial_x + u_x$ and

$$H_1 = -\frac{1}{4} \int u^2 dx$$

The important point is that P_1 is also a Poisson operator. Moreover, it is compatible with P_0 , that is, any linear combination of P_0 and P_1 is still a Poisson operator. Thus, the KdV equation is a bi-Hamiltonian system, that is, it can be seen in two different (but compatible) ways as a Hamiltonian system. Next, we will show how this property can be used to construct an infinite sequence of conserved quantities for the KdV equation, which are in involution with respect to the Poisson brackets $\{\cdot, \cdot\}_0$ and $\{\cdot, \cdot\}_1$ associated with P_0 and P_1 . In particular, the phase space \mathcal{M} of KdV is a bi-Hamiltonian manifold, that is, it has two different (but compatible) Poisson structures. Let us rename $X_1 = X$ the KdV vector field. Since $X = P_0 dH_2 = P_1 dH_1$, one is naturally led to consider the vector fields

$$X_0 = P_0 dH_1, \quad X_2 = P_1 dH_2$$

Explicitly, $X_0(u) = u_x$ and $X_2(u) = (1/16)(u_{xxxxx} - 10uu_{xxx} - 20u_x u_{xx} + 30u^2 u_x)$. One can check that these vector fields are also bi-Hamiltonian. Indeed, $X_0(u) = P_1 dH_0$, with $H_0 = \int u dx$, and

$$X_2 = P_0 dH_3 \quad \text{with} \\ H_3 = -\frac{1}{64} \int \left(u_{xx}^2 + 5uu_x^2 + \frac{5}{2} u^4 \right) dx$$

The functional H_0 is a Casimir of P_0 , that is, $P_0 dH_0 = 0$, so that the iteration ends on this side, but it can be continued indefinitely from the other side, as shown below. For the time being, let us take for granted that there exists an infinite sequence $\{H_k\}_{k \geq 0}$ of functionals such that $P_1 dH_k = P_0 dH_{k+1}$; in other words,

$$\{\cdot, H_k\}_1 = \{\cdot, H_{k+1}\}_0 \tag{12}$$

Such relations are often called Lenard–Magri relations. Then the functionals H_k are in involution with respect to both Poisson brackets. Indeed, for $k > j$, one has

$$\begin{aligned} \{H_j, H_k\}_0 &= \{H_j, H_{k-1}\}_1 = \{H_{j+1}, H_{k-1}\}_0 \\ &= \dots = \{H_k, H_j\}_0 \end{aligned}$$

so that $\{H_j, H_k\}_0 = 0$ for all $j, k \geq 0$, and therefore $\{H_j, H_k\}_1 = 0$ for all $j, k \geq 0$. Hence, these functionals are constants of motion (in involution) for the KdV equation. The Hamiltonian vector fields associated with them are symmetries for the KdV equation; the corresponding evolution equations are called higher-order KdV equations. The set of such equations is the well-known KdV hierarchy. We remark that the existence of a sequence of functionals $\{H_k\}_{k \geq 0}$, fulfilling the Lenard–Magri relations [12] and starting from a Casimir of P_0 , is equivalent to the existence of a Casimir function $H(\lambda) = \sum_{k \geq 0} H_k \lambda^{-k}$ for the Poisson pencil $P_\lambda = P_1 - \lambda P_0$, where λ is a real parameter. A straightforward way (due essentially to Miura, Gardner, and Kruskal) to determine such a Casimir function is to consider the (generalized) Miura map $h \mapsto u = h_x + h^2 - \lambda$. As shown by Kupershmidt and Wilson, it transforms the Poisson structure $(1/2)\partial_x$ (in the variable h) into the Poisson pencil $P_\lambda = -(1/2)\partial_{xxx} + 2(u + \lambda)\partial_x + u_x$. Given u , the Riccati equation

$$h_x + h^2 = u + \lambda \tag{13}$$

admits a unique solution with the asymptotic expansion $h = z + \sum_{k \geq 1} h_k z^{-k}$, where $z^2 = \lambda$. Moreover, the coefficients h_k are differential polynomials in u (i.e., polynomials in u and its x -derivatives) that can be computed by recurrence. Thus, the generalized Miura map can be seen as an invertible transformation. Since the functional $h \mapsto \int h dx$ is a Casimir of the Poisson structure $(1/2)\partial_x$, it follows that if $h(u)$ is the solution of the Riccati equation [13], then $u \mapsto \int h(u) dx$ is a Casimir of the Poisson pencil P_λ . More precisely, one has to introduce the functional $H(\lambda) = z \int h(u) dx$, that turns out to be a Laurent series in λ , because the even coefficients of $h(u)$ are x -derivatives. This is the Casimir function we were looking for. Explicitly, one finds that the first terms of $h(u)$ are

$$\begin{aligned} h_1 &= \frac{1}{2}u, & h_2 &= -\frac{1}{4}u_x, & h_3 &= \frac{1}{8}(u_{xx} - u^2) \\ h_4 &= -\frac{1}{16}(u_{xxx} - 4uu_x) \\ h_5 &= \frac{1}{32}(u_{xxxx} - 6uu_{xx} - 5u_x^2 + 2u^3) \end{aligned}$$

Obviously, h_1 is the density of a Casimir function of P_0 , while h_3 and h_5 are (one-half of) the densities of the

two Hamiltonians H_1 and H_2 of the KdV equation. We conclude this section showing that, as observed by Khesin and Ovsienko (Arnol'd and Khesin 1998), the bi-Hamiltonian structures of KdV have a clear Lie-algebraic origin. Indeed, the second Hamiltonian structure is the Lie–Poisson structure on the dual of the Virasoro algebra, while the first one can be obtained by “freezing” the second one at a suitable point. Let $\mathcal{X}(S^1)$ be the Lie algebra of vector fields on S^1 . The Virasoro algebra is the vector space $\mathfrak{g} = \mathcal{X}(S^1) \oplus \mathbb{R}$ endowed with the Lie-algebra structure

$$\begin{aligned} & \left[\left(f(x) \frac{\partial}{\partial x}, a \right), \left(g(x) \frac{\partial}{\partial x}, b \right) \right] \\ &= \left((f'(x)g(x) - g'(x)f(x)) \frac{\partial}{\partial x}, \right. \\ & \quad \left. \int f'(x)g''(x) dx \right) \end{aligned} \quad [14]$$

It is called a central extension of $\mathcal{X}(S^1)$ since it is obtained by considering the usual commutator between vector fields (up to a sign) and by adding a copy of \mathbb{R} , which turns out to be the center of the Virasoro algebra. Equation [14] gives rise indeed to a Lie-algebra structure because the expression $\int f'(x)g''(x) dx$ defines a 2-cocycle of $\mathcal{X}(S^1)$. The dual space \mathfrak{g}^* of \mathfrak{g} can be considered as the space of the pairs $(u dx \otimes dx, c)$, where $u \in C^\infty(S^1)$ and $c \in \mathbb{R}$. The pairing is obviously given by

$$\left\langle (u dx \otimes dx, c), \left(f \frac{\partial}{\partial x}, a \right) \right\rangle = \int u(x)f(x) dx + ac$$

The Lie–Poisson structure on the dual \mathfrak{g}^* of a Lie algebra \mathfrak{g} is defined as

$$\{F, G\}(X) = \langle X, [dF(X), dG(X)] \rangle \quad [15]$$

where $F, G \in C^\infty(\mathfrak{g}^*)$ and their differentials at $X \in \mathfrak{g}^*$ are seen as elements of \mathfrak{g} . When \mathfrak{g} is the Virasoro algebra and $F(u, c) = \int f(u, c) dx, G(u, c) = \int g(u, c) dx$ are two functionals on \mathfrak{g}^* whose densities f and g are differential polynomials in u , one has

$$\begin{aligned} & \{F, G\}(u, c) \\ &= \left\langle (u dx \otimes dx, c), \left(\left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right) \right. \right. \\ & \quad \left. \left. - \left(\frac{\delta g}{\delta u} \right)' \left(\frac{\delta f}{\delta u} \right) \right) \frac{\partial}{\partial x}, \int \left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right)'' dx \right\rangle \\ &= \int u \left(\left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right) - \left(\frac{\delta g}{\delta u} \right)' \left(\frac{\delta f}{\delta u} \right) \right) dx \\ & \quad + \int c \left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right)'' dx \end{aligned} \quad [16]$$

This is (up to rescaling) the second Poisson bracket of KdV. The KdV equation is therefore an Euler equation, that is, it can be obtained from the Euler equations for the rigid body by replacing the Lie algebra of the rotation group with the Virasoro algebra. To be more precise, the Hamiltonian vector field associated with $H_1(u, c) = -(1/2)(\int u^2 dx + c)$ is

$$u_t + 3uu_x + cu_{xxx} = 0, \quad c_t = 0$$

If $c \neq 0$, this is (up to rescaling) the KdV equation [1]. For $c=0$, we have the Burgers equation (also called dispersionless KdV equation), to be discussed again later on. The first Poisson bracket for the KdV hierarchy can be obtained by “freezing” the Lie–Poisson bracket at the point $((1/2)dx \otimes dx, 0)$ of the dual of the Virasoro algebra. This means that instead of [16] one has to consider

$$\begin{aligned} & \{F, G\}_0(u, c) \\ &= \left\langle \left(\frac{1}{2} dx \otimes dx, 0 \right), \left(\left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right) \right. \right. \\ & \quad \left. \left. - \left(\frac{\delta g}{\delta u} \right)' \left(\frac{\delta f}{\delta u} \right) \right) \frac{\partial}{\partial x}, \int \left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right)'' dx \right\rangle \\ &= \frac{1}{2} \int \left(\left(\frac{\delta f}{\delta u} \right)' \left(\frac{\delta g}{\delta u} \right) - \left(\frac{\delta g}{\delta u} \right)' \left(\frac{\delta f}{\delta u} \right) \right) dx \end{aligned} \quad [17]$$

The corresponding Hamiltonian is $H_2 = (1/2) \int (-u^3 + cu_x^2) dx$. From this (Lie algebraic) point of view, the compatibility between the two Poisson brackets follows from the fact that the pencil $\{\cdot, \cdot\}_\lambda = \{\cdot, \cdot\} - \lambda\{\cdot, \cdot\}_0$ is obtained from the Lie–Poisson bracket $\{\cdot, \cdot\}$ by applying the translation

$$(u dx \otimes dx, c) \mapsto \left(\left(u + \frac{\lambda}{2} \right) dx \otimes dx, c \right)$$

Other Examples

In the previous section, we have presented the bi-Hamiltonian structure of the KdV equation and some of its properties. Now we give two more examples of equations – the Boussinesq equation and the Camassa–Holm equation – admitting a bi-Hamiltonian formulation. We have seen in an earlier section that the system [6] associated with the Boussinesq equation [5] is Hamiltonian with respect to the Poisson structure [7] and the Hamiltonian

$$H_1(u^1, u^2) = \int \left((u^1)^2 + \frac{1}{9}(u^2)^3 - u^1 u_x^2 + \frac{1}{3}(u_x^2)^2 \right) dx$$

A more complicated Poisson structure for this system is

$$P = \begin{pmatrix} A & -3\partial_x^4 + 3u^2\partial_x^2 + 9u^1\partial_x + 3u_x^1 \\ B & -6\partial_x^3 + 6u^2\partial_x + 3u_x^2 \end{pmatrix} \quad [18]$$

with

$$A = 2\partial_x^5 - 4u^2\partial_x^3 - 6u_x^2\partial_x^2 + (2(u^2)^2 + 6u_x^1 - 6u_{xx}^2)\partial_x + (3u_{xx}^1 - 2u_{xxx}^2 + 2u^2u_x^2)$$

and

$$B = 3\partial_x^4 - 3u^2\partial_x^2 + (9u^1 - 6u_x^2)\partial_x + (6u_x^1 - 3u_{xx}^2)$$

It can be obtained by means of the Drinfeld–Sokolov reduction (or also by means of a bi-Hamiltonian reduction) from the Lie–Poisson structure (modified with the cocycle ∂_x) on the space of C^∞ maps from S^1 to the Lie algebra of 3×3 traceless matrices. This is the reason why it is a Poisson structure, compatible with [7]. The system [6] can be written as

$$\begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} = P \begin{pmatrix} (\delta h_2 / \delta u^1) \\ (\delta h_2 / \delta u^2) \end{pmatrix}$$

where $h_2 = (1/3)u_1$ is the density of a Casimir of the Poisson structure [7]. Thus, the Boussinesq equation is a bi-Hamiltonian system and can be shown to possess, like KdV, an infinite sequence of conserved quantities and symmetries, forming the Boussinesq hierarchy. The KdV and the Boussinesq hierarchy are indeed particular examples of Gelfand–Dickey hierarchies (Dickey 2003). They are hierarchies of systems of n equations with n unknown functions and they are related, via the Drinfeld–Sokolov approach, to the Lie algebra $\mathfrak{sl}(n+1)$. As shown by Adler, Dickey, and Gelfand, these hierarchies have a bi-Hamiltonian formulation. Also the generalized KdV equations, associated by Drinfeld and Sokolov with an arbitrary affine Kac–Moody Lie algebra, are bi-Hamiltonian (or are obtained as suitable reductions of bi-Hamiltonian systems). Let us consider now the (dispersionless) Camassa–Holm equation

$$u_t - u_{txx} = -3uu_x + 2u_xu_{xx} + uu_{xxx} \quad [19]$$

which also describes shallow water waves, and possesses remarkable solutions called peakons, since they represent traveling waves with discontinuous first derivative. In order to supply this equation with a (bi-)Hamiltonian structure, one has to perform the change of variable $m = u - u_{xx}$, whose inverse, in the space of period-1 functions, turns out to be given by

$$u(x) = \int_0^x m(y) \sinh(y-x) dy + \frac{1}{2 \sinh(1/2)} \int_0^1 m(y) \cosh\left(y-x-\frac{1}{2}\right) dy$$

The Camassa–Holm equation is then bi-Hamiltonian with respect to the Poisson pair

$$P_1 = \partial_{xxx} - \partial_x, \quad P_2 = 2m\partial_x + m_x$$

Indeed, it can be written as $m_t = P_1 dH_2 = P_2 dH_2$, where

$$H_1 = -\frac{1}{2} \int (u^2 + u_x^2) dx$$

$$H_2 = \frac{1}{2} \int (u^3 + uu_x^2) dx$$

Notice that the Poisson pair of the Camassa–Holm equation can be obtained from that of KdV by moving the cocycle ∂_{xxx} from the second Poisson structure to the first one. Indeed,

$$P_{(a,b,c)} = a\partial_{xxx} + b\partial_x + c(2m\partial_x + m_x) \quad a, b, c \in \mathbb{R} \quad [20]$$

is a family of pairwise compatible Poisson operators. Moreover, we mention that Misiólek has shown that also the Camassa–Holm equation is an Euler equation on the dual of the Virasoro algebra. We conclude this article with a brief discussion concerning the so-called bi-Hamiltonian structures of hydrodynamic type. They play a relevant role in the theory of Frobenius manifolds, that, in turn, have deep relations with many important topics in contemporary mathematics and physics, such as Gromov–Witten invariants and isomonodromic deformations. As we have seen in the earlier section, a Poisson structure of hydrodynamic type is given, on the space of C^∞ maps from S^1 to (an open set of) \mathbb{R}^n , by

$$\{u^{i,x}, u^{j,y}\} = g^{ij}(u)\delta'(x-y) + \Gamma_k^{ij}(u)u_x^k\delta(x-y) \quad [21]$$

where $g^{ij}(u)$ are the contravariant components of a (pseudo-)Riemannian flat metric and the Γ_k^{ij} are the (contravariant) Christoffel symbols of the metric. If two Poisson structures of hydrodynamic type are given, it can be shown that they are compatible if and only if the two corresponding metrics form a flat pencil. This means that their linear combinations (with constant coefficients) are still flat (pseudo-)Riemannian metrics, and that the contravariant Christoffel symbols of the linear combinations are the linear combinations of the contravariant Christoffel symbols of the

two metrics. The simplest example is given by the bi-Hamiltonian formulation of the Burgers (or dispersionless KdV) equation,

$$u_t + 3uu_x = 0$$

that we have already encountered. We know that this equation is Hamiltonian with respect to the (Lie-)Poisson operator $2u\partial_x + u_x$, with Hamiltonian function $H_1 = -(1/2) \int u^2 dx$, and with respect to the Poisson operator ∂_x , with Hamiltonian function $H_2 = -(1/2) \int u^3 dx$. This also means that the bi-Hamiltonian structure of the Burgers equation comes from the family [20]. The first Hamiltonian structure corresponds to the standard metric on \mathbb{R} , that is, $du \otimes du$, whereas the second one is given by the metric $(2u)^{-1} du \otimes du$.

See also: Classical r -Matrices, Lie Bialgebras, and Poisson Lie Groups; Hamiltonian Fluid Dynamics; Infinite-Dimensional Hamiltonian Systems; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems: Overview; Korteweg–de Vries Equation and Other Modulation Equations; Multi-Hamiltonian Systems; Recursion Operators in Classical Mechanics; Solitons and Kac–Moody Lie Algebras; Toda Lattices; WDVV Equations and Frobenius Manifolds.

Billiards in Bounded Convex Domains

S Tabachnikov, Pennsylvania State University, University Park, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Billiard Flow and Billiard Ball Map

The billiard system describes the motion of a free particle inside a domain with elastic reflection off the boundary. More precisely, a billiard table is a Riemannian manifold M with a piecewise smooth boundary, for example, a domain in the plane. The point moves along a geodesic line with a constant speed until it hits the boundary. At a smooth boundary point, the billiard ball reflects so that the tangential component of its velocity remains the same, while the normal component changes its sign. This means that both energy and momentum are conserved. In dimension 2, this collision is described by a well-known law of geometrical optics: the angle of incidence equals the angle of reflection. Thus, the theory of billiards has much in common with geometrical optics. If the billiard ball hits a corner, its further motion is not defined.

The billiard reflection law satisfies a variational principle. Let A and B be fixed points in the billiard

Further Reading

- Arnol'd VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. New York: Springer.
- Błaszak M (1998) *Multi-Hamiltonian Theory of Dynamical Systems*. Berlin: Springer.
- Dickey LA (2003) *Soliton Equations and Hamiltonian Systems*, 2nd edn. River Edge: World Scientific.
- Dorfman I (1993) *Dirac Structures and Integrability of Nonlinear Evolution Equations*. Chichester: Wiley.
- Drinfeld VG and Sokolov VV (1985) Lie algebras and equations of Korteweg–de Vries type. *Journal of Soviet Mathematics* 30: 1975–2036.
- Dubrovin BA (1996) Geometry of 2D topological field theories. In: Donagi R *et al.* (ed.) *Integrable Systems and Quantum Groups (Montecatini Terme, 1993)*, Lecture Notes in Mathematics, vol. 1620, pp. 120–348. Berlin: Springer.
- Dubrovin BA, Krichever IM, and Novikov SP (2001) Integrable systems. I. In: Arnol'd VI (ed.) *Encyclopaedia of Mathematical Sciences. Dynamical Systems IV*, pp. 177–332. Berlin: Springer.
- Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Berlin: Springer.
- Magri F, Falqui G, and Pedroni M (2003) The method of Poisson pairs in the theory of nonlinear PDEs. In: Conte R *et al.* (ed.) *Direct and Inverse Methods in Nonlinear Evolution Equations*, Lecture Notes in Physics, vol. 632, pp. 85–136. Berlin: Springer.
- Marsden JE and Ratiu TS (1999) *Introduction to Mechanics and Symmetry*, 2nd edn. New York: Springer.
- Olver PJ (1993) *Applications of Lie Groups to Differential Equations*, 2nd edn. New York: Springer.

table and let AXB be a billiard trajectory from A to B with reflection at a boundary point X . Then, the position of a variable point X extremizes the length AXB . This is the Fermat principle of geometrical optics.

In this article, we discuss billiards in bounded convex domains with smooth boundary, also called Birkhoff billiards. A related article treats billiards in polygons (*see* Polygonal Billiards).

The billiard flow is defined as a continuous-time dynamical system. The time- t billiard transformation acts on unit tangent vectors to M which constitute the phase space of the billiard flow, and the manifold M is its configuration space. Thus, the billiard flow is the geodesic flow on a manifold with boundary.

It is useful to reduce the dimensions by one and to replace continuous time by discrete one, that is, to replace the billiard flow by a mapping, called the billiard ball map and denoted by T . The phase space of the billiard ball map consists of unit tangent vectors (x, v) with the foot point x on the boundary of M and the inward direction v . A vector (x, v) moves along the geodesic through x in the direction of v to the next point of its intersection x_1 with the boundary ∂M , and then v reflects in ∂M to the new

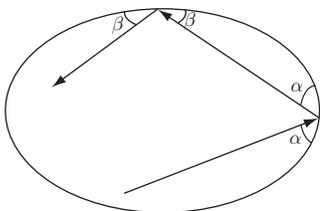


Figure 1 Billiard ball map.

inward vector v_1 . Then, one has: $T(x, v) = (x_1, v_1)$. For a convex M , the map T is continuous. If M is n -dimensional, then the dimension of the phase space of the billiard ball map is $2n - 2$.

Equivalently, and more in the spirit of geometrical optics, one considers \mathcal{L} , the space of oriented geodesics (rays of light) that intersect the billiard table. This space of lines is in one-to-one correspondence with the phase space of the billiard ball map: to an inward unit vector (x, v) there corresponds the oriented line through x in the direction v (Figure 1).

The space of rays \mathcal{L} carries a canonical symplectic structure, that is, a closed nondegenerate differential 2-form. In the Euclidean case, this symplectic structure ω is defined as follows. Given an oriented line ℓ in \mathbb{R}^n , let q be the unit vector along ℓ and p be the vector obtained by dropping the perpendicular from the origin to ℓ . Then, $\omega = dp \wedge dq = \sum dp_i \wedge dq_i$. This construction identifies \mathcal{L} with the cotangent bundle of the unit sphere: q is a unit vector and p is a (co)tangent vector at q , and ω identifies with the canonical symplectic structure of T^*S^{n-1} . In the general case of a Riemannian manifold M , the symplectic structure on the space of oriented geodesics is obtained from that on T^*M by symplectic reduction.

One has an important result: the billiard ball map preserves the symplectic structure $T^*(\omega) = \omega$. As a consequence, T is also measure preserving. In the planar case, one has the following explicit formula for this measure. Let t be an arc length parameter along the boundary of the billiard table and let $\alpha \in [0, \pi]$ be the angle made by the unit vector with this boundary. Then, (α, t) are coordinates in the phase space, identified with the cylinder, and the invariant measure is $\sin \alpha d\alpha dt$.

As a consequence, the total area of the phase space equals $2L$ where L is the perimeter length of the boundary of the billiard table, and the mean free path equals $\pi A/L$, where A is the area of the billiard table. In the general n -dimensional case, the mean free path equals

$$\frac{\text{vol}(S^{n-1})}{\text{vol}(B^{n-1})} \frac{\text{vol}(M)}{\text{vol}(\partial M)}$$

where S^{n-1} and B^{n-1} are the unit sphere and the unit disk in Euclidean spaces.

Existence and Nonexistence of Caustics

Given a plane billiard table, a caustic is a curve inside the table such that if a segment of a billiard trajectory is tangent to this curve then so is each reflected segment. Caustics correspond to invariant circles of the billiard ball map (i.e., invariant curves that go around the phase cylinder): such an invariant circle is a one-parameter family of oriented lines, and the respective caustic is their envelop. An envelop may have cusp-like singularities but if the boundary of the billiard table is a smooth curve with positive curvature then a caustic, sufficiently close to the boundary, is smooth and convex.

One can recover the table from a caustic by the following string construction. Let γ be a caustic. Wrap a closed nonstretchable string around γ , pull it tight at a point and move this point around γ to obtain a new curve Γ . Then, γ is a caustic for the billiard inside Γ . Note that this construction has one parameter, the length of the string.

The following useful “mirror equation” relates various quantities depicted in Figure 2:

$$\frac{1}{a} + \frac{1}{b} = \frac{2k}{\sin \alpha}$$

where k is the curvature of the boundary at the impact point.

Do caustics exist for every convex billiard table? This is important to know, in particular, because the existence of a caustic implies that the billiard ball map is not ergodic. The answer is given by a theorem of Lazutkin: *if the boundary of the billiard table is sufficiently smooth and its curvature never vanishes, then there exists a collection of smooth caustics in the vicinity of the billiard curve whose union has a positive area*. Originally this theorem asked for 553 continuous derivatives; later this was reduced to six. This result uses the techniques of the KAM (Kolmogorov–Arnol’d–Moser) theory. The

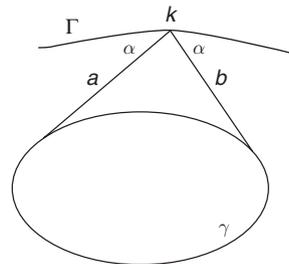


Figure 2 String construction and mirror equation.

crucial fact is that, in appropriate coordinates, the billiard ball map is approximated, near the boundary of the phase cylinder, by the integrable map $(x, y) \mapsto (x + y, y)$.

On the other hand, by a theorem of Mather, if the curvature of a convex smooth billiard curve vanishes at some point, then this billiard ball map has no invariant circles. This result belongs to the well-developed theory of area-preserving twist maps of the cylinder, of which the billiard ball map is an example.

Integrable Billiards

Let a plane billiard table be an ellipse with foci F_1 and F_2 . It is known since antiquity that a billiard ball shot from F_1 reflects to F_2 . A generalization of this optical property of the ellipse is the following theorem: *a billiard trajectory inside an ellipse forever remains tangent to a fixed confocal conic*. More precisely, if a segment of a billiard trajectory does not intersect the segment F_1F_2 , then all the segments of this trajectory do not intersect F_1F_2 and are all tangent to the same ellipse with foci F_1 and F_2 ; and if a segment of a trajectory intersects F_1F_2 , then all the segments of this trajectory intersect F_1F_2 and are all tangent to the same hyperbola with foci F_1 and F_2 .

It follows that confocal ellipses are the caustics of the billiard inside an ellipse. In particular, a neighborhood of the boundary of such a billiard table is foliated by caustics. A long-standing conjecture, attributed to Birkhoff, asserts that if a neighborhood of a strictly convex smooth boundary of a billiard table is foliated by caustics, then this table is an ellipse. This conjecture remains open. The best result in this direction is a theorem of Bialy: *if almost every phase point of the billiard ball map in a strictly convex billiard table belongs to an invariant circle, then the billiard table is a disk*.

The multidimensional analogs of the optical properties of an ellipse are as follows. Consider an ellipsoid M in \mathbb{R}^n given by the equation

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \dots + \frac{x_n^2}{a_n^2} = 1 \tag{1}$$

and define the confocal family of quadrics M_λ by the equation

$$\frac{x_1^2}{a_1^2 + \lambda} + \frac{x_2^2}{a_2^2 + \lambda} + \dots + \frac{x_n^2}{a_n^2 + \lambda} = 1$$

where λ is a real parameter. The topological type of M_λ changes as λ passes the values $-a_i^2$.

One has the following theorem: *a billiard trajectory inside M remains tangent to fixed $(n - 1)$ confocal quadrics*. A similar and closely related result holds for the geodesic curves on M : the tangent lines to a fixed geodesic on M are tangent to $(n - 2)$ other fixed quadrics, confocal with M . For a triaxial ellipsoid, this theorem goes back to Jacobi.

Explicit formulas for the integrals of the billiard in an n -dimensional ellipsoid [1] are as follows. Let (x, v) be a phase point, a unit inward tangent vector whose foot point x lies on the boundary. The following functions are invariant under the billiard ball map:

$$F_i(x, v) = v_i^2 + \sum_{j \neq i} \frac{(v_j x_i - v_i x_j)^2}{a_j^2 - a_i^2}, \quad i = 1, \dots, n$$

these functions are not independent: $F_1 + \dots + F_n = 1$.

In fact, the integrals F_i Poisson-commute (with respect to the Poisson bracket associated with the symplectic structure in the phase space of the billiard ball map that was described above). According to the Arnol'd–Liouville theorem, this complete integrability of the billiard inside an ellipsoid implies that the phase space is foliated by invariant tori and, in appropriate coordinates, the map on each torus is a parallel translation.

Similar results on complete integrability hold for billiards inside quadrics in spaces of constant positive or negative curvature. The former is the intersection of a quadratic cone with the unit sphere, and the latter with the unit pseudosphere.

Periodic Orbits

Periodic billiard trajectories inside a planar billiard table correspond to inscribed polygons of extremal perimeter length. When counting periodic trajectories, one does not distinguish between polygons obtained from each other by cyclic permutation or reversing the order of the vertices. In other words, one counts the orbits of the dihedral group D_n acting on n -periodic billiard polygons.

An additional topological characteristic of a periodic billiard trajectory is the rotation number defined as follows. Assume that the boundary γ of a billiard table is parametrized by the unit circle and consider a polygon (x_1, x_2, \dots, x_n) inscribed in γ . For all i , one has $x_{i+1} = x_i + t_i$ with $t_i \in (0, 1)$. Since the polygon is closed, $t_1 + \dots + t_n \in \mathbb{Z}$. This integer, that takes values from 1 to $n - 1$, is called the rotation number of the polygon and denoted by ρ . Changing the orientation of a polygon replaces the

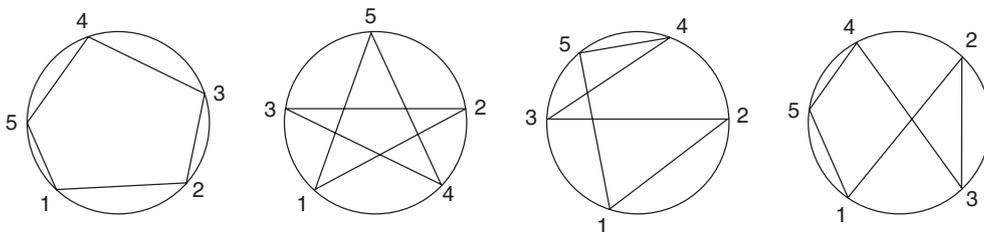


Figure 3 Rotation numbers of periodic trajectories.

rotation number ρ by $n - \rho$. The leftmost 5-periodic trajectory in Figure 3 has $\rho = 1$ and the other three $\rho = 2$.

The following theorem is due to Birkhoff: for every $n \geq 2$ and $\rho \leq \lfloor (n - 1)/2 \rfloor$, coprime with n , there exist two geometrically distinct n -periodic billiard trajectories with the rotation number ρ . For example, there are at least two 2-periodic billiard trajectories inside every smooth oval: one is the diameter, the longest chord, and another one is of minimax type, similar to the minor axis of an ellipse.

In higher dimensions, lower bounds on the number of periodic billiard trajectories inside strictly convex domains with smooth boundaries were obtained only recently by Farber and the present author. Here is one of the results: for a generic billiard table in \mathbf{R}^m , the number of n -periodic trajectories is not less than $(n - 1)(m - 1)$. The proof consists in using the Morse theory to estimate below the number of critical points of the perimeter length function on the space of inscribed n -gons and its quotient space by the dihedral group D_n , and the main difficulty is in describing the topology of these spaces.

Returning to convex smooth planar billiards, the following conjecture remains open for a long time: the set of n -periodic points of the billiard ball map has zero measure. This is easy for $n = 2$; for $n = 3$ this is a theorem by M Rychlik. The motivation for this question comes from spectral geometry. In particular, according to a theorem of Ivrii, the above conjecture implies the Weyl conjecture on the second term for the spectral asymptotics of the Laplacian in a bounded domain with the Dirichlet or Neumann boundary conditions.

Length Spectrum

The set of lengths of the closed trajectories in a convex billiard M is called the length spectrum of M . There is a remarkable relation between the length spectrum and the spectrum of the Laplace operator in M with the Dirichlet boundary condition:

$\Delta f = \lambda f, f|_{\partial M} = 0$. From the physical point of view, the eigenvalues λ are the eigenfrequencies of the membrane M with a fixed boundary. Roughly speaking, one can recover the length spectrum from that of the Laplacian. More precisely, the following theorem of K Anderson and R Melrose holds:

$$\sum_{\lambda_i \in \text{spec } \Delta} \cos(t\sqrt{-\lambda_i})$$

is a well-defined generalized function (distribution) of t , smooth away from the length spectrum. That is, if $l > 0$ belongs to the singular support of this distribution, then there exists either a closed billiard trajectory of length l , or a closed geodesic of length l in the boundary of the billiard table.

This relation between the Laplacian and the length spectrum is due to the fact that geometric optics is not a very accurate description of light. In wave optics, light is considered as electromagnetic waves, and geometric optics gives a realistic approximation only when the wave length is small. This small-wave approximation is based on the assumption that the waves are locally almost harmonic, while their amplitudes change slowly from point to point. The substitution of such a function into the corresponding PDEs gives, in the first approximation, the equations of wave fronts, that is, of geometric optics.

Here is another spectral result concerning a smooth strictly convex plane domain, due to S Marvizi and R Melrose. Let L_n be the supremum and l_n the infimum of the perimeters of simple billiard n -gons. Then,

$$\lim_{n \rightarrow \infty} n^k (L_n - l_n) = 0$$

for any positive k . Furthermore, L_n has an asymptotic expansion, as $n \rightarrow \infty$,

$$L_n \sim l + \sum_{i=1}^{\infty} \frac{c_i}{n^{2i}}$$

where l is the length of the boundary of billiard table and c_i are constants, depending on the curvature of the boundary.

Acknowledgments

This work was partially supported by NSF.

See also: Adiabatic Piston; Hamiltonian Systems: Obstructions to Integrability; Hyperbolic Billiards; Integrable Discrete Systems; Integrable Systems and Algebraic Geometry; Optical Caustics; Integrable Systems: Overview; Polygonal Billiards; Semiclassical Spectra and Closed Orbits; Separatrix Splitting; Stability Theory and KAM.

Further Reading

- Chernov N and Markarian R, *Theory of Chaotic Billiards* (to appear).
 Farber M and Tabachnikov S (2002) Topology of cyclic configuration spaces and periodic orbits of multi-dimensional billiards. *Topology* 41: 553–589.

Black Hole Mechanics

A Ashtekar, Pennsylvania State University,
 University Park, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Over the last 30 years, black holes have been shown to have a number of surprising properties. These discoveries have revealed unforeseen relations between the otherwise distinct areas of general relativity, quantum physics, and statistical mechanics. This interplay, in turn, led to a number of deep puzzles at the very foundations of physics. Some have been resolved while others continue to baffle physicists. The starting point of these fascinating developments was the discovery of laws of black hole mechanics by Bardeen, Bekenstein, Carter, and Hawking. They dictate the behavior of black holes in equilibrium, under small perturbations away from equilibrium, and in fully dynamical situations. While they are consequences of classical general relativity alone, they have a close similarity with the laws of thermodynamics. The origin of this seemingly strange coincidence lies in quantum physics. For further discussion, see Asymptotic Structure and Conformal Infinity; Loop Quantum Gravity; Quantum Geometry and Its Applications; Quantum Field Theory in Curved Spacetime; Stationary Black Holes.

The focus of this article is just on black hole mechanics. The discussion is divided into three parts. In the first, we will introduce the notions of event horizons and black hole regions and discuss properties

- Gutkin E (2003) Billiard dynamics: a survey with the emphasis on open problems. *Regular and Chaotic Dynamics* 8: 1–13.
 Katok A and Hasselblatt B (1995) *Introduction to the Modern Theory of Dynamical Systems*. Cambridge: Cambridge University Press.
 Kozlov V and Treshchev D (1991) *Billiards. A Genetic Introduction to the Dynamics of Systems with Impacts*. Providence: American Mathematical Society.
 Lazutkin V (1993) *KAM Theory and Semiclassical Approximations to Eigenfunctions*. Berlin: Springer.
 Moser J (1980) *Various Aspects of Integrable Hamiltonian Systems*. Progress in Mathematics, vol. 8, pp. 233–289. Basel: Birkhäuser.
 Siburg KF (2004) *The Principle of Last Action in Geometry and Dynamics*, Lecture Notes in Mathematics, vol. 1844. Berlin: Springer.
 Sinai Ya (1976) *Introduction to Ergodic Theory*. Princeton: Princeton University Press.
 Tabachnikov S (1995) *Billiards*, Société Math. de France, Panoramas et Synthèses, No 1.
 Tabachnikov S (2005) *Geometry and Billiards*. American Mathematical Society (to appear).

of globally stationary black holes. In the second, we will consider black holes which are themselves in equilibrium but in surroundings which may be time dependent. Finally, in the third part, we summarize what is known in the fully dynamical situations. For simplicity, all manifolds and fields are assumed to be smooth and, unless otherwise stated, spacetime is assumed to be four dimensional, with a metric of signature $- , + , + , +$, and the cosmological constant is assumed to be zero. An arrow under a spacetime index denotes the pullback of that index to the horizon.

Global Equilibrium

To capture the intuitive notion that black hole is a region from which signals cannot escape to the asymptotic part of spacetime, one needs a precise definition of future infinity. The standard strategy is to use Penrose's conformal boundary \mathcal{I}^+ . A black hole region \mathcal{B} of a spacetime (\mathbb{M}, g_{ab}) is defined as $\mathcal{B} = \mathbb{M} \setminus I^-(\mathcal{I}^+)$, where I^- denotes “chronological past.” The boundary $\partial\mathcal{B}$ of the black hole region is called the “event horizon” and denoted by \mathcal{E} . Thus, \mathcal{E} is the boundary of the past of \mathcal{I}^+ . It therefore follows that \mathcal{E} is a null 3-surface, ruled by future inextendible null geodesics without caustics. If the spacetime is globally hyperbolic, an “instant of time” is represented by a Cauchy surface M . The intersection of \mathcal{B} with M may have several disjoint components, each representing a black hole at that instant of time. If M' is a Cauchy surface to the future of M , the number of disjoint components of $M' \cup \mathcal{B}$ in the causal future of $M \cup \mathcal{B}$ must be less than or equal to those of $M \cup \mathcal{B}$

(see [Hawking and Ellis \(1973\)](#)). Thus, black holes can merge but can not bifurcate. (By a time reversal, i.e., by replacing \mathcal{J}^+ with \mathcal{J}^- and I^- with I^+ , one can define a white hole region \mathcal{W} . However, here we will focus only on black holes.)

A spacetime (\mathbb{M}, g_{ab}) is said to be stationary (i.e., time independent) if g_{ab} admits a Killing field t^a that represents an asymptotic time translation. By convention, t^a is assumed to be unit at infinity. (\mathbb{M}, g_{ab}) is said to be axisymmetric if g_{ab} admits a Killing field ϕ^a generating an $SO(2)$ isometry. By convention ϕ^a is normalized such that the affine length of its integral curves is 2π . Stationary spacetimes with nontrivial $\mathbb{M} \setminus I^-(\mathcal{J}^+)$ represent black holes which are in global equilibrium. In the Einstein–Maxwell theory in four dimensions, there exists a unique three-parameter family of stationary black hole solutions, generally parametrized by mass m , angular momentum J , and electric charge Q . This is the celebrated Kerr–Newman family. Therefore, in general relativity a great deal of work on black holes has focused on these solutions and perturbations thereof. The Kerr–Newman family is axisymmetric and furthermore, its metric has the property that the 2-flats spanned by the Killing fields t^a and ϕ^a are orthogonal to a family of 2-surfaces. This property is called “ t – ϕ orthogonality.” These features of Kerr–Newman space-times are widely used in black hole physics. Note however that uniqueness fails in higher dimensions, and also in the presence of nonabelian gauge fields or rings of perfect fluids around black holes in four dimensions. In mathematical physics, there is significant literature on the new stationary black hole solutions in Einstein–Yang–Mills–Higgs theories. These are called “hairy black holes.” Research on stationary black hole solutions with rings received a boost by a recent discovery that these black holes can violate the Kerr inequality $J \leq Gm^2$ between angular momentum J and mass m .

A null 3-manifold \mathcal{K} in \mathbb{M} is said to be a “Killing horizon” if g_{ab} admits a Killing field K^a which is everywhere normal to \mathcal{K} . On a Killing horizon, one can show that the acceleration of K^a is proportional to K^a itself:

$$K^a \nabla_a K^b = \kappa K^b \quad [1]$$

The proportionality function κ is called “surface gravity.” We will show in the next section that if a mild energy condition holds on \mathcal{K} , then κ must be constant. Note that if we rescale K^a via $K^a \rightarrow cK^a$, where c is a constant, surface gravity also rescales as $\kappa \rightarrow c\kappa$.

In the Kerr–Newman family, the event horizon is a Killing horizon. More generally, if an axisymmetric, stationary black hole spacetime (\mathbb{M}, g_{ab})

satisfies the t – ϕ orthogonality property, its event horizon \mathcal{E} is a Killing horizon. (Although one can envisage stationary black holes in which these additional symmetry conditions are not met, this possibility has been ignored in black hole mechanics on stationary spacetimes. Quasilocal horizons, discussed below, do not require any spacetime symmetries.) In these cases, the normalization freedom in K^a is fixed by requiring that K^a have the form

$$K^a = t^a + \Omega \phi^a \quad [2]$$

on the horizon, where Ω is a constant, called the “angular velocity of the horizon.” The resulting κ is called the surface gravity of the black hole. It is remarkable that κ is constant for all such black holes, even when their horizon is highly distorted (i.e., far from being spherically symmetric) either due to rotation or due to external matter fields. This is analogous to the fact that the temperature of a thermodynamical system in equilibrium is constant, independently of the details of the system. In analogy with thermodynamics, constancy of κ is referred to as the “zeroth law of black hole mechanics.”

Next, let us consider an infinitesimal perturbation δ within the three-parameter Kerr–Newman family. A simple calculation shows that the changes in the Arnowitt–Deser–Misner (ADM) mass m , angular momentum J , and the total charge Q of the spacetime and in the area a of the horizon are constrained via

$$\delta m = \frac{\kappa}{8\pi G} \delta a + \Omega \delta J + \Phi \delta Q \quad [3]$$

where the coefficients κ, Ω, Φ are black hole parameters, $\Phi = A_a K^a$ being the electrostatic potential at the horizon. The last two terms, $\Omega \delta J$ and $\Phi \delta Q$, have the interpretation of “work” required to spin the black hole up by an amount δJ or to increase its charge by δQ . Therefore, [3] has a striking resemblance to the first law, $\delta E = T \delta S + \delta W$, of thermodynamics if (as the zeroth law suggests) κ is made proportional to the temperature T , and the horizon area a to the entropy S . Therefore, [3] and its generalizations discussed below are referred to as the “first law of black hole mechanics.”

In Kerr–Newman spacetimes, the only contribution to the stress–energy tensor comes from the Maxwell field. [Bardeen *et al.* \(1973\)](#) consider stationary black holes with matter such as perfect fluids in the exterior region and stationary perturbations δ thereof. Using Einstein’s equations, they show that the form [3] of the first law does not change; the only modification is addition of certain matter terms on the right-hand side which can be

interpreted as the work δW done on the total system. A generalization in another direction was made by Iyer and Wald (1994) using Noether currents. They allow nonstationary perturbations and, more importantly, drop the restriction to general relativity. Instead, they consider a wide class of diffeomorphism-invariant Lagrangian densities $L(g_{ab}, R_{abcd}, \nabla_a R_{bcde}, \dots, \Phi^{\dots}, \nabla_a \Phi^{\dots}, \dots)$ which depend on the metric g_{ab} , matter fields Φ^{\dots} , and a finite number of derivatives of the Riemann tensor and matter fields. Finally, they restrict themselves to $\kappa \neq 0$. In this case, on the maximal analytic extension of the spacetime, the Killing field K^a vanishes on a 2-sphere S_o called the bifurcate horizon. Then, [3] is generalized to

$$\delta m = \frac{\kappa}{2\pi} \delta S_{\text{hor}} + \delta W \quad [4]$$

Here δW again represents “work terms” and S_{hor} is given by

$$S_{\text{hor}} = -2\pi \oint_{S_o} \frac{\delta L}{\delta R_{abcd}} n_{ab} n_{cd} \quad [5]$$

where n_{ab} is the binormal to S_o (with $n_{ab} n^{ab} = -2$), and the functional derivative inside the integral is evaluated by formally viewing the Riemann tensor as a field independent of the metric. For the Einstein–Hilbert action, this yields $S_{\text{hor}} = a/4G$ and one recovers [3].

These results are striking. However, the underlying assumptions have certain unsatisfactory aspects. First, although the laws are meant to refer just to black holes, one assumes that the entire spacetime is stationary. In thermodynamics, by contrast, one only assumes that the system under consideration is in equilibrium, not the whole universe. Second, in the first law, quantities a, Ω, Φ are evaluated at the horizon while M, J are evaluated at infinity and include contributions from possible matter fields outside the black hole. A more satisfactory law of black hole mechanics would involve attributes of the black hole alone. Finally, the notion of the event horizon is extremely global and teleological since it explicitly refers to \mathcal{J}^+ . An event horizon may well be developing in the very room you are sitting today in anticipation of a gravitational collapse in the center of our galaxy which may occur a billion years hence. This feature makes it impossible to generalize the first law to fully dynamical situations and relate the change in the event horizon area to the flux of energy and angular momentum falling across it. Indeed, one can construct explicit examples of dynamical black holes in which an event horizon \mathcal{E} forms and grows in the flat part of a spacetime where nothing happens

physically. These considerations call for a replacement of \mathcal{E} by a quasilocal horizon which leads to a first law involving only horizon attributes, and which can grow only in response to the influx of energy. Such horizons are discussed in the next two sections.

Local Equilibrium

The key idea here is drop the requirement that spacetime should admit a stationary Killing field and ask only that the intrinsic horizon geometry be time independent. Consider a null 3-surface Δ in a spacetime (\mathbb{M}, g_{ab}) with a future-pointing normal field ℓ^a . The pullback $q_{ab} := g_{ab}$ of the spacetime metric to Δ is the intrinsic, degenerate “metric” of Δ with signature $0, +, +$. The first condition is that it be “time independent,” that is, $\mathcal{L}_\ell q_{ab} = 0$ on Δ . Then by restriction, the spacetime derivative operator ∇ induces a natural derivative operator D on Δ . While D is compatible with q_{ab} , that is, $D_a q_{bc} = 0$, it is not uniquely determined by this property because q_{ab} is degenerate. Thus, D has extra information, not contained in q_{ab} . The pair (q_{ab}, D) is said to determine the intrinsic geometry of the null surface Δ . This notion leads to a natural definition of a horizon in local equilibrium. Let Δ be a null, three-dimensional submanifold of (\mathbb{M}, g_{ab}) with topology $\mathbb{S} \times \mathbb{R}$, where \mathbb{S} is compact and without boundary.

Definition 1 Δ is said to be “isolated horizon” if it admits a null normal ℓ^a such that:

- (i) $\mathcal{L}_\ell q_{ab} = 0$ and $[\mathcal{L}_\ell, D] = 0$ on Δ and
- (ii) $-T^a_b \ell^b$ is a future pointing causal vector on Δ .

One can show that, generically, this null normal field ℓ^a is unique up to rescalings by positive constants.

Both conditions are local to Δ . In particular, (\mathbb{M}, g_{ab}) is not required to be asymptotically flat and there is no longer any teleological feature. Since Δ is null and $\mathcal{L}_\ell q_{ab} = 0$, the area of any of its cross sections is the same, denoted by a_Δ . As one would expect, one can show that there is no flux of gravitational radiation or matter across Δ . This captures the idea that the black hole itself is in equilibrium. Condition (ii) is a rather weak “energy condition” which is satisfied by all matter fields normally considered in classical general relativity. The nontrivial condition is (i). It extracts from the notion of a Killing horizon just a “tiny part” that refers only to the intrinsic geometry of Δ . As a result, every Killing horizon \mathcal{K} is, in particular, an isolated horizon. However, a spacetime with an isolated horizon Δ can admit gravitational radiation and dynamical matter fields away from Δ . In fact, as a family of Robinson–Trautman spacetimes illustrates,

gravitational radiation could even be present arbitrarily close to Δ . Because of these possibilities, there are many nontrivial examples and the transition from event horizons of stationary spacetimes to isolated horizons represents a significant generalization of black hole mechanics. (In fact, the derivation of the zeroth and the first law requires slightly weaker assumptions, encoded in the notion of a “weakly isolated horizon” (Ashtekar *et al.* 2000, 2001).)

An immediate consequence of the requirement $\mathcal{L}_\ell q_{ab} = 0$ is that there exists a 1-form ω_a on Δ such that $D_a \ell^b = \omega_a \ell^b$. Following the definition of κ on a Killing horizon, the surface gravity $\kappa_{(\ell)}$ of (Δ, ℓ) is defined as $\kappa_{(\ell)} = \omega_a \ell^a$. Again, under $\ell^a \rightarrow c \ell^a$, we have $\kappa_{(c\ell)} = c \kappa_\ell$. Together with Einstein’s equations, the two conditions of Definition 1 imply $\mathcal{L}_\ell \omega_a = 0$ and $\ell^a D_{[a} \omega_{b]} = 0$. The Cartan identity relating the Lie and exterior derivative now yields

$$D_a(\omega_b \ell^b) \equiv D_a \kappa_{(\ell)} = 0 \quad [6]$$

Thus, surface gravity is constant on every isolated horizon. This is the zeroth law, extended to horizons representing local equilibrium. In the presence of an electromagnetic field, Definition 1 and the field equations imply $\mathcal{L}_\ell F_{ab} = 0$ and $\ell^a F_{ab} = 0$. The first of these equations implies that one can always choose a gauge in which $\mathcal{L}_\ell A_a = 0$. By Cartan identity it then follows that the electrostatic potential $\Phi_{(\ell)} := A_a \ell^a$ is constant on the horizon. This is the Maxwell analog of the zeroth law.

In this setting, the first law is derived using a Hamiltonian framework (Ashtekar *et al.* 2000, 2001). For concreteness, let us assume that we are in the asymptotically flat situation and the only gauge field present is electromagnetic. One begins by restricting oneself to horizon geometries such that Δ admits a rotational vector field φ^a satisfying $\mathcal{L}_\varphi q_{ab} = 0$. (In fact for black hole mechanics, it suffices to assume only that $\mathcal{L}_\varphi \epsilon_{ab} = 0$, where ϵ_{ab} is the intrinsic area 2-form on Δ . The same is true on dynamical horizons discussed in the next section.) One then constructs a phase space Γ of gravitational and matter fields such that (1) \mathbb{M} admits an internal boundary Δ which is an isolated horizon; and (2) all fields satisfy asymptotically flat boundary conditions at infinity. Note that the horizon geometry is allowed to vary from one phase-space point to another; the pair (q_{ab}, D) induced on Δ by the spacetime metric only has to satisfy Definition 1 and the condition $\mathcal{L}_\varphi q_{ab} = 0$.

Let us begin with angular momentum. Fix a vector field ϕ^a on \mathbb{M} which coincides with the fixed φ^a on Δ and is an asymptotic rotational symmetry at infinity. (Note that ϕ^a is not restricted in any way in the bulk.) Lie derivatives of gravitational and

matter fields along ϕ^a define a vector field $\mathbf{X}(\phi)$ on Γ . One shows that it is an infinitesimal canonical transformation, that is, satisfies $\mathcal{L}_{\mathbf{X}(\phi)} \Omega = 0$, where Ω is the symplectic structure on Γ . The Hamiltonian $H(\phi)$ generating this canonical transformation is given by

$$\begin{aligned} H(\phi) &= J_\Delta^{(\phi)} - J_\infty^{(\phi)} \\ J_\Delta^{(\phi)} &= -\frac{1}{8\pi G} \oint_S (\omega_a \phi^a) \epsilon - \frac{1}{4\pi} \oint_S (A_a \phi^a)^* F \end{aligned} \quad [7]$$

where $J_\infty^{(\phi)}$ is the ADM angular momentum at infinity, S is any cross section of Δ , and ϵ the area element thereon. The term $J_\Delta^{(\phi)}$ is independent of the choice of S made in its evaluation and interpreted as the “horizon angular momentum.” It has numerous properties that support this interpretation. In particular, it yields the standard angular momentum expression in Kerr–Newman spacetimes.

To define horizon energy, one has to introduce a “time-translation” vector field t^a . At infinity, t^a must tend to a unit time translation. On Δ , it must be a symmetry of q_{ab} . Since ℓ^a and φ^a are both horizon symmetries, $t^a = c \ell^a + \Omega \varphi^a$ on Δ , for some constants c and Ω . However, unlike ϕ^a , the restriction of t^a to Δ cannot be fixed once and for all but must be allowed to vary from one phase-space point to another. In particular, on physical grounds, one expects Ω to be zero at a phase-space point representing a nonrotating black hole but nonzero at a point representing a rotating black hole. This freedom in the boundary value of t^a introduces a qualitatively new element. The vector field $\mathbf{X}(t)$ on Γ defined by the Lie derivatives of gravitational and matter fields does not, in general, satisfy $\mathcal{L}_{\mathbf{X}(t)} \Omega = 0$; it need not be an infinitesimal canonical transformation. The necessary and sufficient condition is that $(\kappa_{(c\ell)}/8\pi G) \delta a_\Delta + \Omega \delta J_\Delta + \Phi_{(c\ell)} \delta Q_\Delta$ be an exact variation. That is, $\mathbf{X}(t)$ generates a Hamiltonian flow if and only if there exists a function $E_\Delta^{(t)}$ on Γ such that

$$\delta E_\Delta^{(t)} = \frac{\kappa_{(c\ell)}}{8\pi G} \delta a_\Delta + \Omega \delta J_\Delta + \Phi_{(c\ell)} \delta Q_\Delta \quad [8]$$

This is precisely the first law. Thus, the framework provides a deeper insight into the origin of the first law: it is the necessary and sufficient condition for the evolution generated by t^a to be Hamiltonian. Equation [8] is a genuine restriction on the choice of phase-space functions c and Ω , that is, of restrictions to Δ of evolution fields t^a . It is easy to verify that \mathbb{M} admits many such vector fields. Given one, the Hamiltonian $H(t)$ generating the time evolution along t^a takes the form

$$H(t) = E_\infty^{(t)} - E_\Delta^{(t)} \quad [9]$$

re-enforcing the interpretation of $E_{\Delta}^{(t)}$ as the horizon energy.

In general, there is a multitude of first laws, one for each vector field t^a , the evolution along which preserves the symplectic structure. In the Einstein–Maxwell theory, given any phase-space point, one can choose a canonical boundary value t_o^a exploiting the uniqueness theorem. $E_{\Delta}^{(t_o)}$ is then called the horizon mass and denoted simply by m_{Δ} . In the Kerr–Newman family, $H(t_o)$ vanishes and m_{Δ} coincides with the ADM mass m_{∞} . Similarly, if ϕ^a is chosen to be a global rotational Killing field, $J_{\Delta}^{(\phi)}$ equals $J_{\infty}^{(\phi)}$. However, in more general spacetimes where there is matter field or gravitational radiation outside Δ , these equalities do not hold; m_{Δ} and J_{Δ} represent quantities associated with the horizon alone while the ADM quantities represent the total mass and angular momentum in the spacetime, including contributions from matter fields and gravitational radiation in the exterior region. In the first law [8], only the contributions associated with the horizon appear.

When the uniqueness theorem fails, as, for example, in the Einstein–Yang–Mills–Higgs theory, first laws continue to hold but the horizon mass m_{Δ} becomes ambiguous. Interestingly, these ambiguities can be exploited to relate properties of hairy black holes with those of the corresponding solitons. (For a summary, see [Ashtekar and Krishnan \(2004\)](#).)

Dynamical Situations

A natural question now is whether there is an analog of the second law of thermodynamics. Using event horizons, Hawking showed that the answer is in the affirmative (see [Hawking and Ellis \(1973\)](#)). Let (\mathbb{M}, g_{ab}) admit an event horizon \mathcal{E} . Denote by ℓ^a a geodesic null normal to \mathcal{E} . Its expansion is defined as $\theta_{(\ell)} := q^{ab} \nabla_a \ell_b$, where q^{ab} is any inverse of the degenerate intrinsic metric q_{ab} on \mathcal{E} , and determines the rate of change of the area element of \mathcal{E} along ℓ^a . Assuming that the null energy condition and Einstein’s equations hold, the Raychaudhuri equation immediately implies that if $\theta_{(\ell)}$ were to become negative somewhere it would become infinite within a finite affine parameter. Hawking showed that, if there is a globally hyperbolic region containing $I^-(\mathcal{J}^+) \cup \mathcal{E}$ – that is, if there are no naked singularities – this can not happen, whence $\theta(\ell) \geq 0$ on \mathcal{E} . Hence, if a cross section S_2 of \mathcal{E} is to the future of a cross section S_1 , we must have $a_{S_2} \geq a_{S_1}$. Thus, in any (i.e., not necessarily infinitesimal) dynamical process, the change Δa in the horizon area is always non-negative. This result is known as the “second law of black hole mechanics.” As in the first law, the analog of entropy is the horizon area.

It is tempting to ask if there is a local physical process directly responsible for the growth of area. For event horizons, the answer is in the negative since they can grow in a flat portion of spacetime. However, one can introduce quasilocal horizons also in the dynamical situations and obtain the desired result ([Ashtekar and Krishnan 2003](#)). These constructions are strongly motivated by earlier ideas introduced by [Hayward \(1994\)](#).

Definition 2 A three-dimensional spacelike submanifold \mathcal{H} of (\mathbb{M}, g_{ab}) is said to be a “dynamical horizon” if it admits a foliation by compact 2-manifolds \mathbb{S} (without boundary) such that:

- (i) the expansion $\theta_{(\ell)}$ of one (future directed) null normal field ℓ^a to \mathbb{S} vanishes and the expansion of the other (future directed) null normal field, n^a is negative; and
- (ii) $-T^a_b \ell^b$ is a future pointing causal vector on \mathcal{H} .

One can show that this foliation of \mathcal{H} is unique and that \mathbb{S} is either a 2-sphere or, under degenerate and physically over-restrictive conditions, a 2-torus. Each leaf \mathbb{S} is a marginally trapped surface and referred to as a “cut” of \mathcal{H} . Unlike event horizons \mathcal{E} , dynamical horizons \mathcal{H} are locally defined and do not display any teleological feature. In particular, they cannot lie in a flat portion of spacetime. Dynamical horizons commonly arise in numerical simulations of evolving black holes as world tubes of apparent horizons. As the black hole settles down, \mathcal{H} asymptotes to an isolated horizon Δ , which tightly hugs the asymptotic future portion of the event horizon. However, during the dynamical phase, \mathcal{H} typically lies well inside \mathcal{E} .

The two conditions in [Definition 2](#) immediately imply that the area of cuts of \mathcal{H} increases monotonically along the “outward direction” defined by the projection of ℓ^a on \mathcal{H} . Furthermore, this change turns out to be directly related to the flux of energy falling across \mathcal{H} . Let R denote the “radius function” on \mathcal{H} so that the area of any cut \mathbb{S} is given by $a_{\mathbb{S}} = 4\pi R^2$. Let N denote the norm of $\partial_a R$ and $\Delta\mathcal{H}$, the portion of \mathcal{H} bounded by two cross sections \mathbb{S}_1 and \mathbb{S}_2 . The appropriate energy turns out to be associated with the vector field $N\ell^a$, where ℓ^a is normalized such that its projection on \mathcal{H} is the unit normal $\hat{\tau}^a$ to the cuts \mathbb{S} . In the generic and physically interesting case when \mathbb{S} is a 2-sphere, the Gauss and the Codazzi (i.e., constraint) equations imply

$$\begin{aligned} \frac{1}{2G}(R_2 - R_1) &= \int_{\Delta\mathcal{H}} T_{ab} N \ell^a \hat{\tau}^b d^3V + \frac{1}{16\pi G} \\ &\times \int_{\Delta\mathcal{H}} N \left(\sigma_{ab} \sigma^{ab} + 2\zeta_a \zeta^a \right) d^3V \quad [10] \end{aligned}$$

Here $\hat{\tau}^a$ is the unit normal to \mathcal{H} , σ^{ab} the shear of ℓ^a (i.e., the tracefree part of $q^{am}q^{bm}\nabla_m\ell_n$), and $\zeta^a = q^{ab}\hat{\tau}^c\nabla_c\ell_b$, where q^{ab} is the projector onto the tangent space of the cuts \mathbb{S} . The first integral on the right-hand side can be directly interpreted as the flux across $\Delta\mathcal{H}$ of matter–energy (relative to the vector field N^{ℓ^a}). The second term is purely geometric and is interpreted as the flux of energy carried by gravitational waves across $\Delta\mathcal{H}$. It has several properties which support this interpretation. Thus, not only does the second law of black hole mechanics hold for a dynamical horizon \mathcal{H} , but the “cause” of the increase in the area can be directly traced to physical processes happening near \mathcal{H} .

Another natural question is whether the first law [8] can be generalized to fully dynamical situations, where δ is replaced by a finite transition. Again, the answer is in the affirmative. We will outline the idea for the case when there are no gauge fields on \mathcal{H} . As with isolated horizons, to have a well-defined notion of angular momentum, let us suppose that the intrinsic 3-metric on \mathcal{H} admits a rotational Killing field φ . Then, the angular momentum associated with any cut \mathbb{S} is given by

$$J_{\mathbb{S}}^{(\varphi)} = -\frac{1}{8\pi G} \oint_{\mathbb{S}} K_{ab}\varphi^a\hat{\tau}^b d^2V \equiv \frac{1}{8\pi G} \oint_{\mathbb{S}} j^{(\varphi)} d^2V \quad [11]$$

where K_{ab} is the extrinsic curvature of \mathcal{H} in (\mathbb{M}, g_{ab}) and $j^{(\varphi)}$ is interpreted as “the angular momentum density.” Now, in the Kerr family, the mass, surface gravity, and the angular velocity can be unambiguously expressed as well-defined functions $\bar{m}(a, J)$, $\bar{\kappa}(a, J)$, and $\bar{\Omega}(a, J)$ of the horizon area a and angular momentum J . The idea is to use these expressions to associate mass, surface gravity, and angular velocity with each cut of \mathcal{H} . Then, a surprising result is that the difference between the horizon masses associated with cuts \mathbb{S}_1 and \mathbb{S}_2 can be expressed as the integral of a locally defined flux across the portion $\Delta\mathcal{H}$ of \mathcal{H} bounded by \mathcal{H}_1 and \mathcal{H}_2 :

$$\bar{m}_2 - \bar{m}_1 = \frac{1}{8\pi G} \int_{\Delta\mathcal{H}} \bar{\kappa} da + \frac{1}{8\pi G} \left\{ \oint_{\mathbb{S}_2} \bar{\Omega} j^{(\varphi)} d^2V - \oint_{\mathbb{S}_1} \bar{\Omega} j^{(\varphi)} d^2V - \int_{\bar{\Omega}_1}^{\bar{\Omega}_2} d\bar{\Omega} \oint_{\mathbb{S}} j^{(\varphi)} d^2V \right\} \quad [12]$$

If the cuts \mathbb{S}_2 and \mathbb{S}_1 are only infinitesimally separated, this expression reduces precisely to the standard first law involving infinitesimal variations. Therefore, [12] is an integral generalization of the first law.

Let us conclude with a general perspective. On the whole, in the passage from event horizons in stationary spacetimes to isolated horizons and then to dynamical horizons, one considers increasingly more realistic situations. In all the three cases, the analysis has been extended to allow the presence of

a cosmological constant Λ . (The only significant change is that the topology of cuts \mathbb{S} of dynamical horizons is restricted to be \mathbb{S}^2 if $\Lambda > 0$ and is completely unrestricted if $\Lambda < 0$.) In the first two frameworks, results have also been extended to higher dimensions. Since the notions of isolated and dynamical horizons make no reference to infinity, these frameworks can be used also in spatially compact spacetimes. The notion of an event horizon, by contrast, does not naturally extend to these spacetimes. On the other hand, the generalization [4] of the first law [3] is applicable to event horizons of stationary spacetimes in a wide class of theories while so far the isolated and dynamical horizon frameworks are tied to general relativity (coupled to matter satisfying rather weak energy conditions). From a mathematical physics perspective, extension to more general theories is an important open problem.

See also: Asymptotic Structure and Conformal Infinity; Branes and Black Hole Statistical Mechanics; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Geometric Flows and the Penrose Inequality; Loop Quantum Gravity; Minimal Submanifolds; Quantum Field Theory in Curved Spacetime; Quantum Geometry and its Applications; Random Algebraic Geometry, Attractors and Flux Vacua; Shock Wave Refinement of the Friedman–Robertson–Walker Metric; Stationary Black Holes.

Further Reading

- Ashtekar A, Beetle C, and Lewandowski J (2001) Mechanics of rotating black holes. *Physical Review* 64: 044016 (gr-qc/0103026).
- Ashtekar A, Fairhurst S, and Krishnan B (2000) Isolated horizons: Hamiltonian evolution and the first law. *Physical Review D* 62: 104025 (gr-qc/0005083).
- Ashtekar A and Krishnan B (2003) Dynamical horizons and their properties. *Physical Review D* 68: 104030 (gr-qc/0308033).
- Ashtekar A and Krishnan B (2004) Isolated and dynamical horizons and their applications. *Living Reviews in Relativity* 10: 1–78 (gr-qc/0407042).
- Bardeen JW, Carter B, and Hawking SW (1973) The four laws of black hole mechanics. *Communications in Mathematical Physics* 31: 161.
- DeWitt BS and DeWitt CM (eds.) (1972) *Black Holes*. Amsterdam: North-Holland.
- Frolov VP and Novikov ID (1998) *Black Hole Physics*. Dordrecht: Kluwer.
- Hawking SW and Ellis GFR (1973) *Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
- Hayward S (1994) General laws of black hole dynamics. *Physical Review D* 49: 6467–6474.
- Iyer V and Wald RM (1994) Some properties of noether charge and a proposal for dynamical black hole entropy. *Physical Review D* 50: 846–864.
- Wald RM (1994) *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.

Boltzmann Equation (Classical and Quantum)

M Pulvirenti, Università di Roma “La Sapienza,”
Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Ludwig Boltzmann (1872) established an evolution equation to describe the behavior of a rarefied gas, starting from the mathematical model of elastic balls and using mechanical and statistical considerations. The importance of this equation is twofold. First, it provides a reduced description (as well as the hydrodynamical equations) of the microscopic world. Second, it is also an important tool for the applications, especially for dilute fluids when the hydrodynamical equations fail to hold.

The starting point of the Boltzmann analysis is to abandon the study of the gas in terms of the detailed motion of molecules which constitute it because of their large number. Instead, it is better to investigate a function $f(x, v)$, which is the probability density of a given particle, where x and v denote its position and velocity. Actually, $f(x, v)dx dv$ is often confused with the fraction of molecules falling in the cell of the phase space of size $dx dv$ around x, v . The two concepts are not exactly the same, but they are asymptotically equivalent (when the number of particles is diverging) if a law of large numbers holds.

The Boltzmann equation is the following:

$$(\partial_t + v \cdot \nabla_x)f = Q(f, f) \quad [1]$$

where Q , the collision operator, is defined by eqn [2]:

$$Q(f, f) = \int_{\mathbb{R}^3} dv_1 \int_{S_+^2} dn(v - v_1) \cdot n \times [f(x, v')f(x, v'_1) - f(x, v)f(x, v_1)] \quad [2]$$

and

$$\begin{aligned} v' &= v - n[n \cdot (v - v_1)] \\ v'_1 &= v_1 + n[n \cdot (v - v_1)] \end{aligned} \quad [3]$$

Moreover, n (the impact parameter) is a unitary vector and $S_+^2 = \{n | n \cdot (v - v_1) \geq 0\}$.

Note that v', v'_1 are the outgoing velocities after a collision of two elastic balls with incoming velocities v and v_1 and centers x and $x + m$, r being the diameter of the spheres. Obviously, the collision takes place if $n \cdot (v - v_1) \geq 0$. Equations [3] are a consequence of the conservation of total energy, momentum, and angular momentum. Note also that r does not enter in eqn [1] as a parameter.

As fundamental features of eqn [1], we have the conservation in time of the following five quantities

$$\int dx \int dv f(x, v; t) v^\alpha \quad [4]$$

with $\alpha = 0, 1, 2$, expressing conservation of the probability, momentum, and energy.

From now on we shall set $\int = \int_{\mathbb{R}^3}$ for notational simplicity.

Moreover, Boltzmann introduced the (kinetic) entropy defined as

$$H(f) = \int dx \int dv f \log f(x, v) \quad [5]$$

and proved the famous H -theorem asserting the decreasing of $H(f(t))$ along the solutions to eqn [1].

Finally, in the case of bounded domains or homogeneous solutions ($f = f(v; t)$ is independent of x), the distribution defined for some $\beta > 0, \rho > 0$, and $u \in \mathbb{R}^3$ by

$$M(x, v) = \frac{\rho}{(2\pi/\beta)^{3/2}} e^{-(\beta/2)|v-u|^2} \quad [6]$$

called Maxwellian distribution, is stationary for the evolution given by eqn [1]. In addition, M minimizes H among all distributions with given total mass ρ , given mean velocity u , and mean energy. The parameter β is interpreted as the inverse temperature.

In conclusion, Boltzmann was able to introduce not only an evolutionary equation with the remarkable properties expressing mass, momentum, and energy conservation, but also the trend to the thermal equilibrium. In other words, he tried to conciliate the Newton's laws with the second principle of thermodynamics.

The Boltzmann Heuristic Argument

Thus, we want to find an evolution equation for the quantity $f(x, v; t)$. The molecular system we are considering consists of N identical particles of diameter r in the whole space \mathbb{R}^3 . We denote by $x_1, v_1, \dots, x_N, v_N$ a state of the system, where x_i and v_i indicate the position and the velocity of the particle i . The particles cannot overlap (i.e., the centers of two particles cannot be at a distance smaller than the particle diameter r).

The particles are moving freely up to the first instance of contact, that is, the first time when two particles (say particles i and j) arrive at a distance r . Then the pair interacts when an elastic collision occurs. This means that they change instantaneously

their velocities, according to the conservation of the energy and linear and angular momentum. More precisely, the velocities after a collision with incoming velocities v and v_1 are those given by formula [3]. After the first collision, the system evolves by iterating the procedure. Here we neglect triple collisions because they are unlikely. The evolution equation for a tagged particle is then of the form

$$(\partial_t + v \cdot \nabla_x)f = \text{Coll} \quad [7]$$

where Coll denotes the variation of f due to the collisions.

We have

$$\text{Coll} = G - L \quad [8]$$

where L and G (the loss and gain terms, respectively) are the negative and positive contributions to the variation of f due to the collisions. More precisely, $L dx dv dt$ is the probability of the test particle to disappear from the cell $dx dv$ of the phase space because of a collision in the time interval $(t, t + dt)$ and $G dx dv dt$ is the probability to appear in the same time interval for the same reason. Let us consider the sphere of center x with radius r and a point $x + nr$ over the surface, where n denotes the generic unit vector. Consider also the cylinder with base area $dS = r^2 dn$ and height $|V|dt$ along the direction of $V = v_2 - v$.

Then a given particle (say particle 2) with velocity v_2 can contribute to L because it can collide with the test particle in the time dt , provided it is localized in the cylinder and if $V \cdot n \leq 0$. Therefore, the contribution to L due to the particle 2 is the probability of finding such a particle in the cylinder (conditioned to the presence of the first particle in x). This quantity is $f_2(x, v, x + nr, v_2) |(v_2 - v) \cdot n| r^2 dn dv_2 dt$, where f_2 is the joint distribution of two particles. Integrating in dn and dv_2 , we obtain that the total contribution to L due to any predetermined particle is

$$r^2 \int dv_2 \int_{S_+^2} dn f_2(x, v, x + nr, v_2) |(v_2 - v) \cdot n| \quad [9]$$

where S_+^2 is the unit hemisphere $(v_2 - v) \cdot n < 0$. Finally, we obtain the total contribution multiplying by the total number of particles:

$$L = (N - 1)r^2 \int dv_2 \times \int_{S_+^2} dn f_2(x, v, x + nr, v_2) |(v_2 - v) \cdot n| \quad [10]$$

The gain term can be derived analogously by considering that we are looking at particles which have velocities v and v_2 after the collisions so

that we have to integrate over the hemisphere $S_+^2 = \{(v_2 - v) \cdot n > 0\}$:

$$G = (N - 1)r^2 \int dv_2 \times \int_{S_+^2} dn f_2(x, v, x + nr, v_2) |(v_2 - v) \cdot n| \quad [11]$$

Summing G and $-L$, we get

$$\text{Coll} = (N - 1)r^2 \int dv_2 \times \int_{S_+^2} dn f_2(x, v, x + nr, v_2) (v_2 - v) \cdot n \quad [12]$$

which, however, is not a very useful expression because the time derivative of f is expressed in terms of another object, namely f_2 . An evolution equation for f_2 will imply f_3 , the joint distribution of three particles, and so on, up to we include the total particle number N . Here the basic main assumption of Boltzmann enters, namely that two given particles are uncorrelated if the gas is rarefied, namely

$$f(x, v, x_2, v_2) = f(x, v)f(x_2, v_2) \quad [13]$$

Condition [13], referred to as the propagation of chaos, seems contradictory at first sight: if two particles collide, correlations are created. Even though we could assume eqn [13] at some time, if the test particle collides with particle 2, such an equation cannot be satisfied anymore after the collision.

Before discussing the propagation of chaos hypothesis, we first analyze the size of the collision operator. We remark that, in practical situations for a rarefied gas, the combination $Nr^3 \approx 10^{-4} \text{ cm}^3$ (i.e., the volume occupied by the particles) is very small, while $Nr^2 = O(1)$. This implies that $G = O(1)$. Therefore, since we are dealing with a very large number of particles, we are tempted to perform the limit $N \rightarrow \infty$ and $r \rightarrow 0$ in such a way that $r^2 = O(N^{-1})$. As a consequence, the probability that two tagged particles collide (which is of the order of the surface of a ball, i.e., $O(r^2)$) is negligible. However, the probability that a given particle performs a collision with any one of the remaining $N - 1$ particles (which is $O(Nr^2) = O(1)$) is not negligible. Therefore, condition [13] is referring to two preselected particles (say particles 1 and 2), so that it is not unreasonable to conceive that it holds in the limiting situation in which we are working.

However, we cannot insert [13] in [12] because this latter equation refers to instants before and after the collision and, if we know that a collision took place, we certainly cannot invoke eqn [13]. Hence, it is more convenient to assume eqn [13] in the loss term and work over the gain term to keep advantage

of the factorization property which will be assumed only before the collision.

Coming back to eqn [11] for the outgoing pair velocities v, v_2 (satisfying the condition $(v_2 - v) \cdot n > 0$), we make use of the continuity property

$$f_2(x, v, x + nr, v_2) = f_2(x, v', x + nr, v'_2) \quad [14]$$

where the pair v', v'_2 is pre-collisional. On f_2 expressed before the collision, we can reasonably apply condition [13] and obtain

$$\begin{aligned} G - L &= (N - 1)r^2 \int dv_2 \int_{S_+^2} dn(v - v_2) \cdot n \\ &\quad \times [f(x, v')f(x - nr, v'_2) \\ &\quad - f(x, v)f(x + nr, v_2)] \end{aligned} \quad [15]$$

after a change $n \rightarrow -n$ in the gain term, using the notation S_+^2 for the hemisphere $\{n | (v_2 - v) \cdot n \geq 0\}$. This transforms the pair v', v'_2 from a pre-collisional to a post-collisional pair.

Finally, in the limit $N \rightarrow \infty$, $r \rightarrow 0$, $Nr^2 = \lambda^{-1}$, we find

$$\begin{aligned} (\partial_t + v \cdot \nabla_x)f \\ = \lambda^{-1} \int dv_2 \int_{S_+^2} dn(v - v_2) \cdot n \\ \quad \times [f(x, v')f(x, v'_2) - f(x, v)f(x, v_2)] \end{aligned} \quad [16]$$

The parameter λ , called mean free path, represents, roughly speaking, the typical length a particle can cover without undergoing any collision. In eqns [1] and [2], we just chose $\lambda = 1$.

Equation [16] (or, equivalently, eqns [1] and [2]) is the Boltzmann equation for hard spheres. Such an equation has a statistical nature, and it is not equivalent to the Hamiltonian dynamics from which it has been derived. Indeed, the H -theorem shows that such an equation is not reversible in time as expected of any law of mechanics.

This concludes the heuristic preliminary analysis of the Boltzmann equation. We certainly know that the above arguments are delicate and require a more rigorous and deeper analysis. If we want the Boltzmann equation not to be a phenomenological model, derived by *ad hoc* assumptions and justified only by its practical relevance, but rather that it is a consequence of a mechanical model, we must derive it rigorously. In particular, the propagation of chaos should be not a hypothesis but the statement of a theorem.

Beyond the Hard Spheres

The heuristic arguments we have developed so far can be extended to different potentials than that of the hard-sphere systems. If the particles interact via

a two-body interaction $V = V(r)$, the resulting Boltzmann equation is eqn [1], with

$$Q(f, f) = \int dv_1 \int_{S_+^2} dn B(v - v_1; n) [f'f'_1 - ff_1] \quad [17]$$

where we are using the usual shorthand notation:

$$\begin{aligned} f' &= f(x, v'), \quad f'_1 = f(x, v'_1), \quad f = f(x, v), \\ f_1 &= f(x, v_1) \end{aligned} \quad [18]$$

and $B = B(v - v_1; n)$ is a suitable function of the relative velocity $v - v_1$ and the impact parameter n , which is proportional to the cross section relative to the potential V . Another equivalent, sometimes more convenient, way, to express eqn [17] is

$$\begin{aligned} Q(f, f) &= \int dv_1 \int dv' \int dv'_1 W(v, v_1 | v', v'_1) \\ &\quad [f'f'_1 - ff_1] \end{aligned} \quad [19]$$

with

$$\begin{aligned} W(v, v_1 | v', v'_1) \\ = w(v, v_1 | v', v'_1) \times \delta(v + v_1 - v' - v'_1) \\ \quad \times \delta\left(\frac{1}{2}(v^2 + v_1^2 - (v')^2 - (v'_1)^2)\right) \end{aligned} \quad [20]$$

where w is a suitable kernel. All the qualitative properties, such as the conservation laws and the H -theorem, are obviously still valid.

Consequences

The Boltzmann equation provoked a debate involving Loschmidt, Zermelo, and Poincaré, who outlined inconsistencies between the irreversibility of the equation and the reversible character of the Hamiltonian dynamics. Boltzmann argued the statistical nature of his equation and his answer to the irreversibility paradox was that “most” of the configurations behave as expected by the thermodynamical laws. However, he did not have the probabilistic tools for formulating in a precise way the statements of which he had a precise intuition.

Grad (1949) stated clearly the limit $N \rightarrow \infty$, $r \rightarrow 0$, $Nr^2 \rightarrow \text{const.}$, where N is the number of particles and r is the diameter of the molecules, in which the Boltzmann equation is expected to hold. This limit is usually called the Boltzmann–Grad limit (B–G limit in the sequel).

The problem of a rigorous derivation of the Boltzmann equation was an open and challenging problem for a long time. Lanford (1975) showed that, although for a very short time, the Boltzmann equation can be derived starting from the mechanical model of the hard-sphere system. The proof has a deep content but is relatively simple from a technical viewpoint.

Existence

The mathematical study of the Boltzmann equation starts with the problem of proving the existence of the solutions. One would like to be able to show that, for all (or at least for a physically significant family of) initial distributions (which are positive and summable functions) with finite momentum, energy, and entropy, there exists a unique solution to eqn [1] with the same mass, momentum, and energy as of the initial distribution. Moreover, the entropy should decrease and the solution should approach the right Maxwellian as $t \rightarrow \infty$. The problem, in such a generality, is still unsolved, but several results in this direction have been achieved since the pioneering works due to Carleman (1933) for the homogeneous equation. Actually, there are satisfactory results for some special situations, such as the homogeneous solutions (independent of x) close to the equilibrium, to the vacuum, or to homogeneous data. The most general result we have up to now is, unfortunately, not constructive. This is due to Di Perna and Lions (1989), who showed the existence of suitable weak solutions to eqn [1]. However, we still do not know whether such solutions, which preserve mass and momentum, and satisfy the H -theorem, are unique and also preserve the energy.

Hydrodynamics

The derivation of hydrodynamical equations from the Boltzmann equation is a problem as old as the equation itself and, in fact, it goes back to Maxwell and Hilbert. Preliminary to the discussion of the hydrodynamic limit, we establish a few properties of the collision kernel.

It is a well-known fact that the only solution to the equation

$$Q(f, f) = 0 \quad [21]$$

is a local Maxwellian, namely

$$\begin{aligned} f(x, v) &:= M(x, v) \\ &= \frac{\rho(x)}{(2\pi T(x))^{3/2}} e^{-|v-u(x)|^2/2T(x)} \end{aligned} \quad [22]$$

where the local parameters $\rho, \rho u$, and T satisfy the relations

$$\int M \, dv = \rho \quad [23]$$

$$\int v M = \rho u \quad [24]$$

$$\frac{1}{2} \int v^2 M \, dv = \frac{3}{2} \rho T + \frac{1}{2} \rho u^2 \quad [25]$$

Moreover, the only solution to the equation

$$\int b(v) Q(f, f) \, dv = 0 \quad [26]$$

is any linear combination of the quantities $(1, v, v^2)$, called collision invariants. The last property obviously corresponds to the mass, momentum, and energy conservation.

With this in mind, consider a change of variables in the Boltzmann equation [1], passing from microscopic to macroscopic variables, $x \rightarrow \varepsilon x$, $t \rightarrow \varepsilon t$. Here ε is a small scale parameter expressing the ratio between the typical inter-particle distances and the typical distances over which the macroscopic equations are varying. Such a change yields

$$(\partial_t + v \cdot \nabla_x) f_\varepsilon = \frac{1}{\varepsilon} Q(f_\varepsilon, f_\varepsilon) \quad [27]$$

We need to allow the small parameter ε (mean free path or the Knudsen number) to tend to zero. In order to eliminate the singularity on the right-hand side of [27], we multiply both sides by the collision invariants v^α with $\alpha = 0, 1, 2$, and obtain the five equations:

$$\int dv v^\alpha (\partial_t + v \cdot \nabla_x) f_\varepsilon = 0 \quad [28]$$

On the other hand, if f_ε converges to f , as $\varepsilon \rightarrow 0$, necessarily $Q(f, f) = 0$ and hence $f = M$. Therefore, we expect that in the limit $\varepsilon \rightarrow 0$,

$$\int dv v^\alpha (\partial_t + v \cdot \nabla_x) M = 0 \quad [29]$$

Equation [29] fixes a relation among the fields ρ, u, T as functions of x and t . A standard computation gives us the Euler equations for compressible gas

$$\partial_t \rho + \operatorname{div}(\rho u) = 0 \quad [30]$$

$$\partial_t u + (u \cdot \nabla) u + \frac{1}{\rho} \nabla p = 0 \quad [31]$$

$$\partial_t T + (u \cdot \nabla) T + \frac{2}{3} T \nabla u = 0 \quad [32]$$

where the pressure p is related to the density ρ and the temperature T by the perfect gas law

$$p = \rho T \quad [33]$$

In order to make the above arguments rigorous, Hilbert (1916) developed a useful tool, called the Hilbert expansion, to control the limiting procedure.

Namely, he expressed a formal solution to eqn [27] in the form of a power series expansion:

$$f_\varepsilon = \sum_{j \geq 0} f_j \varepsilon^j \quad [34]$$

where f_0 is the local Maxwellian, with the parameters ρ, u, T satisfying the Euler equations. All the other coefficients f_j of the developments can be determined by recurrence, inverting suitable operators. However, the series is not expected to be convergent, so that the way to show the validity of the hydrodynamical limit rigorously is to truncate the expansion and to control the remainder. The first result in this direction was obtained by Caflisch (1980). However, this approach is based on the regularity of the solutions to the Euler equations, which is known to hold only for short times since shocks can be formed. How to approximate the shocks in terms of a kinetic description is still a difficult and open problem.

Note that the hydrodynamical picture of the Boltzmann equation just means that we are looking at the solutions of this equation at a suitable macroscopic scale. The rarefaction hypothesis underlying the Boltzmann description is reflected in the law of perfect gas, which states that the particles, in the local thermal equilibrium, are free.

Stationary Problems

Stationary non-Maxwellian solutions to the Boltzmann equation should describe stationary nonequilibrium states exhibiting nontrivial flows. In spite of the physical relevance of these problems, not many complete mathematical results are, at the moment, available. Among them, there is the traveling-wave problem, which can be formulated in the following way. We look for a solution $f = f(x - ct, v), f: \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^+$, constant in form but traveling with a constant velocity $c > 0$, to

$$(v_1 - c)f' = \mathcal{Q}(f, f) \quad [35]$$

where v_1 is the first component of v and f' denotes the spatial derivative of f . Equation [35] must be complemented by the boundary conditions which are $f \rightarrow M_\pm$, as $x \rightarrow \infty$, where M_\pm are the right and left Maxwellians, namely two prescribed equilibrium situations at infinity. The parameters (density, mean velocity, and temperature) of the Maxwellians, however, cannot be chosen arbitrarily. Indeed, the conservations of the mass, momentum, and energy (which are properties of \mathcal{Q}) imply the conservations (in x) of the fluxes of these quantities. Hence, we have to impose five equations that relate

the upstream and the downstream values of the densities, mean velocities, and temperatures. Such relations are known in gas dynamics as the Rankine–Hugoniot conditions. A solution of this problem has been found by Caflisch and Nikolaenko (1983) in case of a weak shock (namely, when M_+ and M_- are close) by using Hilbert expansion techniques. More recently, Liu and Yu (2004) established also stability and positivity of this solution.

Quantum Kinetic Theory

Uehling and Uhlenbeck (1933) introduced the following kinetic equation for describing a large system of weakly interacting bosons or fermions:

$$\begin{aligned} (\partial_t + v \cdot \nabla_x)f &= \int dv_1 \int dv' \int dv'_1 W(v, v_1 | v', v'_1) \\ &\quad \times \{ (1 \pm f)(1 \pm f_1)f'f'_1 \\ &\quad - (1 \pm f')(1 \pm f'_1)ff_1 \} \end{aligned} \quad [36]$$

Here the $+/-$ sign, stand for bosons/fermions, respectively, and

$$\begin{aligned} W(v, v_1 | v', v'_1) &= (\hat{V}(v' - v) - \hat{V}(v' - v_1))^2 \delta(v + v_1 - v' - v'_1) \\ &\quad \times \delta\left(\frac{1}{2}(v^2 + v_1^2 - (v')^2 - (v'_1)^2)\right) \end{aligned} \quad [37]$$

Moreover,

$$\hat{V}(p) = 4\pi \int dx e^{ip \cdot x} \quad [38]$$

where V is the interaction potential. Note that eqn [37] is the expression of the cross section of a quantum scattering in the Born approximation.

The unknown $f = f(x, v; t)$ in eqn [37] is the expected number of molecules falling in the unit (quantum) cell of the phase space. This function is proportional to the one-particle Wigner function, introduced by Wigner (1932) to handle kinetic problems in quantum mechanics, and defined as (setting $\hbar = 1$):

$$\frac{1}{(2\pi)^3} \int dy e^{iy \cdot v} \rho\left(x + \frac{1}{2}y; x - \frac{1}{2}y\right)$$

where $\rho(x; z)$ is the kernel of a one-particle density matrix. Basically, the Wigner function is an equivalent way to describe a state of a quantum system. For instance, eqn [40] below expresses the equilibrium distributions for bosons and fermions in terms of Wigner functions. In general, the Wigner functions, due to the uncertainty principle, are real but not necessarily positive; however, the integral with respect to x and v gives the probability

distributions of the velocity and the position, respectively. In the kinetic regime, in which we are interested, the scales are mesoscopic, namely the typical quantum oscillations are on a scale much smaller than the characteristic scales of the problem, so that we expect that f should be a genuine probability distribution, since the Heisenberg principle does not play an essential role. However, the interaction occurs on a microscopic scale, so that we expect that the statistics play a role in addition to the quantum rules for the scattering.

In this framework, the entropy functional is

$$H(f) = \int dx \int dv [f(x, v) \log f(x, v) \mp (1 \pm f(x, v)) \log(1 \pm f(x, v))] \quad [39]$$

It is decreasing along the solutions to eqn [35] and it is also minimized (among the distributions with given mass, momentum, and energy) by the equilibria

$$M(v) = \frac{z}{e^{(\beta/2)|v-u|^2} \mp z} \quad [40]$$

namely the Bose–Einstein and the Fermi–Dirac distributions, respectively. Here $\beta > 1$ and $z > 0$ are the inverse temperature and the activity, respectively. Note that, for the Bose–Einstein distribution, $z < 1$. This creates, in a sense, an inconsistency with eqn [36]. Indeed, assuming $u=0$ and an initial distribution $f = f_0(v)$ with the density larger than the maximal density allowed by eqn [40], namely

$$\rho_c := \int dv \frac{1}{e^{(\beta/2)v^2} - 1} \quad [41]$$

it cannot converge to any equilibrium. In order to overcome this difficulty related to the Bose condensation, one can enlarge the definition of the equilibria family by setting

$$M(v) = \frac{1}{e^{(\beta/2)v^2} - 1} + \mu\delta(v) \quad [42]$$

to take care of excess of mass by means of a condensate component. However, it is not clear whether eqn [36] can actually describe the Bose condensation since its derivation from the Schrödinger equation requires, just from the very beginning, the existence of bosonic quasifree states which can be constructed only if the density is moderate. Further analyses are certainly needed to clarify the situation. A rigorous derivation of the Uehling and Uhlenbeck equation is, up to now, far from being obtained even for short times; nevertheless, such an equation is extensively used in the applications. Equation [36] concerns a weakly interacting gas of quantum particles. From a mathematical viewpoint, it is expected to be valid in the so-called weak-coupling

limit, which consists in scaling space and time and the interaction potential ϕ as

$$x \rightarrow \varepsilon x, \quad t \rightarrow \varepsilon t, \quad \phi \rightarrow \sqrt{\varepsilon} \phi \quad [43]$$

where $\varepsilon^{-1} = N^{1/3}$ is a parameter diverging when the number of particles N tends to infinity.

We mention, incidentally, that under such a scaling, a classical system is described by a transport equation, called Fokker–Planck–Landau equation, with a diffusion operator in the velocity space.

The B–G limit considered for classical particle systems is different from that considered here for weakly interacting quantum systems. It is actually equivalent to rescaling space and time according to

$$x \rightarrow \varepsilon x, \quad t \rightarrow \varepsilon t \quad [44]$$

leaving the interaction unscaled but, in order to control the total interaction, we make the density diverging gently as $\varepsilon^{-1} = N^{1/2}$.

A quantum system under such a scaling is expected to be described by a Boltzmann equation [1] with the collision operator \mathcal{Q} computed with the full quantum cross section. Now we do not have any effect of the statistics because in this rarefaction limit these corrections disappear. On the other hand, the cross section is that arising from the analysis of the quantum scattering. Since we do not rescale the interaction, all the other terms in the Born expansion of the cross section play a role. This kind of Boltzmann equation is a good description of a rarefied gas in which quantum effects are not negligible.

See also: Adiabatic Piston; Evolution Equations: Linear and Nonlinear; Gravitational N -Body Problem (Classical); Interacting Particle Systems and Hydrodynamic Equations; Kinetic Equations; Multiscale Approaches; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach; Quantum Dynamical Semigroups.

Further Reading

- Balesku R (1978) *Equilibrium and Nonequilibrium Statistical Mechanics*. Moscow: Mir (distributed by Imported Publications, Chicago, Ill).
- Caflich RE (1980) The fluid dynamical limit of the nonlinear Boltzmann equation. *Communications of Pure and Applied Mathematics* 33: 651–666.
- Caflich RE and Nicolaenko B (1983) Shock waves and the Boltzmann equation. Nonlinear partial differential equations. *Contemporary Mathematics* 17: 35–44.
- Carleman T (1933) Sur la théorie de l'équation intégral-différentielle de Boltzmann. *Acta Mathematica* 60: 91–146.
- Cercignani C (1998) *Ludwig Boltzmann. The Man Who Trusted Atoms*. Oxford: Oxford University Press.
- Cercignani C, Illner R, and Pulvirenti M (1994) The Mathematical Theory of Dilute Gases. *Springer Series in Applied Mathematics*, vol. 106. New York: Springer.

- Di Perna RJ and Lions P-L (1989) On the Cauchy problem for the Boltzmann equations: Global existence and weak stability. *Annals of Mathematics* 130: 321–366.
- Grad H (1949) On the kinetic theory of rarified gases. *Communications in Pure and Applied Mathematics* 2: 331–407.
- Hilbert D (1916) Begründung der Kinetischen Gastheorie. *Mathematische Annalen* 72: 331–407.
- Lanford OE III (1975) Time evolution of large classical systems. In: Ehlers J, Hepp K, and Weidenmüller HA (eds.) *Lecture Notes in Physics*, vol. 38, pp. 1–111. Berlin: Springer.

- Liu T-P and Yu S-H (2004) Boltzmann equation: micro–macro decompositions and positivity of shock profiles. *Communications in Mathematical Physics* 246(1): 133–179.
- Spohn H (1994) Quantum kinetic equations. In: Fannes M, Maes C, and Verbeure A (eds.) *On Three Levels: Micro, Meso and Macro Approaches in Physics*. New York: Plenum.
- Uehling EA and Uhlenbeck GE (1933) Transport phenomena in Einstein–Bose and Fermi–Dirac gases. I. *Physical Reviews* 43: 552–561.
- Wigner EP (1932) On the quantum correction for thermodynamic equilibrium. *Physical Reviews* 40: 749–759.

Bose–Einstein Condensates

F Dalfovo, L P Pitaevskii, and S Stringari,
Università di Trento, Povo, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

In 1924 the Indian physicist S N Bose introduced a new statistical method to derive the blackbody radiation law in terms of a gas of light quanta (photons). His work, together with the contemporary de Broglie’s idea of matter–wave duality, led A Einstein to apply the same statistical approach to a gas of N indistinguishable particles of mass m . An amazing result of his theory was the prediction that below some critical temperature a finite fraction of all the particles condense into the lowest-energy single-particle state. This phenomenon, named Bose–Einstein condensation (BEC), is a consequence of purely statistical effects. For several years, such a prediction received little attention, until 1938, when F London argued that BEC could be at the basis of the superfluid properties observed in liquid ^4He below 2.17 K. A strong boost to the investigation of Bose–Einstein condensates was given in 1995 by the observation of BEC in dilute gases confined in magnetic traps and cooled down to temperatures of the order of a few nK. Differently from superfluid helium, these gases allow one to tune the relevant parameters (confining potential, particle density, interactions, etc.), so to make them an ideal test-ground for concepts and theories on BEC.

What Is BEC?

In nature, particles have either integer or half-integer spin. Those having half-integer spin, like electrons, are called fermions and obey the Fermi–Dirac statistics; those having integer spin are called bosons and obey the Bose–Einstein statistics. Let us consider a system of N bosons. In order to introduce the concept of BEC on a

general ground, one can start with the definition of the one-body density matrix

$$n^{(1)}(\mathbf{r}, \mathbf{r}') = \langle \hat{\Psi}^\dagger(\mathbf{r}) \hat{\Psi}(\mathbf{r}') \rangle \quad [1]$$

The quantities $\hat{\Psi}^\dagger(\mathbf{r})$ and $\hat{\Psi}(\mathbf{r})$ are the field operators which create and annihilate a particle at point \mathbf{r} , respectively; they satisfy the bosonic commutation relations

$$[\hat{\Psi}(\mathbf{r}), \hat{\Psi}^\dagger(\mathbf{r}')] = \delta(\mathbf{r} - \mathbf{r}'), \quad [\hat{\Psi}(\mathbf{r}), \hat{\Psi}(\mathbf{r}')] = 0 \quad [2]$$

If the system is in a pure state described by the N -body wave function $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$, then the average [1] is taken following the standard rules of quantum mechanics and the one-body density matrix can be written as

$$n^{(1)}(\mathbf{r}, \mathbf{r}') = N \int d\mathbf{r}_2 \dots d\mathbf{r}_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}', \mathbf{r}_2, \dots, \mathbf{r}_N) \quad [3]$$

involving the integration over the $N - 1$ variables $\mathbf{r}_2, \dots, \mathbf{r}_N$. In the more general case of a statistical mixture of pure states, expression [3] must be averaged according to the probability for a system to occupy the different states.

Since $n^{(1)}(\mathbf{r}, \mathbf{r}') = (n^{(1)}(\mathbf{r}', \mathbf{r}))^*$ the quantity $n^{(1)}$, when regarded as a matrix function of its indices \mathbf{r} and \mathbf{r}' , is Hermitian. It is therefore always possible to find a complete orthonormal basis of single-particle eigenfunctions, $\varphi_i(\mathbf{r})$, in terms of which the density matrix takes the diagonal form

$$n^{(1)}(\mathbf{r}, \mathbf{r}') = \sum_i n_i \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}') \quad [4]$$

The real eigenvalues n_i are subject to the normalization condition $\sum_i n_i = N$ and have the meaning of occupation numbers of the single-particle states φ_i . BEC occurs when one of these numbers (say, n_0) becomes macroscopic, that is, when $n_0 \equiv N_0$ is a number of order N , all the others remaining of order 1.

In this case eqn [4] can be conveniently rewritten in the form

$$n^{(1)}(\mathbf{r}, \mathbf{r}') = N_0 \varphi_0^*(\mathbf{r}) \varphi_0(\mathbf{r}') + \sum_{i \neq 0} n_i \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}') \quad [5]$$

and the state represented by $\varphi_0(\mathbf{r})$ is called Bose–Einstein condensate. This definition is rather general, since it applies to any macroscopic ($N \gg 1$) system of indistinguishable bosons independently of mutual interactions and external fields.

The one-body density matrix [1] contains information on important physical observables. By setting $\mathbf{r} = \mathbf{r}'$ one finds the diagonal density of the system

$$n(\mathbf{r}) \equiv n^{(1)}(\mathbf{r}, \mathbf{r}) = \langle \hat{\Psi}^\dagger(\mathbf{r}) \hat{\Psi}(\mathbf{r}) \rangle \quad [6]$$

with $N = \int d\mathbf{r} n(\mathbf{r})$. The off-diagonal components can instead be used to calculate the momentum distribution

$$n(\mathbf{p}) = \langle \hat{\Psi}^\dagger(\mathbf{p}) \hat{\Psi}(\mathbf{p}) \rangle \quad [7]$$

where $\hat{\Psi}(\mathbf{p}) = (2\pi\hbar)^{-3/2} \int d\mathbf{r} \hat{\Psi}(\mathbf{r}) \exp[-i\mathbf{p} \cdot \mathbf{r}/\hbar]$ is the field operator in momentum representation. By inserting this expression for $\hat{\Psi}(\mathbf{p})$ into eqn [7] one finds

$$n(\mathbf{p}) = \frac{1}{(2\pi\hbar)^3} \int d\mathbf{R} ds n^{(1)}\left(\mathbf{R} + \frac{\mathbf{s}}{2}, \mathbf{R} - \frac{\mathbf{s}}{2}\right) e^{-i\mathbf{p} \cdot \mathbf{s}/\hbar} \quad [8]$$

where $\mathbf{s} = \mathbf{r} - \mathbf{r}'$ and $\mathbf{R} = (\mathbf{r} + \mathbf{r}')/2$.

Let us consider a uniform system of N particles in a volume V and take the thermodynamic limit $N, V \rightarrow \infty$ with density N/V kept fixed. The eigenfunctions of the density matrix are plane waves and the lowest-energy state has zero momentum, $\mathbf{p} = 0$, and constant wave function $\varphi_0(\mathbf{r}) = V^{-1/2}$. BEC in this state implies a macroscopic number of particles having zero momentum and constant density N_0/V . The density matrix only depends on $\mathbf{s} = \mathbf{r} - \mathbf{r}'$ and can be written as

$$n^{(1)}(\mathbf{s}) = \frac{N_0}{V} + \frac{1}{V} \sum_{\mathbf{p} \neq 0} n_{\mathbf{p}} e^{-i\mathbf{p} \cdot \mathbf{s}/\hbar} \quad [9]$$

In the $s \rightarrow \infty$ limit, the sum on the right vanishes due to destructive interference between different plane waves, but the first term survives. One thus finds that, in the presence of BEC, the one-body density matrix tends to a constant finite value at large distances. This behavior is named *off-diagonal long-range order*, since it involves the off-diagonal components of the density matrix. Its counterpart in momentum space is the appearance of a singular term at $\mathbf{p} = 0$:

$$n(\mathbf{p}) = N_0 \delta(\mathbf{p}) + \sum_{\mathbf{p}' \neq 0} n_{\mathbf{p}'} \delta(\mathbf{p} - \mathbf{p}') \quad [10]$$

The sum on the right is the number of noncondensed particles ($N - N_0$), and the quantity N_0/N is called condensate fraction.

If the system is not uniform, the eigenfunctions of the density matrix are no longer plane waves but, provided N is sufficiently large, the concept of BEC is still well defined, being associated with the occurrence of a macroscopic occupation of a single-particle eigenfunction $\varphi_0(\mathbf{r})$ of the density matrix. Thus, the condensed bosons can be described by means of the function $\Psi(\mathbf{r}) = \sqrt{N_0} \varphi_0(\mathbf{r})$, which is a classical complex field playing the role of an *order parameter*. This is the analog of the classical limit of quantum electrodynamics, where the electromagnetic field replaces the microscopic description of photons. The function Ψ may also depend on time and can be written as

$$\Psi(\mathbf{r}, t) = |\Psi(\mathbf{r}, t)| e^{iS(\mathbf{r}, t)} \quad [11]$$

Its modulus determines the contribution of the condensate to the diagonal density [6], while the phase S is crucial in characterizing the coherence and superfluid properties of the system. The order parameter [11], also named *macroscopic wave function* or *condensate wave function*, is defined only up to a constant phase factor. One can always multiply this function by the numerical factor $e^{i\alpha}$ without changing any physical property. This reflects the gauge symmetry exhibited by all the physical equations of the problem. Making an explicit choice for the value of the order parameter, and hence for the phase, corresponds to a formal breaking of gauge symmetry.

BEC in Ideal Gases

Once we have defined what is a Bose–Einstein condensate, the next question is when such a condensation occurs in a given system. The ideal Bose gas provides the simplest example. So, let us consider a gas of noninteracting bosons described by the Hamiltonian $\hat{H} = \sum_i \hat{H}_i^{(1)}$, where the Schrödinger equation $\hat{H}_i^{(1)} \varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r})$ gives the spectrum of single-particle wave functions and energies. One can define an occupation number n_i as the number of particles in the state with energy ϵ_i . Thus, any given state of the many-body system is specified by a set $\{n_i\}$. The mean occupation numbers, \bar{n}_i , can be calculated by using the standard rules of statistical mechanics. For instance, by considering a grand canonical ensemble at temperature T , one finds

$$\bar{n}_i = \{\exp[\beta(\epsilon_i - \mu)] - 1\}^{-1} \quad [12]$$

with $\beta = 1/(k_B T)$. The chemical potential μ is fixed by the normalization condition $\sum_i \bar{n}_i = N$, where N is the average number of particles in the gas. For $T \rightarrow \infty$ the chemical potential is negative and large. It increases monotonically when T is lowered. Let us call ϵ_0 the lowest single-particle level in the spectrum. If at some critical temperature T_c the normalization condition can be satisfied with $\mu \rightarrow \epsilon_0^-$, then the occupation of the lowest state, $\bar{n}_0 = N_0$, becomes of order N and BEC is realized. Below T_c the normalization condition must be replaced with $N = N_0 + N_T$, where $N_T = \sum_{i \neq 0} \bar{n}_i$ is the number of particles out of the condensate, that is, the *thermal* component of the gas. Whether BEC occurs or not, and what is the value of T_c depends on the dimensionality of the system and the type of single-particle spectrum.

The simplest case is that of a gas confined in a cubic box of volume $V = L^3$ with periodic boundary conditions, where $\hat{H}^{(1)} = -(\hbar^2/2m)\nabla^2$. The eigenfunctions are plane waves $\varphi_{\mathbf{p}}(\mathbf{r}) = V^{-1/2} \exp[-i\mathbf{p} \cdot \mathbf{r}/\hbar]$, with energy $\epsilon_{\mathbf{p}} = p^2/2m$ and momentum $\mathbf{p} = 2\pi\hbar\mathbf{n}/L$. Here \mathbf{n} is a vector whose components n_x, n_y, n_z are 0 or \pm integers. The lowest eigenvalue has zero energy ($\epsilon_0 = 0$) and zero momentum. The mean occupation numbers are given by $\bar{n}_{\mathbf{p}} = \{\exp[\beta(p^2/2m - \mu)] - 1\}^{-1}$. In the thermodynamic limit ($N, V \rightarrow \infty$ with N/V kept constant), one can replace the sum $\sum_{\mathbf{p}}$ with the integral $\int d\epsilon \rho(\epsilon)$, where $\rho(\epsilon) = (2\pi)^{-2} V (2m/\hbar^2)^{3/2} \sqrt{\epsilon}$ is the density of states. In this way, one can calculate the thermal component of the gas as a function of T , finding the critical temperature

$$k_B T_c = \frac{2\pi\hbar^2}{m} \left(\frac{N}{V\zeta(3/2)} \right)^{2/3} \quad [13]$$

where ζ is the Riemann zeta function and $\zeta(3/2) \simeq 2.612$. For $T > T_c$, one has $\mu < 0$ and $N_T = N$. For $T < T_c$ one instead has $\mu = 0, N_T = N - N_0$ and

$$N_0(T) = N[1 - (T/T_c)^{3/2}] \quad [14]$$

The critical temperature turns out to be fully determined by the density N/V and by the mass of the constituents. These results were first obtained by A Einstein in his seminal paper and used by F London in the context of superfluid helium. We notice that the replacement of the sum with an integral in the above derivation is justified only if the thermal energy $k_B T$ is much larger than the energy spacing between single-particle levels, that is, if $k_B T \gg \hbar^2/2mV^{2/3}$. Is also worth noticing that the above expression for T_c can be written as $\lambda_T^3 N/V \simeq 2.612$, where $\lambda_T = [2\pi\hbar^2/(mk_B T)]^{1/2}$ is the thermal de Broglie wavelength. This is

equivalent to saying that BEC occurs when the mean distance between bosons is of the order of their de Broglie wavelength.

Another interesting case, which is relevant for the recent experiments with BEC in dilute gases confined in magnetic and/or optical traps, is that of an ideal gas subject to harmonic potentials. Let us consider, for simplicity, an isotropic external potential $V_{\text{ext}}(\mathbf{r}) = (1/2)m\omega_{\text{ho}}^2 r^2$. The single-particle Hamiltonian is $\hat{H}^{(1)} = -(\hbar^2/2m)\nabla^2 + V_{\text{ext}}(\mathbf{r})$ and its eigenvalues are $\epsilon_{n_x, n_y, n_z} = (n_x + n_y + n_z + 3/2)\hbar\omega_{\text{ho}}$. The corresponding density of states is $\rho(\epsilon) = (1/2)(\hbar\omega_{\text{ho}})^{-3} \epsilon^2$. A natural thermodynamic limit for this system is obtained by letting $N \rightarrow \infty$ and $\omega_{\text{ho}} \rightarrow 0$, while keeping the product $N\omega_{\text{ho}}^3$ constant. The condition for BEC to occur is that μ approaches the value $\epsilon_{000} = (3/2)\hbar\omega_{\text{ho}}$ from below by cooling the gas down to T_c . Following the same procedure as for the uniform gas, one finds

$$k_B T_c = \hbar\omega_{\text{ho}} [N/\zeta(3)]^{1/3} = 0.94\hbar\omega_{\text{ho}} N^{1/3} \quad [15]$$

and

$$N_0(T) = N[1 - (T/T_c)^3] \quad [16]$$

Notice that the condensate is not uniform in this case, since it corresponds to the lowest eigenfunction of the harmonic oscillator, which is a Gaussian of width $a_{\text{ho}} = [\hbar/(m\omega_{\text{ho}})]^{1/2}$. Correspondingly, the condensate in the momentum space is also a Gaussian, of width a_{ho}^{-1} . This implies that, differently from the gas in a box, here the condensate can be seen both in coordinate and momentum space in the form of a narrow distribution emerging from a wider thermal component. Finally, results [15] and [16] remain valid even for anisotropic harmonic potentials, with trapping frequencies ω_x, ω_y , and ω_z , provided the frequency ω_{ho} is replaced by the geometric average $(\omega_x \omega_y \omega_z)^{1/3}$.

BEC in Interacting Gases

Actual condensates are made of interacting particles. The full many-body Hamiltonian is

$$\hat{H} = \int d\mathbf{r} \hat{\Psi}^\dagger(\mathbf{r}) \hat{H}_0 \hat{\Psi}(\mathbf{r}) + \frac{1}{2} \int d\mathbf{r}' d\mathbf{r} \hat{\Psi}^\dagger(\mathbf{r}) \hat{\Psi}^\dagger(\mathbf{r}') V(\mathbf{r} - \mathbf{r}') \hat{\Psi}(\mathbf{r}') \hat{\Psi}(\mathbf{r}) \quad [17]$$

where $V(\mathbf{r} - \mathbf{r}')$ is the particle–particle interaction and $\hat{H}_0 = -(\hbar^2/2m)\nabla^2 + V_{\text{ext}}(\mathbf{r})$. Differently from the case of ideal gases, \hat{H} is no longer a sum of single-particle Hamiltonians. However, the general definitions given in the section “What is BEC?” are still valid. In particular, the one-body density matrix, in the presence of BEC, can be separated as in eqn [5]. One

can write $n^{(1)}(\mathbf{r}, \mathbf{r}') = \Psi^*(\mathbf{r})\Psi(\mathbf{r}') + \tilde{n}^{(1)}(\mathbf{r}, \mathbf{r}')$, where Ψ is the order parameter of the condensate ($\Psi^*(\mathbf{r})\Psi(\mathbf{r}')$ being of order N), while $\tilde{n}^{(1)}(\mathbf{r}, \mathbf{r}')$ vanishes for large $|\mathbf{r} - \mathbf{r}'|$. This is equivalent to say that the bosonic field operator splits in two parts,

$$\hat{\Psi}(\mathbf{r}) = \Psi(\mathbf{r}) + \delta\hat{\Psi}(\mathbf{r}) \quad [18]$$

where the first term is a complex function and the second one is the field operator associated with the noncondensed particles. This decomposition is particularly useful when the depletion of the condensate, that is, the fraction of noncondensed particles, is small. This happens when the interaction is weak, but also for particles with arbitrary interaction, provided the gas is dilute. In this case, one can expand the many-body Hamiltonian by treating the operator $\delta\hat{\Psi}$ as a small quantity.

A suitable strategy consists in writing the Heisenberg equation for the evolution of the field operators, $i\hbar\partial_t\hat{\Psi} = [\hat{\Psi}, \hat{H}]$, using the many-body Hamiltonian [17]:

$$\begin{aligned} i\hbar\partial_t\hat{\Psi}(\mathbf{r}, t) &= \left(\hat{H}_0 + \int d\mathbf{r}' \hat{\Psi}^\dagger(\mathbf{r}', t) V(\mathbf{r} - \mathbf{r}') \hat{\Psi}(\mathbf{r}', t) \right) \\ &\quad \times \hat{\Psi}(\mathbf{r}, t) \end{aligned} \quad [19]$$

The zeroth-order is thus obtained by replacing the operator $\hat{\Psi}$ with the classical field Ψ . In the integral containing the interaction $V(\mathbf{r} - \mathbf{r}')$, this replacement is, in general, a poor approximation when short distances ($\mathbf{r} - \mathbf{r}'$) are involved. In a dilute and cold gas, one can nevertheless obtain a proper expression for the interaction term by observing that, in this case, only binary collisions at low energy are relevant and these collisions are characterized by a single parameter, the s -wave scattering length, a , independently of the details of the two-body potential. This allows one to replace $V(\mathbf{r} - \mathbf{r}')$ in \hat{H} with an effective interaction $V(\mathbf{r} - \mathbf{r}') = g\delta(\mathbf{r} - \mathbf{r}')$, where the coupling constant g is given by $g = 4\pi\hbar^2 a/m$. The scattering length can be measured with several experimental techniques or calculated from the exact two-body potential. Using this pseudopotential and replacing the operator $\hat{\Psi}$ with the complex function Ψ in the Heisenberg equation of motion, one gets

$$\begin{aligned} i\hbar\partial_t\Psi(\mathbf{r}, t) &= \left(-\frac{\hbar^2\nabla^2}{2m} + V_{\text{ext}}(\mathbf{r}) + g|\Psi(\mathbf{r}, t)|^2 \right) \Psi(\mathbf{r}, t) \end{aligned} \quad [20]$$

This is known as Gross–Pitaevskii (GP) equation and it was first introduced in 1961. It has the form of a *nonlinear Schrödinger equation*, the nonlinearity coming from the mean-field term, proportional to

$|\Psi|^2$. It has been derived assuming that N is large while the fraction of noncondensed atoms is negligible. On the one hand, this means that quantum fluctuations of the field operator have to be small, which is true when $n|a|^3 \ll 1$, where n is the particle density. In fact, one can show that, at $T=0$ the quantum depletion of the condensate is proportional to $(n|a|^3)^{1/2}$. On the other hand, thermal fluctuations have also to be negligible and this means that the theory is limited to temperatures much lower than T_c . Within these limits, one can identify the total density with the condensate density.

The stationary solution of eqn [20] corresponds to the condensate wave function in the ground state. One can write $\Psi(\mathbf{r}, t) = \Psi_0(\mathbf{r}) \exp(-i\mu t/\hbar)$, where μ is the chemical potential. Then the GP equation [20] becomes

$$\left(-\frac{\hbar^2\nabla^2}{2m} + V_{\text{ext}}(\mathbf{r}) + g|\Psi_0(\mathbf{r})|^2 \right) \Psi_0(\mathbf{r}) = \mu\Psi_0(\mathbf{r}) \quad [21]$$

where $n(\mathbf{r}) = |\Psi_0(\mathbf{r})|^2$ is the particle density. The same equation can be obtained by minimizing the energy of the system written as a functional of the density:

$$E[n] = \int d\mathbf{r} \left[\frac{\hbar^2}{2m} |\nabla\sqrt{n}|^2 + nV_{\text{ext}}(\mathbf{r}) + \frac{gn^2}{2} \right] \quad [22]$$

The first term on the right corresponds to the quantum kinetic energy coming from the uncertainty principle; it is usually named “quantum pressure” and vanishes for uniform systems.

The next order in $\delta\hat{\Psi}$ gives the excited states of the condensate. In a uniform gas the ground-state order parameter, Ψ_0 , is a constant and the first-order expansion of \hat{H} was introduced by N Bogoliubov in 1947. In particular, he found an elegant way to diagonalize the Hamiltonian by using simple linear combinations of particle creation and annihilation operators. These are known as Bogoliubov’s transformations and stay at the basis of the concept of *quasiparticle*, one of the most important concepts in quantum many-body theory.

A generalization of Bogoliubov’s approach to the case of nonuniform condensates is obtained by considering small deviations around the ground state in the form

$$\Psi(\mathbf{r}, t) = e^{-i\mu t/\hbar} [\Psi_0(\mathbf{r}) + u(\mathbf{r})e^{-i\omega t} + v^*(\mathbf{r})e^{i\omega t}] \quad [23]$$

Inserting this expression into eqn [20] and keeping terms linear in the complex functions u and v , one gets

$$\hbar\omega u(\mathbf{r}) = [\hat{H}_0 - \mu + 2g\Psi_0^2(\mathbf{r})]u(\mathbf{r}) + g\Psi_0^2(\mathbf{r})v(\mathbf{r}) \quad [24]$$

$$-\hbar\omega v(\mathbf{r}) = [\hat{H}_0 - \mu + 2g\Psi_0^2(\mathbf{r})]v(\mathbf{r}) + g\Psi_0^2(\mathbf{r})u(\mathbf{r}) \quad [25]$$

These coupled equations allow one to calculate the energies $\varepsilon = \hbar\omega$ of the excitations. They also give the so-called quasiparticle amplitudes u and v , which obey the normalization condition

$$\int d\mathbf{r}[u_i^*(\mathbf{r})u_j(\mathbf{r}) - v_i^*(\mathbf{r})v_j(\mathbf{r})] = \delta_{ij}$$

In a uniform gas, u and v are plane waves and one recovers the famous Bogoliubov's spectrum

$$\hbar\omega = \left[\frac{\hbar^2 q^2}{2m} \left(\frac{\hbar^2 q^2}{2m} + 2gn \right) \right]^{1/2} \quad [26]$$

where q is the wave vector of the excitations. For large momenta the spectrum coincides with the free-particle energy $\hbar^2 q^2/2m$. At low momenta, it instead gives the phonon dispersion $\omega = cq$, where $c = [gn/m]^{1/2}$ is the Bogoliubov sound velocity. The transition between the two regimes occurs when the excitation wavelength is of the order of the *healing length*,

$$\xi = [8\pi na]^{-1/2} = \hbar/(mc\sqrt{2}) \quad [27]$$

which is an important length scale for superfluidity. When the order parameter is forced to vanish at some point (by an impurity, a wall, etc.), the healing length provides the typical distance over which it recovers its bulk value. In a nonuniform condensate the excitations are no longer plane waves but, at low energy, they have still a phonon-like character, in the sense that they involve a collective motion of the condensate.

The GP equation [20] is the starting point for an accurate mean-field description of BEC in dilute cold gases, which is rigorous at $T=0$ and for $n|a|^3 \ll 1$. Static and dynamics properties of condensates in different geometries can be calculated by solving the GP equation numerically or using suitable approximated methods. The inclusion of effects beyond mean field is a highly nontrivial and interesting problem. A rather extreme case is represented by liquid ^4He , which is a dense system where the interaction between atoms causes a large depletion of the condensate even at $T=0$ (N_0/N being less than 10%) and thus a full many-body treatment is required for its rigorous description. Nevertheless, even in this case, the general definitions of the section “What is BEC?” are still useful.

Superfluidity and Coherence

With the word superfluidity, one summarizes a complex of macroscopic phenomena occurring in quantum fluids under particular conditions: persistent currents, equilibrium states at rest in rotating

vessels, viscousless motion, quantized vorticity, and others. These features can also be observed in BEC. The link between BEC and superfluidity is given by the phase of the order parameter [11]. To understand this point, let us consider a uniform system. If $\hat{\Psi}(\mathbf{r}, t)$ is a solution of the Heisenberg equation [19] with $V_{\text{ext}} = 0$, then

$$\hat{\Psi}'(\mathbf{r}, t) = \hat{\Psi}(\mathbf{r} - \mathbf{v}t, t) \exp \left[\frac{i}{\hbar} \left(m\mathbf{v} \cdot \mathbf{r} - \frac{1}{2} m\mathbf{v}^2 t \right) \right] \quad [28]$$

where \mathbf{v} is a constant vector, is also a solution. This equation gives the Galilean transformation of the field operator and also applies to its condensate component Ψ . At equilibrium, the ground-state order parameter is given by $\Psi_0 = \sqrt{n} \exp(-i\mu t/\hbar)$, where n is a constant independent of \mathbf{r} . In a frame where the condensate moves with velocity \mathbf{v} , the order parameter instead takes the form $\Psi_0 = \sqrt{n} \exp(iS)$, with $S(\mathbf{r}, t) = \hbar^{-1} [m\mathbf{v} \cdot \mathbf{r} - (m\mathbf{v}^2/2 + \mu)t]$. The velocity of the condensate can thus be identified with the gradient of the phase S :

$$\mathbf{v}(\mathbf{r}, t) = \frac{\hbar}{m} \nabla S(\mathbf{r}, t) \quad [29]$$

This definition is also valid for \mathbf{v} varying slowly in space and time. The modulus of the order parameter plays a minor role in this definition and it is not necessary to assume the gas to be dilute and close to $T=0$. Indeed, the relation [29] between the velocity field and the phase of the order parameter also applies in the presence of large quantum depletion, as in superfluid ^4He , and at $T \neq 0$. In this case, n should not be identified with the condensate density. Conversely, in dilute gases at $T=0$, n is the condensate density and the velocity [29] can be simply obtained by applying the usual definition of current density operator, \hat{j} , to the order parameter [11].

The velocity [29] describes a potential flow and corresponds to a collective motion of many particles occupying a single quantum state. Being equal to the gradient of a scalar function, it is irrotational ($\nabla \times \mathbf{v}_s = 0$) and satisfies the Onsager–Feynman quantization condition $\oint \mathbf{v}_s \cdot d\mathbf{l} = \kappa \hbar/m$, with κ non-negative integer. These conditions are not satisfied by a classical fluid, where the hydrodynamic velocity field, $\mathbf{v}(\mathbf{r}, t) = \mathbf{j}(\mathbf{r}, t)/n(\mathbf{r}, t)$, is the average over many different states and does not correspond to a potential flow.

By using the definition of the phase S and velocity \mathbf{v} , together with particle conservation, one can show that the dynamics of a condensate, as far as macroscopic motions are concerned, is governed by the hydrodynamic equations of an irrotational

nonviscous fluid. Within the mean-field theory, this can be easily seen by rewriting the GP equation [20] in terms of the density $n = |\Psi|^2$ and the velocity [29]. Neglecting the quantum pressure term $\nabla^2 \sqrt{n}$ (hence limiting the description to length scales larger than the healing length ξ), one gets

$$\frac{\partial}{\partial t} n + \nabla \cdot (vn) = 0 \quad [30]$$

and

$$m \frac{\partial}{\partial t} v + \nabla \left(V_{\text{ext}} + \mu(n) + \frac{mv^2}{2} \right) = 0 \quad [31]$$

with the local chemical potential $\mu(n) = gn$. These equations have the typical structure of the dynamic equations of superfluids at zero temperature and can be viewed as the $T=0$ case of the more general Landau's two-fluid theory.

One of the most striking evidences of superfluidity is the observation of quantized vortices, that is, vortices obeying the Onsager–Feynman quantization condition. A vast literature is devoted to vortices in superfluid helium and, more recently, vortices have also been produced and studied in condensates of ultracold gases, including nice configurations of many vortices in regular triangular lattices, similar to the Abrikosov lattices in superconductors. Other phenomena, such as the reduction of the moment of inertia, the occurrence of Josephson tunneling through barriers, the existence of thresholds for dissipative processes (Landau criterion), and others, are typical subjects of intense investigation.

Another important consequence of the fact that BEC is described by an order parameter with a well-defined phase is the occurrence of coherence effects which, in different words, mean that condensates behave like *matter waves*. For instance, one can measure the phase difference between two condensates by means of interference. This can be done in coordinate space by confining two condensates in two potential minima, a and b , at a distance d . Let us take d along z and assume that, at $t=0$, the order parameter is given by the linear combination $\Psi(\mathbf{r}) = \Psi_a(\mathbf{r}) + \exp(i\phi)\Psi_b(\mathbf{r})$ with Ψ_a and Ψ_b real and without overlap. Then let us switch off the confining potentials so that the condensates expand and overlap. If the overlap occurs when the density is small enough to neglect interactions, the motion is ballistic and the phase of each condensate evolves as $S(\mathbf{r}, t) \simeq mr^2/(2\hbar t)$, so that $\mathbf{v} = \mathbf{r}/t$. This implies a relative phase $\phi + S(x, y, z + d/2) - S(x, y, z - d/2) = \phi + mdz/\hbar t$. The total density $n = |\Psi|^2$ thus exhibits periodic modulations along z with wavelength $\hbar t/md$. This interference pattern has indeed

been observed in condensates of ultracold atoms. In these systems it was also possible to measure the coherence length, that is, the distance $|\mathbf{r} - \mathbf{r}'|$ at which the one-body density vanishes and the phase of the order parameter is no more well defined. In most situations, the coherence length turns out to be of the order of, or larger than the size of the condensates. However, interesting situations exist when the coherence length is shorter but the system still preserves some features of BEC (quasicondensates).

Final Remarks

Bose–Einstein condensates of ultracold atoms are easily manipulated by changing and tuning the external potentials. This means, for instance, that one can prepare condensates in different geometries, including very elongated (quasi-1D) or disk-shaped (quasi-2D) condensates. This is conceptually important, since BEC in lower dimensions is not as simple as in three dimensions: thermal and quantum fluctuations play a crucial role, superfluidity must be properly re-defined, and very interesting limiting cases can be explored (Tonks–Girardeau regime, Luttinger liquid, etc.). Another possibility is to use laser beams to produce standing waves acting as an external periodic potential (optical lattice). Condensates in optical lattices behave as a sort of perfect crystal, whose properties are the analog of the dynamic and transport properties in solid-state physics, but with controllable spacing between sites, no defects and tunable lattice geometry. One can investigate the role of phase coherence in the lattice, looking, for instance, at Josephson effects as in a chain of junctions. By tuning the lattice depth one can explore the transition from a superfluid phase and a Mott-insulator phase, which is a nice example of quantum phase transition. Controlling cold atoms in optical lattice can be a good starting point for application in quantum engineering, interferometry, and quantum information.

Another interesting aspect of BECs is that the key equation for their description in mean-field theory, namely the GP equation [20], is a nonlinear Schrödinger equation very similar to the ones commonly used, for instance, in nonlinear quantum optics. This opens interesting perspectives in exploiting the analogies between the two fields, such as the occurrence of dynamical and parametric instabilities, the possibility to create different types of solitons, the occurrence of nonlinear processes like, for example, higher harmonic generation and mode mixing.

A relevant part of the current research also involves systems made of mixtures of different gases, Bose–Bose or Fermi–Bose, and many activities with ultracold atoms now involve fermionic gases, where BEC can

also be realized by condensing molecules of fermionic pairs. An extremely active research now concerns the BCS–BEC crossover, which can be obtained in Fermi gases by tuning the scattering length (and hence the interaction) by means of Feshbach resonances.

Ten years after the first observation of BEC in ultracold gases, it is almost impossible to summarize all the researches done in this field. A large amount of work has already been devoted to characterize the condensates and several new lines have been opened. Rather detailed review articles and books are already available for the interested readers.

See also: Interacting Particle Systems and Hydrodynamic Equations; Quantum Phase Transitions; Quantum Statistical Mechanics: Overview; Renormalization: Statistical Mechanics and Condensed Matter; Superfluids; Variational Techniques for Ginzburg–Landau Energies.

Further Reading

Cornell EA and Wieman CE (2002) Nobel lecture: Bose–Einstein condensation in a dilute gas, the first 70 years and some recent experiments. *Reviews of Modern Physics* 74: 875.

Bosons and Fermions in External Fields

E Langmann, KTH Physics, Stockholm, Sweden

© 2006 Elsevier Ltd. All rights reserved.

Introduction

In this article we discuss quantum theories which describe systems of nondistinguishable particles interacting with external fields. Such models are of interest also in the nonrelativistic case (in quantum statistical mechanics, nuclear physics, etc.), but the relativistic case has additional, interesting complications: relativistic models are genuine quantum field theories, that is, quantum theories with an infinite number of degrees of freedom, with nontrivial features like divergences and anomalies. Since interparticle interactions are ignored, such models can be regarded as a first approximation to more complicated theories, and they can be studied by mathematically precise methods.

Models of relativistic particles in external electromagnetic fields have received considerable attention in the physics literature, and interesting phenomena like the Klein paradox or particle–antiparticle pair creation in overcritical fields have been studied; see [Rafelski et al. \(1978\)](#) for an extensive review. We will not discuss these physics questions but only

- Dalfovo F, Giorgini S, Pitaevskii LP, and Stringari S (1999) Theory of Bose–Einstein condensation in trapped gases. *Reviews of Modern Physics* 71: 463.
- Griffin A, Snoko DW, and Stringari S (1995) *Bose–Einstein Condensation*. Cambridge: Cambridge University Press.
- Huang K (1987) *Statistical Mechanics*, 2nd edn. New York: Wiley.
- Inguscio M, Stringari S, and Wieman CE (1999) *Bose–Einstein Condensation in Atomic Gases*, Proceedings of the International School of Physics “Enrico Fermi,” Course CXL. Amsterdam: IOS Press.
- Ketterle W (2002) Nobel lecture: when atoms behave as waves: Bose–Einstein condensation and the atom laser. *Reviews of Modern Physics* 74: 1131.
- Landau LD and Lifshitz EM (1980) *Statistical Physics*, Part 1. Oxford: Pergamon Press.
- Leggett AJ (2001) Bose–Einstein condensation in the alkali gases: some fundamental concepts. *Reviews of Modern Physics* 73: 307.
- Lifshitz EM and Pitaevskii LP (1980) *Statistical Physics*, Part 2. Oxford: Pergamon Press.
- Pethick CJ and Smith H (2002) *Bose–Einstein Condensation in Dilute Gases*. Cambridge: Cambridge University Press.
- Pitaevskii LP and Stringari S (2003) *Bose–Einstein Condensation*. Oxford: Clarendon Press.

describe some prototype examples and a general Hamiltonian framework which has been used in mathematically precise work on such models. The general framework for this latter work is the mathematical theory of Hilbert space operators (see, e.g., [Reed and Simon \(1975\)](#)), but in our discussion we try to avoid presupposing knowledge of that theory. As mentioned briefly in the end, this work has had close relations to various topics of recent interest in mathematical physics, including anomalies, infinite-dimensional geometry and group theory, conformal field theory, and noncommutative geometry.

We restrict our discussion to spin-0 bosons and spin-1/2 fermions, and we will not discuss models of particles in external gravitational fields but only refer the interested reader to [DeWitt \(2003\)](#). We also only mention in passing that external field problems have also been studied using functional integral approaches, and mathematically precise work on this can be found in the extensive literature on determinants of differential operators.

Examples

Consider the Schrödinger equation describing a nonrelativistic particle of mass m and charge e

moving in three-dimensional space and interacting with an external vector and scalar potentials \mathbf{A} and ϕ , respectively,

$$i\partial_t\psi = H\psi, \quad H = \frac{1}{2m}(-i\nabla + e\mathbf{A})^2 - e\phi \quad [1]$$

(we set $\hbar = c = 1$, $\partial_t = \partial/\partial t$, and ψ, ϕ , and \mathbf{A} can depend on the space and time variables $\mathbf{x} \in \mathbb{R}^3$ and $t \in \mathbb{R}$). This is a standard quantum-mechanical model, with ψ the one-particle wave function allowing for the usual probabilistic interpretation.

One interesting generalization to the relativistic regime is the Klein–Gordon equation

$$\left[(i\partial_t + e\phi)^2 - (-i\nabla + e\mathbf{A})^2 - m^2\right]\psi = 0 \quad [2]$$

with a \mathbb{C} -valued function ψ . There is another important relativistic generalization, the Dirac equation

$$[(i\partial_t + e\phi) - (-i\nabla + e\mathbf{A}) \cdot \boldsymbol{\alpha} + m\beta]\psi = 0 \quad [3]$$

with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and β Hermitian 4×4 matrices satisfying the relations

$$\alpha_i\alpha_j + \alpha_j\alpha_i = \delta_{ij}, \quad \alpha_i\beta = -\beta\alpha_i, \quad \beta^2 = 1 \quad [4]$$

and a \mathbb{C}^4 -valued function ψ (we also write 1 for the identity). These two relativistic equations differ by the transformation properties of ψ under Lorentz transformations: in [2] it transforms like a scalar and thus describes spin-0 particles, and it transforms like a spinor describing spin-1/2 particles in [3]. While these equations are natural relativistic generalizations of the Schrödinger equation, they no longer allow to consistently interpret ψ as one-particle wave functions. The physical reason is that, in a relativistic theory, high-energy processes can create particle–antiparticle pairs, and this makes the restriction to a fixed particle number inconsistent. This problem can be remedied by constructing a many-body model allowing for an arbitrary number of particles and antiparticles. The requirement that this many-body model should have a ground state is an important ingredient in this construction.

It is obviously of interest to formulate and study many-body models of nondistinguishable particles already in the nonrelativistic case. An important empirical fact is that such particles come in two kinds, bosons and fermions, distinguished by their exchange statistics (we ignore the interesting possibility of exotic statistics). For example, the fermion many-particle version of [1] for suitable ϕ and \mathbf{A} is a useful model for electrons in a metal. An elegant method to go from the one- to the many-particle description is the formalism of second quantization: one promotes ψ to a quantum field operator with

certain (anti-) commutator relations, and this is a convenient way to construct the appropriate many-particle Hilbert space, Hamiltonian, etc. In the nonrelativistic case, this formalism can be regarded as an elegant reformulation of a pedestrian construction of a many-body quantum-mechanical model, which is useful since it provides convenient computational tools. However, this formalism naturally generalizes to the relativistic case where the one-particle model no longer has an acceptable physical interpretation, and one finds that one can nevertheless give a consistent physical interpretation to [2] and [3] provided that ψ are interpreted as quantum field operators describing bosons and fermions. This particular exchange statistics of the relativistic particles is a special case of the spin-statistics theorem: *integer-spin particles are bosons and half-integer spin particles are fermions*. While many structural features of this formalism are present already in the simpler nonrelativistic models, the relativistic models add some nontrivial features typical for quantum field theories.

In the following, we discuss a precise mathematical formulation of the quantum field theory models described above. We emphasize the functorial nature of this construction, which makes manifest that it also applies to other situations, for example, where the bosons and fermions are also coupled to a gravitational background, are considered in other spacetime dimensions than $3 + 1$, etc.

Second Quantization: Nonrelativistic Case

Consider a quantum system of nondistinguishable particles where the quantum-mechanical description of one such particle is known. In general, this one-particle description is given by a Hilbert space h and one-particle observables and transformations which are self-adjoint and unitary operators on h , respectively. The most important observable is the Hamiltonian H . We will describe a general construction of the corresponding many-body system.

Example As a motivating example we take the Hilbert space $h = L^2(\mathbb{R}^3)$ of square-integrable functions $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^3$, and the Hamiltonian H in [1]. A specific example for a unitary operator on h is the gauge transformation $(Uf)(\mathbf{x}) = \exp(i\chi(\mathbf{x}))f(\mathbf{x})$ with χ a smooth, real-valued function on \mathbb{R}^3 .

In this example, the corresponding wave functions for N identical such particles are the L^2 -functions $f_N(\mathbf{x}_1, \dots, \mathbf{x}_N)$, $\mathbf{x}_j \in \mathbb{R}^3$. It is obvious how to extend

one-particle observables and transformations to such N -particle states: for example, the N -particle Hamiltonian corresponding to H in [1] is

$$H_N = \sum_{j=1}^N \frac{1}{2m} (-i\nabla_{\mathbf{x}_j} + e\mathbf{A}(t, \mathbf{x}_j))^2 - e\phi(t, \mathbf{x}_j) \quad [5]$$

and the N -particle gauge transformation U_N is defined through multiplication with $\prod_{j=1}^N \exp(i\chi(\mathbf{x}_j))$.

For systems of indistinguishable particles it is enough to restrict to wave functions which are even or odd under particle exchanges,

$$\begin{aligned} f_N(\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_k, \dots, \mathbf{x}_N) \\ = \pm f_N(\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_j, \dots, \mathbf{x}_N) \end{aligned} \quad [6]$$

for all $1 \leq j < k \leq N$, with the upper and lower signs corresponding to bosons and fermions, respectively (this empirical fact is usually taken as a postulate in nonrelativistic many-body quantum physics). It is convenient to define the zero-particle Hilbert space as \mathbb{C} (complex numbers) and to introduce a Hilbert space containing states with all possible particle numbers: this so-called Fock space contains all states

$$\begin{pmatrix} f_0 \\ f_1(\mathbf{x}_1) \\ f_2(\mathbf{x}_1, \mathbf{x}_2) \\ f_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \\ \vdots \end{pmatrix} \quad [7]$$

with $f_0 \in \mathbb{C}$. The definition of H_N and U_N then naturally extends to this Fock space; see below.

General Construction

The construction of Fock spaces and many-particle observables and transformations just outlined in a specific example is conceptually simple. An alternative, more efficient construction method is to use “quantum fields,” which we denote as $\psi(\mathbf{x})$ and $\psi^\dagger(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^3$. They can be fully characterized by the following (anti-) commutator relations:

$$[\psi(\mathbf{x}), \psi^\dagger(\mathbf{y})]_{\mp} = \delta^3(\mathbf{x} - \mathbf{y}), \quad [\psi(\mathbf{x}), \psi(\mathbf{y})]_{\mp} = 0 \quad [8]$$

where $[a, b]_{\mp} \equiv ab \mp ba$, with the commutator and anticommutators (upper and lower signs, respectively) corresponding to the boson and fermion case, respectively. It is convenient to “smear” these fields with one-particle wave functions and define

$$\begin{aligned} \psi(f) &= \int_{\mathbb{R}^3} d^3x \overline{f(\mathbf{x})} \psi(\mathbf{x}) \\ \psi^\dagger(f) &= \int_{\mathbb{R}^3} d^3x \psi^\dagger(\mathbf{x}) f(\mathbf{x}) \end{aligned} \quad [9]$$

for all $f \in \mathcal{h}$. Then the relations characterizing the field operators can be written as

$$\begin{aligned} [\psi(f), \psi^\dagger(g)]_{\mp} &= (f, g) \\ [\psi(f), \psi(g)]_{\mp} &= 0 \\ \forall f, g \in \mathcal{h} \end{aligned} \quad [10]$$

where

$$(f, g) = \int_{\mathbb{R}^3} d^3x \overline{f(\mathbf{x})} g(\mathbf{x})$$

is the inner product in \mathcal{h} . The Fock space $\mathcal{F}_{\mp}(\mathcal{h})$ can then be defined by postulating that it contains a normalized vector Ω called “vacuum” such that

$$\psi(f)\Omega = 0 \quad \forall f \in \mathcal{h} \quad [11]$$

and that all $\psi^{(\dagger)}(f)$ are operators on $\mathcal{F}_{\mp}(\mathcal{h})$ such that $\psi^\dagger(f) = \psi(f)^*$, where $*$ is the Hilbert space adjoint. Indeed, from this we conclude that $\mathcal{F}_{\mp}(\mathcal{h})$, as vector space, is generated by

$$f_1 \wedge f_2 \wedge \dots \wedge f_N \equiv \psi^\dagger(f_1) \psi^\dagger(f_2) \dots \psi^\dagger(f_N) \Omega \quad [12]$$

with $f_j \in \mathcal{h}$ and $N = 0, 1, 2, \dots$, and that the Hilbert space inner product of such vectors is

$$\begin{aligned} \langle f_1 \wedge f_2 \wedge \dots \wedge f_N, g_1 \wedge g_2 \wedge \dots \wedge g_M \rangle \\ = \delta_{N,M} \sum_{P \in S_N} (\pm 1)^{|P|} \prod_{j=1}^N (f_j, g_{P_j}) \end{aligned} \quad [13]$$

with S_N the permutation group, with $(+1)^{|P|} = 1$ always, and $(-1)^{|P|} = +1$ and -1 for even and odd permutations, respectively. The many-body Hamiltonian $q(H)$ corresponding to the one-particle Hamiltonian H can now be defined by the following relations:

$$q(H)\Omega = 0, \quad [q(H), \psi^\dagger(f)] = \psi^\dagger(Hf) \quad [14]$$

for all $f \in \mathcal{h}$ such that Hf is defined. Indeed, this implies that

$$\begin{aligned} q(H)f_1 \wedge f_2 \wedge \dots \wedge f_N \\ = \sum_{j=1}^N f_1 \wedge f_2 \wedge \dots \wedge (Hf_j) \wedge \dots \wedge f_N \end{aligned} \quad [15]$$

which defines a self-adjoint operator on $\mathcal{F}_{\mp}(\mathcal{h})$, and it is easy to check that this coincides with our down-to-earth definition of H_N above. Similarly, the many-body transformation $Q(U)$ corresponding to a one-particle transformation U can be defined as

$$Q(U)\Omega = \Omega, \quad Q(U)\psi^\dagger(f) = \psi^\dagger(Uf)Q(U) \quad [16]$$

for all $f \in \mathcal{h}$, which implies that

$$\begin{aligned} Q(U)f_1 \wedge f_2 \wedge \dots \wedge f_N \\ = (Uf_1) \wedge (Uf_2) \wedge \dots \wedge (Uf_N) \end{aligned} \quad [17]$$

and thus coincides with our previous definition of U_N .

While we presented the construction above for a particular example, it is important to note that it actually does not make reference to what the one-particle formalism actually is. For example, if we had a model of particles on a space \mathcal{M} given by some “nice” manifold of any dimension and with M internal degrees of freedom, we would take $h = L^2(\mathcal{M}) \otimes \mathbb{C}^M$ and replace [9] by

$$\psi(f) = \int_{\mathcal{M}} d\mu(\mathbf{x}) \sum_{j=1}^M \overline{f_j(\mathbf{x})} \psi_j(\mathbf{x}) \quad [18]$$

and its Hermitian conjugate, with the measure μ on \mathcal{M} defining the inner product in h ,

$$(f, g) = \int d\mu(\mathbf{x}) \sum_j \overline{f_j(\mathbf{x})} g_j(\mathbf{x})$$

With that, all formulas after [9] hold true as they stand. Given any one-particle Hilbert space h with inner product (\cdot, \cdot) , observable H , and transformation U , the formulas above define the corresponding Fock spaces $\mathcal{F}_{\mp}(h)$ and many-body observable $q(H)$ and transformation $Q(U)$. It is also interesting to note that this construction has various beautiful general (functorial) properties: the set of one-particle observables has a natural Lie algebra structure with the Lie bracket given by the commutator (strictly speaking: i times the commutator, but we drop the common factor i for simplicity). The definitions above imply that

$$[q(A), q(B)] = q([A, B]) \quad [19]$$

for one-particle observables A, B , that is, the above-mentioned Lie algebra structure is preserved under this map q . In a similar manner, the set of one-particle transformations has a natural group structure preserved by the map Q ,

$$Q(U)Q(V) = Q(UV), \quad Q(U)^{-1} = Q(U^{-1}) \quad [20]$$

Moreover, if A is self-adjoint, then $\exp(iA)$ is unitary, and one can show that

$$Q(\exp(iA)) = \exp(iq(A)) \quad [21]$$

For later use, we note that, if $\{f_n\}_{n \in \mathbb{Z}}$ is some complete, orthonormal basis in h , then operators A on h can be represented by infinite matrices $(A_{mn})_{m, n \in \mathbb{Z}}$ with $A_{mn} = (f_m, Af_n)$, and

$$q(A) = \sum_{m, n} A_{mn} \psi_m^\dagger \psi_n \quad [22]$$

where $\psi_n^{(\dagger)} = \psi^{(\dagger)}(f_n)$ obey

$$[\psi_m, \psi_n^\dagger]_{\mp} = \delta_{m, n}, \quad [\psi_m, \psi_n^\dagger]_{\mp} = 0 \quad [23]$$

for all m, n . We also note that, in our definition of $q(A)$, we made a convenient choice of normalization, but there is no physical reason to not choose a different normalization and define

$$q'(A) = q(A) - b(A) \quad [24]$$

where b is some linear function mapping self-adjoint operators A to real numbers. For example, one may wish to use another reference vector $\tilde{\Omega}$ instead of Ω in the Fock space, and then would choose $b(A) = \langle \tilde{\Omega}, q(A)\tilde{\Omega} \rangle$. Then the relations in [19] are changed to

$$[q'(A), q'(B)] = q'([A, B]) + S_0(A, B) \quad [25]$$

where $S_0(A, B) = b([A, B])$. However, the \mathbb{C} -number term $S_0(A, B)$ in the relations [25] is trivial, since it can be removed by going back to $q(A)$.

Physical Interpretation

The Fock space $\mathcal{F}_{\mp}(h)$ is the direct sum of subspaces of states with different particle numbers N ,

$$\mathcal{F}_{\mp}(h) = \bigoplus_{N=0}^{\infty} h_{\mp}^{(N)} \quad [26]$$

where the zero-particle subspace $h_{\mp}^{(0)} = \mathbb{C}$ is generated by the vacuum Ω , and $h_{\mp}^{(N)}$ is the N -particle subspace generated by the states $f_1 \wedge f_2 \wedge \cdots \wedge f_N$, $f_j \in h$. We note that

$$\mathcal{N} \equiv q(1) \quad [27]$$

is the “particle-number operator,” $\mathcal{N}F_N = NF_N$ for all $F_N \in h_{\mp}^{(N)}$. The field operators obviously change the particle number: $\psi^\dagger(f)$ increases the particle number by one (maps $h_{\mp}^{(N)}$ to $h_{\mp}^{(N+1)}$), and $\psi(f)$ decreases it by one. Since every $f \in h$ can be interpreted as one-particle state, it is natural to interpret $\psi^\dagger(f)$ and $\psi(f)$ as “creation” and “annihilation” operators, respectively: they create and annihilate one particle in the state $f \in h$. It is important to note that, in the fermion case, [10] implies that $\psi^\dagger(f)^2 = 0$, which is a mathematical formulation of the Pauli exclusion principle: *it is not possible to have two fermions in the same one-particle state*. In the boson case, there is no such restriction. Thus, even though the formalisms used to describe boson and fermion systems look very similar, they describe dramatically different physics.

Applications

In our example, the many-body Hamiltonian $\mathcal{H}_0 \equiv q(H)$ can also be written in the following suggestive form:

$$\mathcal{H}_0 = \int d^3x \psi^\dagger(\mathbf{x})(H\psi)(\mathbf{x}) \quad [28]$$

and similar formulas hold true for other observables and other Hilbert spaces $h = L^2(\mathcal{M}) \otimes \mathbb{C}^n$. It is rather easy to solve the model defined by such Hamiltonian: all necessary computations can be reduced to one-particle computations. For example, in the static case, where A and ϕ are time independent, a main quantity of interest in statistical physics is the free energy

$$\mathcal{E} \equiv -\beta^{-1} \log(\text{tr}(\exp(-\beta[\mathcal{H}_0 - \mu\mathcal{N}]))) \quad [29]$$

where $\beta > 0$ is the inverse temperature, μ the chemical potential, and the trace over the Fock space $\mathcal{F}_{\mp}(h)$. One can show that

$$\mathcal{E} = \pm \text{tr}(\beta^{-1} \log(1 \mp \exp(-\beta[H - \mu]))) \quad [30]$$

where the trace is over the one-particle Hilbert space h . Thus, to compute \mathcal{E} , one only needs to find the eigenvalues of H .

It is important to mention that the framework discussed here is not only for external field problems but can be equally well used to formulate and study more complicated models with interparticle interactions. For example, while the model with the Hamiltonian \mathcal{H}_0 above is often too simple to describe systems in nature, it is easy to write down more realistic models, for example, the Hamiltonian

$$\begin{aligned} \mathcal{H} = \mathcal{H}_0 + (e^2/2) \int d^3x \int d^3y \psi^\dagger(\mathbf{x})\psi^\dagger(\mathbf{y}) \\ \times |\mathbf{x} - \mathbf{y}|^{-1} \psi(\mathbf{y})\psi(\mathbf{x}) \end{aligned} \quad [31]$$

describes electrons in an external electromagnetic field interacting through Coulomb interactions. This illustrates an important point which we would like to stress: the task in quantum theory is twofold, namely to formulate and to solve (exact or otherwise) models. Obviously, in the nonrelativistic case, it is equally simple to formulate many-body models with and without interparticle interactions, and only the latter are simpler because they are easier to solve: the two tasks of formulating and solving models can be clearly separated. As we will see, in the relativistic case, even the formulation of an external field problem is nontrivial, and one finds that one cannot formulate the model without at least partially solving it. This is a common feature of quantum field theories making them challenging and interesting.

Relativistic Fermion and Boson Systems

We now generalize the formalism developed in the previous section to the relativistic case.

Field Algebras and Quasifree Representations

In the previous section, we identified the field operators $\psi^{(\dagger)}(f)$ with particular Fock space operators. This is analogous to identifying the operators $p_j = -i\partial_{x_j}$ and $q_j = x_j$ on $L^2(\mathbb{R}^M)$ with the generators of the Heisenberg algebra, as usually done. (We recall: the Heisenberg algebra is the star algebra generated by P_j and Q_j , $j=1,2,\dots,M < \infty$, with the well-known relations

$$\begin{aligned} [P_j, P_k] = -i\delta_{jk}, \quad [P_j, P_k] = [P_j, Q_k] = 0 \\ P_j^\dagger = P_j, \quad Q_j^\dagger = Q_j \end{aligned} \quad [32]$$

for all j, k .) Identifying the Heisenberg algebra with a particular representation is legitimate since, as is well known, all its irreducible representations are (essentially) the same (this statement is made precise by a celebrated theorem due to von Neumann).

However, in case of the algebra generated by the field operators $\psi^{(\dagger)}(f)$, there exist representations which are truly different from the ones discussed in the last section, and such representations are needed to construct relativistic external field problems. It is therefore important to distinguish the fields as generators of an algebra from the operators representing them. We thus define the (boson or fermion) field algebra $\mathcal{A}_{\mp}(h)$ over a Hilbert space h as the star algebra generated by $\Psi^\dagger(f)$, $f \in h$, such that the map $f \rightarrow \Psi(f)$ is linear and the relations

$$\begin{aligned} [\Psi(f), \Psi^\dagger(g)]_{\mp} = (f, g) \\ [\Psi(f), \Psi(g)]_{\mp} = 0 \\ \Psi^\dagger(f)^\dagger = \Psi(f) \end{aligned} \quad [33]$$

are fulfilled for all $f, g \in h$, with \dagger the star operation in $\mathcal{A}_{\mp}(h)$. The particular representation of this algebra discussed in the last section will be denoted by π_0 , $\pi_0(\Psi^{(\dagger)}(f)) = \psi^{(\dagger)}(f)$. Other representations π_{P_-} can be constructed from any projection operators P_- on h , that is, any operator P_- on h satisfying $P_-^* = P_-^2 = P_-$. Writing $\hat{\psi}^{(\dagger)}(f)$ short for $\pi_{P_-}(\Psi^{(\dagger)}(f))$, this so-called quasifree representation is defined by

$$\begin{aligned} \hat{\psi}^\dagger(f) = \psi^\dagger(P_+f) + \psi(\overline{P_-f}) \\ \hat{\psi}(f) = \psi(P_+f) \mp \psi^\dagger(\overline{P_-f}) \end{aligned} \quad [34]$$

where the bar means complex conjugation. It is important to note that, while the star operation is identical with the Hilbert space adjoint $*$ in the fermion case, we have

$$\begin{aligned} \hat{\psi}(f)^\dagger = \psi(Ff)^* \quad \text{with} \\ F = P_+ - P_- \quad \text{for bosons} \end{aligned} \quad [35]$$

where F is a grading operator, that is, $F^* = F$ and $F^2 = 1$. We stress that the “physical” star operation always is $*$, that is, physical observables A obey $A = A^*$.

The present framework suggests to regard quantization as the procedure which amounts to going from a one-particle Hilbert space h to the corresponding field algebra $\mathcal{A}_+(h)$. Indeed, the Heisenberg algebra is identical with the boson field algebra $\mathcal{A}_-(\mathbb{C}^M)$ (since the latter is obviously identical with the algebra of M harmonic oscillators), and thus conventional quantum mechanics can be regarded as boson quantization in the special case where the one-particle Hilbert space is finite dimensional. It is interesting to note that “fermion quantum mechanics” $\mathcal{A}_-(\mathbb{C}^M)$ is the natural framework for formulating and studying lattice fermion and spin systems which play an important role in condensed matter physics.

In the following, we elaborate the naive interpretations of the relativistic equations in [2] and [3] as a quantum theory of one particle, and we discuss why they are unphysical. For simplicity, we assume that the electromagnetic fields ϕ, \mathbf{A} are time independent. We then show that quasifree representations as discussed above can provide physically acceptable many-particle theories. We first consider the Dirac case, which is somewhat simpler.

Fermions

One-particle formalism Recalling that $i\partial_t$ is the energy operator, we define the Dirac Hamiltonian D by rewriting [3] in the following form:

$$i\partial_t\psi = D\psi, \quad D = (-i\nabla + e\mathbf{A}) \cdot \boldsymbol{\alpha} + m\beta - e\phi \quad [36]$$

This Dirac Hamiltonian is obviously a self-adjoint operator on the one-particle Hilbert space $h = L^2(\mathbb{R}^4) \otimes \mathbb{C}^4$, but, different from the Schrödinger Hamiltonian in [1], it is not bounded from below: for any $E_0 > -\infty$, one can find a state f such that the energy expectation value (f, Df) is less than E_0 . This can be easily seen for the simplest case where the external potential vanishes, $\mathbf{A} = \phi = 0$. Then the eigenvalues of D can be computed by Fourier transformation, and one finds

$$E = \pm\sqrt{\mathbf{p}^2 + m^2}, \quad \mathbf{p} \in \mathbb{R}^3 \quad [37]$$

Due to the negative energy eigenvalues we conclude that there is no ground state, and the Dirac Hamiltonian thus describes an unstable system, which is physically meaningless.

To summarize: a (unphysical) one-particle description of relativistic fermions is given by a Hilbert space h together with a self-adjoint Hamiltonian D unbounded from below. Other observables and transformations are given by self-adjoint and unitary operators on h , respectively.

Many-body formalism We now explain how to construct a physical many-body description from these data. To simplify notation, we first assume that D has a purely discrete spectrum (which can be achieved by using a compact space). We can then label the eigenfunctions f_n by integers n such that the corresponding eigenvalues $E_n \geq 0$ for $n \geq 0$ and $E_n < 0$ for $n < 0$. Using the naive representation of the fermion field algebra discussed in the last section, we get (we use the notation introduced in [22])

$$q(D) = \sum_{n \geq 0} |E_n| \psi_n^\dagger \psi_n - \sum_{n < 0} |E_n| \psi_n^\dagger \psi_n \quad [38]$$

which is obviously not bounded from below and thus not physically meaningful. However, $\psi_n^\dagger \psi_n = 1 - \psi_n \psi_n^\dagger$, which suggests that we can remedy this problem by interchanging the creation and annihilation operators for $n < 0$. This is possible: it is easy to see that

$$\hat{\psi}_n \equiv \psi_n \quad \forall n \geq 0 \quad \text{and} \quad \hat{\psi}_n \equiv \psi_n^\dagger \quad \forall n < 0 \quad [39]$$

provides a representation of the algebra in [23]. We thus define

$$\hat{q}(D) \equiv \sum_{n \in \mathbb{Z}} E_n : \hat{\psi}_n^\dagger \hat{\psi}_n : \quad [40]$$

with the so-called normal ordering prescription

$$:\psi_m^\dagger \psi_n : \equiv \psi_m^\dagger \psi_n - \langle \Omega, \psi_m^\dagger \psi_n \Omega \rangle \quad [41]$$

where we made use of the freedom of normalization explained after [23] to eliminate unwanted additive constants. We get $q(D) = \sum_{n \in \mathbb{Z}} |E_n| \psi_n^\dagger \psi_n$, which is manifestly a non-negative self-adjoint operator with Ω as ground state. We thus found a physical many-body description for our model. We can now define for other one-particle observables,

$$\hat{q}(A) \equiv \sum_{n \in \mathbb{Z}} A_{mn} : \hat{\psi}_m^\dagger \hat{\psi}_n : \quad [42]$$

and, by straightforward computations, we obtain

$$[\hat{q}(A), \hat{q}(B)] = \hat{q}([A, B]) + S(A, B) \quad [43]$$

where $S(A, B) = \sum_{m < 0} \sum_{n \geq 0} (A_{mn} B_{nm} - B_{mn} A_{nm})$, that is,

$$S(A, B) = \text{tr}(P_- A P_+ B P_- - P_- B P_+ A P_-) \quad [44]$$

with $P_- = \sum_{n < 0} f_n(f_n, \cdot)$ the projection onto the subspace spanned by the negative energy eigenvectors of D and $P_+ = 1 - P_-$. One can show that $\hat{q}(A)$ is no longer defined for all operators but only if

$$P_- A P_+ \text{ and } P_+ A P_- \text{ are} \\ \text{Hilbert-Schmidt operators} \quad [45]$$

(we recall that a is a Hilbert-Schmidt operator if $\text{tr}(a^* a) < \infty$). The \mathbb{C} -number term $S(A, B)$ in [43] is

often called Schwinger term and, different from the similar term in [25], it is now nontrivial, that is, it is no longer possible to remove it by a redefinition $\hat{q}'(A) = \hat{q}(A) - b(A)$. This Schwinger term is an example of an anomaly, and it has various interesting implications.

In a similar manner, one can construct the many-body transformations $\hat{Q}(U)$ of unitary operators U on h satisfying the very Hilbert–Schmidt condition in [45], and one obtains

$$\hat{Q}(U)\hat{Q}(V) = \chi(U, V)\hat{Q}(UV) \quad [46]$$

with interesting phase-valued functions χ .

More generally, for any one-particle Hilbert space h and Dirac Hamiltonian D , the physical representation is given by the quasifree representation π_{P_-} in [34] with P_- the projection onto the negative energy subspace of D . The results about \hat{q} and \hat{Q} mentioned hold true in any such representation.

Thus the one-particle Hamiltonian D determines which representation one has to use, and one therefore cannot construct the “physical” representation without specific information about D . However, not all these representations are truly different: if there is a unitary operator U on the Fock space $\mathcal{F}_+(h)$ such that

$$U^* \pi_{P_-^{(1)}}(\psi^{(\dagger)}(f))U = \pi_{P_-^{(2)}}(\psi^{(\dagger)}(f)) \quad [47]$$

for all $f \in h$, then the quasifree representations associated with the different projections $P_-^{(1)}$ and $P_-^{(2)}$ are physically equivalent: one could equally well formulate the second model using the representation of the first. Two such quasifree representations are called unitarily equivalent, and a fundamental theorem due to Shale and Stinespring states that *two quasifree representations $\pi_{P_-^{(1,2)}}$ are unitarily equivalent if and only if $P_-^{(1)} - P_-^{(2)}$ is a Hilbert–Schmidt operator* (a similar result holds true in the boson case).

Bosons

One-particle formalism Similarly as for the Dirac case, the solutions of the Klein–Gordon equation in [2] also do not define a physically acceptable one-particle quantum theory with a ground state: the energy eigenvalues in [37] for $A = \phi = 0$ are a consequence the relativistic invariance and thus equally true for the Klein–Gordon case. However, in this case there is a further problem. To find the one-particle Hamiltonian, one can rewrite the second-order equation in [2] as a system of first-order equations,

$$\begin{aligned} i\partial_t \Phi &= K\Phi \\ \Phi &= \begin{pmatrix} \psi \\ \pi^\dagger \end{pmatrix}, \quad K = \begin{pmatrix} C & i \\ -iB^2 & C \end{pmatrix} \end{aligned} \quad [48]$$

with

$$B^2 \equiv (-i\nabla + eA)^2 + m^2, \quad C \equiv -e\phi \quad [49]$$

Thus, one sees that the natural one-particle Hilbert space for the Klein–Gordon equation is $h = L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$; here, and in the following, we identify h with $h_0 \oplus h_0$, $h_0 = L^2(\mathbb{R}^3)$, and use a convenient 2×2 matrix notation naturally associated with that splitting. However, the one-particle Hamiltonian is not self-adjoint but rather obeys

$$K^* = JKJ, \quad J \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad [50]$$

with $*$ the Hilbert space adjoint. It is important to note that J is a grading operator. Thus, we can define a sesquilinear form

$$(f, g)_J \equiv (f, Jg) \quad \forall f, g \in h \quad [51]$$

with (\cdot, \cdot) the standard inner product, and [50] is equivalent to K being self-adjoint with respect to this sesquilinear form; in this case, we say that K is J -self-adjoint. Thus, in the Klein–Gordon case, this sesquilinear form takes the role of the Hilbert space inner product and, in particular, not (Φ, Φ) but $(\Phi, \Phi)_J$ is preserved under time evolution. However, different from $\Phi^\dagger \Phi$, $\Phi^\dagger J \Phi$ is not positive definite, and it is therefore not possible to interpret it as probability density as in conventional quantum mechanics. For consistency, one has to require that one-particle transformations U are unitary with respect to $(\Phi, \Phi)_J$, that is, $U^{-1} = JUJ$. We call such operators J -unitary.

To summarize: a (unphysical) one-particle description of relativistic bosons is given by a Hilbert space of the form $h = h_0 \oplus h_0$, the grading operator J in [50], and a J -self-adjoint Hamiltonian K of the form as in eqn [48], where $B \geq 0$ and C are self-adjoint operators on h_0 . Other observables and transformations are given by J -self-adjoint and J -unitary operators on h , respectively.

Many-body formalism We first consider the quasifree representation $\pi_{P_-^{(0)}}$ of the boson field algebra $\mathcal{A}_-(h)$ so that the grading operator in [35] is equal to J , that is, $P_-^{(0)} = (1 - J)/2$. Writing $\pi_{P_-^{(0)}}(\Psi^{(\dagger)}(f)) = \psi^{(\dagger)}(f)$, one finds that

$$q(A)^* = q(JAJ), \quad Q(U)^* = Q(JU^*J) \quad [52]$$

and thus J -self-adjoint operators and J -unitary operators are mapped to proper observables and transformations. In particular, $q(K)$ is a self-adjoint

operator, which resolves one problem of the one-particle theory. However, $q(K)$ is not bounded from below, and thus $\pi_{P(0)}$ is not yet the physical representation.

The physical representation can be constructed using the operators

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} B^{1/2} & iB^{-1/2} \\ B^{1/2} & iB^{-1/2} \end{pmatrix}, \quad F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad [53]$$

(for simplicity, we restrict ourselves to the case $C=0$ and $B > 0$; we use the calculus of self-adjoint operators here) with the following remarkable properties:

$$\begin{aligned} T^{-1} &= JT^*F \\ TKT^{-1} &= \begin{pmatrix} B & 0 \\ 0 & -B \end{pmatrix} \equiv \hat{K} \end{aligned} \quad [54]$$

One can check that

$$\hat{\psi}^\dagger(f) \equiv \psi^\dagger(Tf), \quad \hat{\psi}(f) \equiv \psi(T^{-1}f) \quad [55]$$

is a quasifree representation π_{P_-} of $\mathcal{A}_-(\hbar)$ with $P_- = (1 - F)/2$. With that the construction of \hat{q} and \hat{Q} is very similar to the fermion case described above (the crucial simplification is that \hat{K} and F now are diagonal). In particular, $\hat{q}(K)$ is a non-negative operator with the ground state Ω , and $\hat{q}(A)$ and $\hat{Q}(U)$ are self-adjoint and unitary for every one-particle observable A and transformation U , respectively. One also gets relations as in [43] and [46].

Related Topics of Recent Interest

The impossibility to construct relativistic quantum-mechanical models played an important role in the early history of quantum field theory, as beautifully discussed in chapter 1 of Weinberg (1995).

The abstract formalism of quasifree representations of fermion and boson field algebras was developed in many papers (see, e.g., Ruijsenaars (1977), Grosse and Langmann (1992), and Langmann (1994) for explicit results on \hat{Q} and χ). A nice textbook presentation with many references can be found in chapter 13 of Gracia-Bondía *et al.* (2001) (this chapter is rather self-contained but mainly restricted to the fermion case).

Based on the Shale–Stinespring theorem, there has been considerable amount of work to investigate whether the quasifree representations associated with different external electromagnetic fields ψ_1, A_1 and ψ_2, A_2 are unitarily equivalent, if and which time-dependent many-body Hamiltonians exist, etc. (see chapter 13 of Gracia-Bondía *et al.* (2001), and references therein).

The infinite-dimensional Lie algebra \mathfrak{g}_2 of Hilbert space operators satisfying the condition in [45] is an interesting infinite-dimensional Lie algebra with a beautiful representation theory. This subject is closely

related to conformal field theory (see, e.g., Kac and Raina (1987) for a textbook presentation and Carey and Ruijsenaars (1987) for a detailed mathematical account within the framework described by us).

It turns out that the mathematical framework discussed in the previous section is sufficient for constructing fully interacting quantum field theories, in particular Yang–Mills gauge theories, in $1+1$ but not in higher dimensions. The reason is that, in $3+1$ dimensions, the one-particle observables A of interest do not obey the Hilbert–Schmidt condition in [45] but only the weaker condition

$$\text{tr}(a^*a)^n < \infty, \quad a = P_\mp AP_\pm \quad [56]$$

with $n=2$, and the natural analog of \mathfrak{g}_2 in $3+1$ dimensions thus seems to be the Lie algebra \mathfrak{g}_{2n} of operators satisfying this condition with $n=2$. Various results on the representation theory of such Lie algebras $\mathfrak{g}_{2n>2}$ have been developed (see Mickelsson (1989), where various interesting relations to infinite-dimensional geometry are also discussed).

As mentioned, the Schwinger term $S(A,B)$ in [44] is an example of an anomaly. Mathematically, it is a nontrivial 2-cocycle of the Lie algebra \mathfrak{g}_2 , and analogs for the groups $\mathfrak{g}_{2n>2}$ have been found. These cocycles provide a natural generalization of anomalies (in the meaning of particle physics) to operator algebras. They not only shed some interesting light on the latter, but also provide a link to notions and results from noncommutative geometry (see, e.g., Gracia-Bondía *et al.* (2001)). We believe that this link can provide a fruitful driving force and inspiration to find ways to deepen our understanding of quantum Yang–Mills theories in $3+1$ dimensions (Langmann 1996).

See also: Anomalies; C^* -Algebras and Their Classification; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Dirac Operator and Dirac Field; Gerbes in Quantum Field Theory; Quantum Field Theory in Curved Spacetime; Quantum n -Body Problem; Superfluids; Two-Dimensional Models.

Further Reading

- Carey AL and Ruijsenaars SNM (1987) On fermion gauge groups, current algebras and Kac–Moody algebras. *Acta Applicandae Mathematicae* 10: 1–86.
- DeWitt B (2003) *The Global Approach to Quantum Field Theory*, International Series of Monographs on Physics, vols. 1 and 2, p. 114. New York: Oxford University Press.
- Gracia-Bondía JM, Várilly JC, and Figueroa H (2001) *Elements of Noncommutative Geometry*, Birkhäuser Advanced Texts: Basel Textbooks. Boston: Birkhäuser.
- Grosse H and Langmann E (1992) A supersversion of quasifree second quantization. *Journal of Mathematical Physics* 33: 1032–1046.
- Kac VG and Raina AK (1987) *Bombay Lectures on Highest Weight Representations of Infinite-Dimensional Lie Algebras*,

- Advanced Series in Mathematical Physics, vol. 2. Teaneck: World Scientific Publishing.
- Langmann E (1994) Cocycles for boson and fermion Bogoliubov transformations. *Journal of Mathematical Physics* 96–112.
- Langmann E (1996) Quantum gauge theories and noncommutative geometry. *Acta Physica Polonica B* 27: 2477–2496.
- Mickelsson J (1989) *Current Algebras and Groups*, Plenum Monographs in Nonlinear Physics. New York: Plenum Press.
- Rafelski J, Fulcher LP, and Klein A (1978) Fermions and bosons interacting with arbitrary strong external fields. *Physics Reports* 38: 227–361.

- Reed M and Simon B (1975) *Methods of Modern Mathematical Physics. II. Fourier Analysis, Self-Adjointness*. New York: Academic Press.
- Ruijsenaars SNM (1977) On Bogoliubov transformations for systems of relativistic charged particles. *Journal of Mathematical Physics* 18: 517–526.
- Weinberg S (1995) *The Quantum Theory of Fields*, vol. I (English summary) Foundations. Cambridge: Cambridge University Press.

Boundaries for Spacetimes

S G Harris, St. Louis University, St. Louis, MO, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

There is a common practice in mathematics of placing a boundary on an object which may not appear to come naturally equipped with one; this is often thought of as adding ideal points to the object. Perhaps the most famous example is the addition of a single “point at infinity” to the complex plane, resulting in the Riemann sphere: this is a boundary point in the sense of providing an ideal endpoint for lines and other endless curves in the plane. Often, there is more than one reasonable way to construct a boundary for a given object, depending on the intent; for instance, the plane is sometimes equipped, not with a single point at infinity, but with a circle at infinity, resulting in a space homeomorphic to a closed disk. Both these boundaries on the plane have useful but different things to tell us about the nature of the plane; the common feature is that, by bringing the infinite reach of the plane within the confines of a more finite object, we are better able to grasp the behavior of the original object.

The general usefulness of the construction of boundaries for an object is to allow behavior of structures in the “completed” object to aid in visualization of behavior in the original object, such as by providing a degree of measurement or other classification of processes at infinity. This utility has not been overlooked for spacetimes. A variety of purposes may be served by various boundary construction methods: providing a locale for singularities (as the spacetime itself is modeled by a smooth manifold with a smooth metric, free of singular points); providing a platform from which to measure global properties such as total energy or angular momentum; displaying in finite form the causal structure at infinity; or providing a compact (or quasicompact) topological envelope for the spacetime while preserving the causal structure.

This article will consider several of the methods that have been used or proposed for constructing boundaries for spacetimes, ranging from the *ad hoc* (but practical) to the universal. Perhaps the simplest way to classify these methods is into those which employ or analyze embeddings of the spacetime in question and those that do not.

Boundaries from Embeddings

General

The simplest and most common method of constructing a boundary for a spacetime M is to find a suitable manifold N (of the same dimension) and an appropriate map $\phi: M \rightarrow N$ which is a topological embedding, that is, a homeomorphism onto its image $\phi(M)$. We can consider \bar{M}_ϕ , the closure of $\phi(M)$ in N , as the ϕ -completion of M , and $\partial_\phi(M) = \bar{M}_\phi - \phi(M)$ as the ϕ -boundary. Typically, this embedding is chosen in such a way that curves of interest in M – such as timelike or null geodesics or causal curves of bounded acceleration – which have no endpoints in M , do have endpoints in $\partial_\phi(M)$; in other words, if $c: [0, \infty) \rightarrow M$ is such a curve of interest, then $\lim_{t \rightarrow \infty} \phi(c(t))$ exists in N .

The common practice, initiated by Penrose in 1967, is to choose N to be another spacetime – often called the unphysical spacetime, while M is considered the spacetime of physical interest – and to require the embedding ϕ to be a conformal mapping, that is, ϕ carries the spacetime metric in M to a scalar multiple of the spacetime metric in N . As conformal maps preserve the local causal structure, leaving unchanged the notions of timelike curve or null curve, this means that \bar{M}_ϕ inherits from N a causal structure which, locally, is an extension of that of M . This allows us to speak of causal relationships within \bar{M}_ϕ , closely related to those in M .

Minkowski Space

The prototypical example is the conformal embedding of Minkowski space into the Einstein static spacetime.

Let \mathbb{R}^n denote Euclidean n -space, \mathbb{S}^n the unit n -sphere, and \mathbb{L}^n Minkowski n -space, that is, \mathbb{R}^n with metric $ds^2 = dx_1^2 + \dots + dx_{n-1}^2 - dt^2$ (so $\mathbb{L}^n = \mathbb{R}^{n-1} \times \mathbb{L}^1$). The n -dimensional Einstein static spacetime is the product spacetime $\mathbb{E}^n = \mathbb{S}^{n-1} \times \mathbb{L}^1$. Consider \mathbb{S}^{n-1} as embedded in $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}^1$. Then the conformal embedding is $\phi: \mathbb{L}^n \rightarrow \mathbb{E}^n$, expressed as $\phi: \mathbb{R}^{n-1} \times \mathbb{L}^1 \rightarrow \mathbb{S}^{n-1} \times \mathbb{L}^1 \subset \mathbb{R}^{n-1} \times \mathbb{R}^1 \times \mathbb{L}^1$ given by $\phi(x, t) = ((x/|x|) \sin \theta, \cos \theta, \tau)$, where $\theta = \tan^{-1}(t + |x|) - \tan^{-1}(t - |x|)$ and $\tau = \tan^{-1}(t + |x|) + \tan^{-1}(t - |x|)$. The boundary $\partial_\phi(\mathbb{L}^n)$ consists of the following: the points $\{\theta + \tau = \pi; 0 < \tau \leq \pi\}$, composed of an \mathbb{S}^{n-2} of null lines coming together at the point $i^+ = (0, 1, \pi)$; a similar cone of null lines $\{\theta - \tau = \pi; -\pi \leq \tau < 0\}$ with vertex at $i^- = (0, 1, -\pi)$; and a single limit-point for both cones at $i^0 = (0, -1, 0)$. The $\tau > 0$ null cone is called \mathfrak{S}^+ (the letter is read “scri” for “script-I”), its counterpart \mathfrak{S}^- (Figures 1 and 2). As all future-directed timelike geodesics in \mathbb{L}^n have i^+ as an endpoint in \mathbb{E}^n , i^+ is called future-timelike infinity; similarly, i^- is past-timelike infinity. Every future-directed null geodesic ends up on \mathfrak{S}^+ , which is thus

termed future-null infinity, and \mathfrak{S}^- is past-null infinity. All spacelike geodesics come to i^0 , spacelike infinity.

For $n=2$, this picture produces the familiar diamond representation of \mathbb{L}^2 (Figure 3): as \mathbb{E}^2 is easily unrolled into another copy of \mathbb{L}^2 (metric

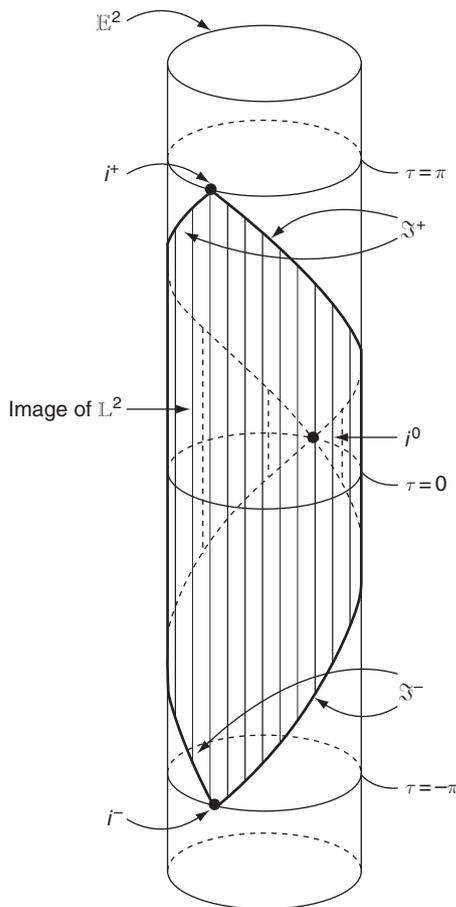


Figure 1 \mathbb{L}^2 conformally embedded in $\mathbb{E}^2 = \mathbb{S}^1 \times \mathbb{L}^1$.

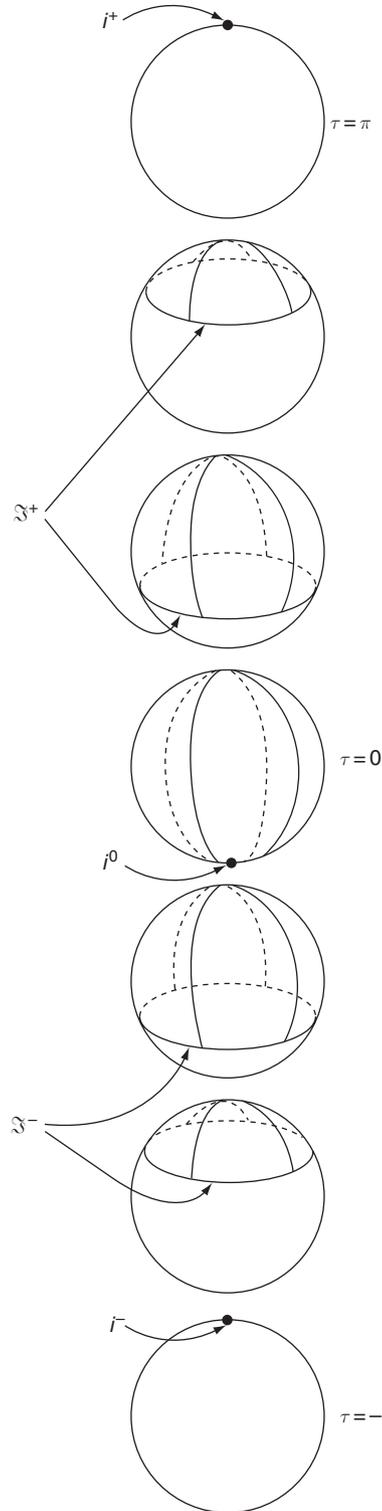


Figure 2 \mathbb{L}^3 conformally embedded in $\mathbb{E}^3 = \mathbb{S}^2 \times \mathbb{L}^1$.

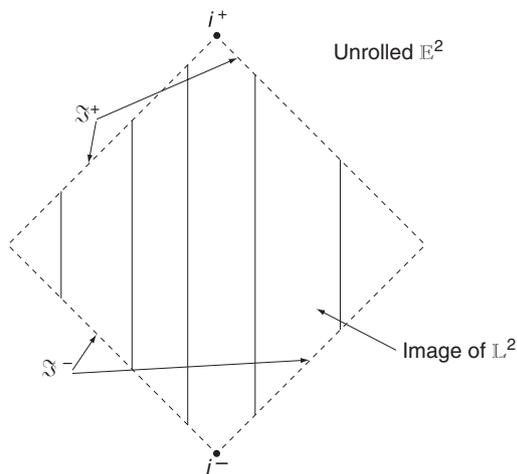


Figure 3 L^2 conformally embedded in unrolled \mathbb{E}^2 , i.e., $\mathbb{R}^1 \times L^1 = L^2$.

$d\theta^2 - d\tau^2$), this means that $\phi(L^2)$ is the region $|\theta| + |\tau| < \pi$ in L^2 ; timelike curves and null geodesics in the original L^2 are the same as in $\phi(L^2)$, and their endpoints in the boundary of the diamond are evident. For higher dimensions, the picture is not as visually obvious, since \mathbb{E}^n cannot be unrolled; but the principle of reading the causal structure at infinity of L^n via its boundary points in \mathbb{E}^n remains the same.

Conformal Embeddings

There have been various formulations designed to emulate the conformal mapping of L^n with respect to spacetimes, which are, in some sense, asymptotically like Minkowski space being conformally mapped into larger spacetimes. A spacetime M with metric g is called asymptotically simple or (alternatively) asymptotically flat if there is a spacetime N with metric h , an embedding $\phi: M \rightarrow N$, and a scalar function Ω defined on N with $\phi^*h = (\Omega \circ \phi)^2g$ (i.e., ϕ is conformal with Ω^2 the conformal factor) and $\Omega = 0$ on $\partial_\phi(M)$, $d\Omega \neq 0$ on $\partial_\phi(M)$, and various other restrictions on Ω , depending on the intent. One can define asymptotic symmetries of M by means of motions within $\partial_\phi(M)$, leading to notions of global energy and angular momentum (see [Hawking and Ellis \(1973\)](#) and [Wald \(1984\)](#) for details).

Classifications of Embeddings

As a general rule, there is no uniqueness in the choice of an embedding ϕ for a spacetime M to construct a boundary, nor in the topology of the resulting boundary $\partial_\phi(M)$, or even of which curves of interest end up having endpoints in the boundary. In an attempt to categorize which embeddings yield equivalent results and what sort of results there are in terms of endpoints of curves, [Scott and Szekeres](#)

(1994) formulated what they called the abstract boundary of a spacetime. This depends on a choice of class of “interesting” curves, each characterizable as having either infinite or finite parameter length; typical choices for this class would be timelike geodesics or causal geodesics or timelike curves of bounded acceleration. For instance, a boundary point may be said to represent a singularity with respect to the chosen class of curves if it is the endpoint of one such curve with finite parameter length; nonsingular points are points at infinity. These classifications do not require conformal embeddings, nor even that the target of the embeddings be spacetimes; they accommodate boundaries of a far more general type than Penrose’s notion stemming from conformal embeddings.

A somewhat different study of boundaries from embeddings has been formulated by [García-Parrado and Senovilla \(2003\)](#), classifying points at infinity and singularities in $\partial_\phi(M)$ for embeddings $\phi: M \rightarrow N$ in which N is a spacetime, ϕ preserves the chronology relation \ll , and there is also a diffeomorphism $\psi: \phi(M) \rightarrow N$ which again preserves \ll (the chronology relation in a spacetime is defined thus: $x \ll y$ if and only if there is a future-directed timelike curve from x to y). This scheme applies more generally than to conformal embeddings, but the requirement for chronology-preserving maps in both directions guarantees a strong sensitivity to causality; it amounts to a mild extension of Penrose’s notion that is often much easier to construct.

Universal Constructions

B-Boundary

Attempts have been made to formulate boundary concepts specifically for defining singularities as ideal endpoints for finite-length geodesics. The most complete venture in this direction is the b-boundary (“b” for “bundle”) of Schmidt ([Hawking and Ellis 1973](#), pp. 276–284). This is a formulation that takes note only of the connection in the linear frames bundle $L(M)$ of a spacetime M (or of any manifold with a linear connection, metric or otherwise); in other words, it takes no particular note of the spacetime metric or even of the causal structure of the spacetime, but only of the notion of parallel translation of tangent vectors along curves. Parallel translation of a frame (a basis for the tangent space) along a curve is used to obtain an *ad hoc* length for the curve by treating the translated frame as positive-definite orthonormal at each point; whether this length is finite or infinite is independent of the choice of the original frame. The Schmidt construction

defines a boundary on M which gives an endpoint for each curve, endless in M , which is finite in that sense: Select a positive-definite metric on $L(M)$, give it a boundary by means of Cauchy completion, and then take the appropriate quotient by the bundle group. This has an appealing universality of application, but the problems of putting it into practice are quite formidable. Also, the fact that it takes no special note of the spacetime character of M suggests that it may not be of particular utility for physical insights.

Causal Boundary: Basics

In 1972 Geroch, Kronheimer, and Penrose (GKP) formulated a notion of boundary – the causal boundary – that is specifically adapted to the causal character of a spacetime M ; indeed, it is defined in such a way that one need know only the chronology relation \ll on M without any further reference to the metric (another way of saying this is that the causal boundary is conformally invariant). Like Schmidt’s b-boundary, the causal boundary is a universal construction, not depending on any extraneous choices; however, although it has an obvious clarity in its causal structure, there are subtleties in the choice of an appropriate topology which are perhaps not yet fully resolved. As this boundary construction appears to embody the best hopes for a practical universal construction, it is detailed here in some depth.

The causal boundary construction applies only to strongly causal spacetimes; essentially, this means that the local causal structure at each point is exactly reflective of the global causal structure.

The basic construction of the causal boundary of a spacetime M starts with two separate parts: the future and past (pre-)boundaries of M , intended as yielding endpoints for, respectively, future- and past-endless causal curves. Part of the difficulty of the causal boundary is knowing how best to meld these two into one; currently, there are several answers to this conundrum.

The elements of the future causal boundary of M are defined in terms of the past-set operator I^- . For a point $x \in M$, the past of x is $I^-(x) = \{y \mid y \ll x\}$; for a set $A \subset M$, $I^-[A] = \bigcup_{x \in A} I^-(x)$. A set $P \subset M$ is called a past set if $I^-[P] = P$; anything of the form $P = I^-[A]$ is a past set, and all past sets have this form. A past set P is an indecomposable past set (IP) if P cannot be written as $P_1 \cup P_2$ for past sets which are proper subsets $P_i \subsetneq P$. IPs come in exactly two varieties: pointlike IPs (PIPs), of the form $I^-(x)$ (Figure 4), and terminal IPs (TIPs), of the form $I^-[c]$ for c a future-endless causal curve (Figure 5). (Of course, any $I^-(x)$ can also be expressed as $I^-[c]$ for c

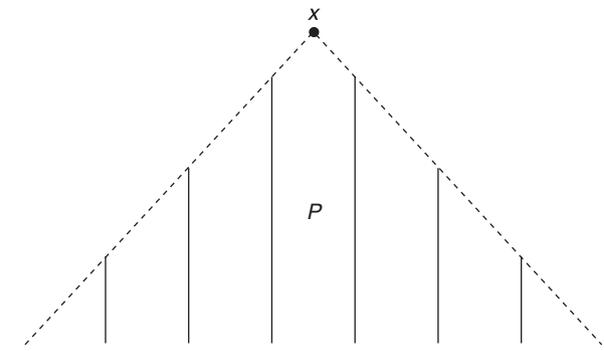


Figure 4 PIP $P = I^-(x)$.

a causal curve ending at x .) The future causal boundary of M , $\hat{\partial}(M)$, consists of all the TIPs of M ; the future causal completion of M is $\hat{M} = \hat{\partial}(M) \cup M$. But that is just a set; the causal structure of M needs to be extended to \hat{M} .

For any $x \in M$ and $P \in \hat{\partial}(M)$, set $x \ll P$ if and only if $x \in P$; set $P \ll x$ if and only if $P \subset I^-(y)$ for some $y \ll x$ ($y \in M$); and for P and Q in $\hat{\partial}(M)$, set $P \ll Q$ if and only if $P \subset I^-(y)$ for some $y \in Q$. If we consider this an extension of the \ll relation on M , then we end up with a relation which, like that on M , is transitive and antireflexive. Furthermore, it has the property that for all $\alpha, \beta \in \hat{M}$, $\alpha \ll \beta$ if and only if for some $x \in M$, $\alpha \ll x \ll \beta$. (One can also amend the chronology relation within M to be more like the definition in the extension; that is not of major import.)

We can also extend the causality relation \prec on M to one on \hat{M} (in M , $x \prec y$ if there is a future-directed

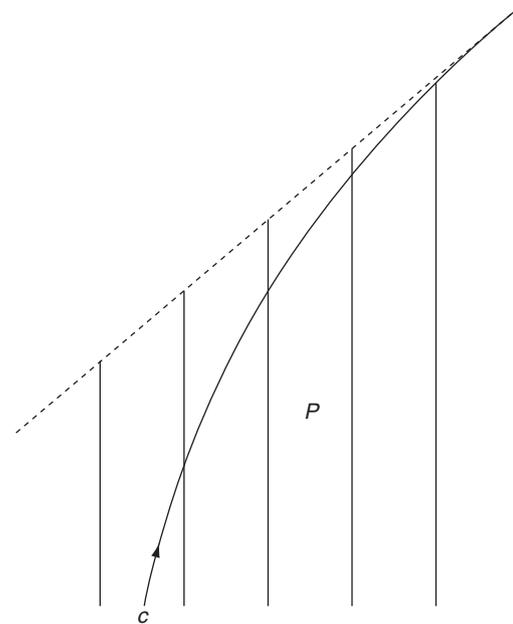


Figure 5 TIP $P = I^-[c]$.

causal curve from x to y): for $x \in M$ and $P, Q \in \hat{\partial}(M)$, $x \prec P$ for $I^-(x) \subset P$, $P \prec x$ for $P \subset I^-(x)$, and $P \prec Q$ for $P \subset Q$.

The intent is to have the elements of $\hat{\partial}(M)$ provide future endpoints for future-endless causal curves in M ; in particular, we want two such curves, c_1 and c_2 , to be assigned the same future endpoint precisely when $I^-[c_1] = I^-[c_2]$. This is accomplished by the simple expedient of defining the future endpoint of a future-endless causal curve c to be $P = I^-[c]$. We do not have a topology on \hat{M} as yet, but it is worth noting that if P is the assigned future endpoint of c , then $I^-(P) = I^-[c]$; this is at least the correct causal behavior for a putative future endpoint of c .

We can perform all the operations above in the time-dual manner, obtaining the past causal boundary $\check{\partial}(M)$, consisting of terminal indecomposable future sets (TIFs), and the past causal completion $\check{M} = \check{\partial}(M) \cup M$. The full causal boundary of M consists of the union of $\hat{\partial}(M)$ with $\check{\partial}(M)$ with some sort of identifications to be made.

As an example of the need for identifications, consider M to be \mathbb{L}^2 with a closed timelike line segment deleted, say $M = \mathbb{L}^2 - \{(0, t) \mid 0 \leq t \leq 1\}$. For $\hat{\partial}(M)$, we have first the boundary elements at infinity: the TIP $i^+ = M$ (the past of the positive time axis) and the set of TIPs making up \mathfrak{S}^+ (the pasts of null lines going out to infinity in \mathbb{L}^2); and then, the boundary elements coming from the deleted points: for each t with $0 < t \leq 1$, two IPs emanating from $(0, t)$, that is, P_t^+ , the past of the null line going pastwards from $(0, t)$ toward $x > 0$, and P_t^- , the past of the null line going pastwards from $(0, t)$ toward $x < 0$; and P_0 , emanating from $(0, 0)$, that is, the past of the negative time axis. Similarly, $\check{\partial}(M)$ consists of i^- , \mathfrak{S}^- , TIFs F_t^+ and F_t^- emanating from $(0, t)$ for $0 \leq t < 1$, and the TIF F_1 emanating from $(0, 1)$. We probably want to make at least the following identifications for each t with $0 < t < 1$, $P_t^+ \equiv F_t^+$ and $P_t^- \equiv F_t^-$; $P_1^+ \equiv F_1 \equiv P_1^-$; and $F_0^+ \equiv P_0 \equiv F_0^-$. This results in a two-sided replacement for the deleted segment; for some purposes, it might be deemed desirable to identify the two sides as one, but a universal boundary is probably a good idea, leaving further identifications as optional quotients of the universal object.

How best to define the appropriate identifications in general is a matter of some controversy. GKP defined a somewhat complicated topology on $\bar{M} = \hat{\partial}(M) \cup \check{\partial}(M) \cup M$, then used an identification intended to result in a Hausdorff space. There are significant problems with this approach in some *outré* spacetimes, as pointed out by Budic and Sachs (1974) and Szabados (1989), both of whom recommended a different set of identifications. But what is

of more concern is that the topology prescribed by GKP is not what might be expected in even the simplest of cases, for example, Minkowski space: $\bar{\mathbb{L}}^n$ needs no identifications among boundary points (no matter whose identification procedure is followed). The GKP topology on $\bar{\mathbb{L}}^n$, restricted to $\hat{\partial}(\mathbb{L}^n)$, is not that of a cone ($S^{n-2} \times \mathbb{R}^1$ with a point added), as is the case for \mathfrak{S}^+ in the conformal embedding into \mathbb{E}^n ; but, instead, each null line in $\hat{\partial}(\mathbb{L}^n)$ (not including i^+) is an open set, and i^+ has no neighborhood in $\hat{\partial}(\mathbb{L}^n)$ save for the entire boundary. This is a topology bearing no relation at all to that of any embedding.

Future Causal Boundary

Construction An alternative approach, initiated by Harris (1998), is to forego the full causal boundary and concentrate only on \hat{M} and \check{M} separately. There is an advantage to this in that the process of future causal completion – that is to say, forming \hat{M} from M – can be made functorial in an appropriate category of “chronological sets”: a set X with a relation \ll which is transitive and antireflexive such that it possesses a countable subset S which is “chronologically dense,” that is, for any $x, y \in X$, there is some $s \in S$ with $x \ll s \ll y$. Any strongly causal spacetime M is a chronological set, as is \hat{M} . The entire construction of the future causal boundary works just as well for a chronological set. The role of a timelike curve in a chronological set is taken by a future chain: a sequence $c = \{x_n\}$ with $x_n \ll x_{n+1}$ for all n . For any future chain c , $I^-[c]$ is an IP, and any IP can be so expressed; but unlike in spacetimes, $I^-(x)$ may or may not be an IP for $x \in X$. Then, \hat{X} is always future complete in the sense that for any future chain c in \hat{X} , there is an element $\alpha \in \hat{X}$ with $I^-(\alpha) = I^-[c]$: for instance, if the chain c lies in X but there is no $x \in X$ with $I^-(x) = I^-[c]$, just let $\alpha = I^-[c]$, which is an element of $\hat{\partial}(X)$. This yields a functor of future completion from the category of chronological sets to the category of future-complete chronological sets, and the embedding $X \rightarrow \hat{X}$ is a universal object in the sense of the category theory; this implies that it is categorically unique and is the minimal future-completion process.

However, it is crucial to have more than the chronology relation operating in what is to be a boundary; topology of some sort is needed. This is accomplished by defining what might be called the future-chronological topology for any chronological set – including for \hat{M} when M is a strongly causal spacetime. This topology is defined by means of a limit-operator \hat{L} on sequences: if X is the chronological set, then for any sequence of points $\sigma = \{x_n\}$ in X , $\hat{L}(\sigma)$ denotes a subset of X which is the set of

limits of σ . It is explicitly recognized that there may be more than one limit of a sequence, as the space may not be Hausdorff; no attempt is made to remove any non-Hausdorffness, as this is viewed as giving important information on how, possibly, two points in the future causal boundary represent very similar and yet not identical pieces of information about the causal structure at infinity. Once the limit operator is in place, the actual topology on X is defined thus: a subset $A \subset X$ is said to be closed if and only if for any sequence $\sigma \subset A, \hat{L}(\sigma) \subset A$ (and open sets are complements of closed sets). This yields the elements of $\hat{L}(\sigma)$ as topological limits of σ .

The definition of \hat{L} is simplest when X has the property that $I^-(x)$ is an IP for any $x \in X$; as this is true for X being either a spacetime M or the future causal completion \hat{M} of a spacetime, the discussion here is restricted to this situation. Let us also make the common assumption that X is past-distinguishing, that is, $I^-(x) = I^-(y)$ implies $x = y$.

Let $\sigma = \{x_n\}$ be a sequence of points in a past-distinguishing chronological set X in which the past of any point is an IP. Then $\hat{L}(\sigma)$ consists of those points x for which (see Figures 6 and 7)

1. for all $y \in I^-(x)$, for n sufficiently large, $y \ll x_n$, and
2. for any IP $P \supseteq I^-(x)$, there is some $z \in P$ such that for n sufficiently large, $z \ll x_n$.

Then the future-chronological topology on X has these features:

1. It is a T_1 topology, that is, points are closed.
2. If $I^-(x) = I^-[c]$ for a future chain $c = \{x_n\}$, then x is a topological limit of the sequence $\{x_n\}$.
3. If $X = M$, a strongly causal spacetime, then the future-chronological topology is precisely the manifold topology.
4. If $X = \hat{M}$, the future causal completion of a strongly causal spacetime M , then the induced topology on M is the manifold topology, $\hat{\partial}(M)$ is a closed subset of \hat{M} , and M is dense in \hat{M} . As per property (2), for any future-endless causal curve c

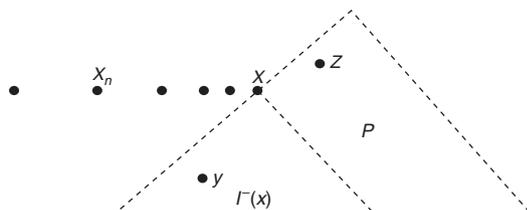


Figure 6 $x \in \hat{L}(\{x_n\})$: for all $y \in I^-(x)$, eventually $y \ll x_n$, and for all IP $P \supseteq I^-(x)$, there is some $z \in P$ such that eventually $z \ll x_n$.

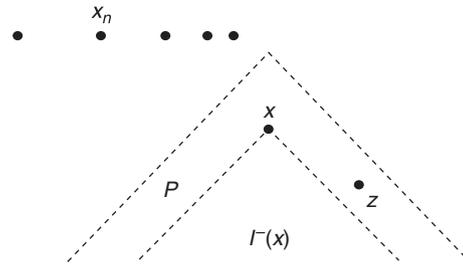


Figure 7 $x \notin \hat{L}(\{x_n\})$: there is some IP $P \supseteq I^-(x)$ such that for all $z \in P$, $z \ll x_n$ for infinitely many n .

in M , the point $I^-[c]$ in $\hat{\partial}(M)$ is the topological endpoint of c in \hat{M} .

5. If $X = \widehat{\mathbb{L}}^n$, then X is homeomorphic to the conformal image of \mathbb{L}^n in \mathbb{E}^n together with \mathfrak{S}^+ and i^+ ; in particular, $\hat{\partial}(\mathbb{L}_n)$ has the topology of a cone.

Examples The future causal boundary with the future-chronological topology can be calculated with a fair degree of success. For instance, if M is conformal to a simple product spacetime $Q \times \mathbb{L}^1$ (Q a Riemannian manifold), then $\hat{\partial}(M)$ is much like $\hat{\partial}(\mathbb{L}^n)$ in that it consists of null or timelike lines factored over a particular boundary construction $\partial(Q)$ on Q , coming together at a single point i^+ (the IP which is all of M); if Q is complete, then these are all null lines, and together they may be called \mathfrak{S}^+ .

The elements of $\partial(Q)$ are defined in terms of the Lipschitz-1 functions on Q known as Busemann functions: if $c: [\alpha, \omega) \rightarrow Q$ is any endless unit-speed curve (typically, $\omega = \infty$), then the Busemann function $b_c: Q \rightarrow \mathbb{R}$ is defined by $b_c(q) = \lim_{s \rightarrow \omega} (s - d(c(s), q))$, where d is the distance function in Q ; this function is either finite for all q or infinite for all q . The set $B(Q)$ of finite Busemann functions has an \mathbb{R} -action defined by $a \cdot b_c = b_{a \cdot c}$, where $(a \cdot c)(s) = c(s + a)$. Then $\partial(Q) = B(Q)/\mathbb{R}$. For any $P \in \hat{\partial}(M)$, the boundary of P , as a subset of $Q \times \mathbb{L}^1 \cong Q \times \mathbb{R}$, is the graph of a Busemann function (the function is b_c for P generated by a null curve projecting to c); and a point $x = (q, t)$ in M can be represented by $\partial(I^-(x))$, which is the graph of the function $t - d(-, q)$. Thus, one could use the function-space topology on $B(Q)$ to topologize \hat{M} ; in that function-space topology $\hat{\partial}(M)$ is a cone on $\partial(Q)$, and \hat{M} , apart from i^+ , is the topological product of \mathbb{R} with $Q \cup \partial(Q)$. The future-chronological topology is sometimes different from the function-space topology, allowing more convergent sequences than the function-space topology does. When this happens, the result is non-Hausdorff, revealing pairs of points in $\hat{\partial}(M)$ which are more closely related to one another than the function-space

topology reveals; but it is still the case that $\hat{\partial}(M)$, apart from i^+ , is fibered by \mathbb{R} over $\partial(Q)$.

If Q is a warped product $Q = (a, b) \times K$ for a compact manifold K with metric $dr^2 + e^{\phi(r)}h$ with h a metric on K , then one can calculate more precisely: if, for instance, ϕ has a minimum in the interior of (a, b) and has suitable growth on either end, then $\partial(Q)$ represents two copies of K (one for each end of $(a, b) \times K$), the future-chronological topology is the same as the function-space topology, and \hat{M} (apart from i^+) is a simple product of \mathbb{R} with $Q \cup \partial(Q)$: $\hat{\partial}(M)$ is precisely a null cone over two copies of K . This applies, for instance, to exterior Schwarzschild, where $K = S^2$; the boundary at one end of exterior Schwarzschild is the usual \mathfrak{S}^+ , and the boundary at the other end is the null cone $\{r = 2m\}$, where exterior attaches to interior Schwarzschild.

Calculations for the future-chronological topology become much easier when $\hat{\partial}(M)$ is purely spacelike, that is, no $P \in \hat{\partial}(M)$ is contained in the past of any other element of \hat{M} . For instance, if M is conformal to a multiwarped product, $Q_1 \times \dots \times Q_m \times (a, b)$ with metric $f_1(t)^2 h_1 + \dots + f_m(t)^2 h_m - dt^2$, where h_i is a Riemannian metric on Q_i , then $\hat{\partial}(M)$ will be purely spacelike if all the Riemannian factors are complete and for each i , $\int_{b^-}^b 1/f_i(t) dt < \infty$; in that case, $\hat{\partial}(M) \cong Q$, where $Q = Q_1 \times \dots \times Q_m$ and $\hat{M} \cong Q \times (a, b)$. This applies, for instance, to interior Schwarzschild, where $Q_1 = \mathbb{R}^1$ and $Q_2 = S^2$, yielding the topology of $\mathbb{R}^1 \times S^2$ for the Schwarzschild singularity.

There is a categorical universality for spacelike boundaries and the future-chronological topology. This means that any other reasonable way of future-completing interior Schwarzschild must yield $\mathbb{R}^1 \times S^2$ or a topological quotient of that for the singularity; and if the result is to be past-distinguishing, $\mathbb{R}^1 \times S^2$ is the only possibility.

Of course, all this can be done in the time-dual fashion, using the past-chronological topology on \check{M} . It would be desirable to combine the future and past causal boundaries with a suitable topology as well as appropriate identifications. There has been some work in that direction.

Causal Boundary: Revisited

Marolf and Ross (2003) have proposed an identification of TIPs and TIFs that relies on the equivalence relation defined by Szabados. For an IP P and IF F , call (P, F) a Szabados pair if $P \subset I^-(x)$ for all $x \in F$, P is maximal among IPs for that property, and dually for F with respect to P . For instance, for any $x \in M$, $(I^-(x), I^+(x))$ is a Szabados pair. The Marolf–Ross version of the causal boundary, $\bar{\partial}(M)$, consists of all Szabados pairs

formed of TIPs and TIFs, plus any TIP or TIF that cannot be paired; this produces an appropriate set of identifications within $\hat{\partial}(M) \cup \bar{\partial}(M)$. The chronology relation on M is extended to $\bar{M} = \bar{\partial}(M) \cup M$ by treating each point x in M as the Szabados pair $(I^-(x), I^+(x))$ and each unpaired IP P as (P, \emptyset) and unpaired IF F as (\emptyset, F) , and then defining $(P, F) \ll (P', F')$ whenever $F \cap P' \neq \emptyset$.

The resulting chronological set is not necessarily either past- or future-distinguishing, but it is (past and future)-distinguishing. The topology they propose places endpoints in $\bar{\partial}(M)$ for all causal curves which are endless in M , but there may be multiple future endpoints for a single future-endless curve. The topology need not be T_1 : points can fail to be closed. For a product spacetime $M = Q \times \mathbb{L}^1$, the Marolf–Ross topology on \bar{M} is always the function-space topology.

As of this writing, there is active research by J L Flores to institute a Marolf–Ross type of identification of $\hat{\partial}(M)$ with $\bar{\partial}(M)$ using a topology that partakes more of the future- and past-chronological topologies.

See also: Asymptotic Structure and Conformal Infinity; Spacetime Topology, Causal Structure and Singularities.

Further Reading

- Budic R and Sachs RK (1974) Causal boundaries for general relativistic space-times. *Journal of Mathematical Physics* 15: 1302–1309.
- García-Parrado A and Senovilla JMM (2003) Causal relationship: a new tool for the causal characterization of Lorentzian manifolds. *Classical and Quantum Gravity* 20: 625–664.
- Geroch RP, Kronheimer EH, and Penrose R (1972) Ideal points in space-time. *Proceedings of the Royal Society of London, Series A* 327: 545–567.
- Harris SG (1998) Universality of the future chronological boundary. *Journal of Mathematical Physics* 39: 5427–5445.
- Harris SG (2000) Topology of the future chronological boundary: universality for spacelike boundaries. *Classical and Quantum Gravity* 17: 551–603.
- Harris SG (2001) Causal boundary for standard static spacetimes. *Nonlinear Analysis* 47: 2971–2981 (Special Edition: Proceedings of the Third World Congress in Nonlinear Analysis).
- Harris SG (2004a) Boundaries on spacetimes: an outline. *Classical and Quantum Gravity* 359: 65–85.
- Harris SG (2004b) Discrete group actions on spacetimes: causality conditions and the causal boundary. *Classical and Quantum Gravity* 21: 1209–1236.
- Harris SG and Dray T (1990) The causal boundary of the trousers space. *Classical and Quantum Gravity* 7: 149–161.
- Hawking SW and Ellis GFR (1973) *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
- Marolf D and Ross SF (2003) A new recipe for causal completions. *Classical and Quantum Gravity* 20: 4085–4118.
- Schmidt BG (1972) Local completeness of the b -boundary. *Communications in Mathematical Physics* 29: 49–54.
- Scott SM and Szekeres P (1994) The abstract boundary – a new approach to singularities of manifolds. *Journal of Geometry and Physics* 13: 223–253.

Szabados LB (1988) Causal boundary for strongly causal spacetimes. *Classical and Quantum Gravity* 5: 121–134.
 Szabados LB (1989) Causal boundary for strongly causal spacetimes: II. *Classical and Quantum Gravity* 6: 77–91.

Wald RM (1984) *General Relativity*. Chicago: University of Chicago Press.

Boundary Conformal Field Theory

J Cardy, Rudolf Peierls Centre for Theoretical Physics, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Boundary conformal field theory (BCFT) is simply the study of conformal field theory (CFT) in domains with a boundary. It gains its significance [1] because, in some ways, it is mathematically simpler: the algebraic and geometric structures of CFT appear in a more straightforward manner; and [2] because it has important applications: in string theory in the physics of open strings and D-branes, and in condensed matter physics in boundary critical behavior and quantum impurity models.

This article, however, describes the basic ideas from the point of view of quantum field theory, without regard to particular applications or to any deeper mathematical formulations.

Review of CFT

Stress Tensor and Ward Identities

Two-dimensional CFTs are massless, local, relativistic renormalized quantum field theories. Usually they are considered in imaginary time, that is, on two-dimensional manifolds with Euclidean signature. In this article, the metric is also taken to be Euclidean, although the formulation of CFTs on general Riemann surfaces is also of great interest, especially for string theory. For the time being, the domain is the entire complex plane.

Heuristically, the correlation functions of such a field theory may be thought of as being given by the Euclidean path integral, that is, as expectation values of products of local densities with respect to a Gibbs measure $Z^{-1} e^{-S_E(\{\psi\})} [d\psi]$, where the $\{\psi(x)\}$ are some set of fundamental local fields, S_E is the Euclidean action, and the normalization factor Z is the partition function. Of course, such an object is not in general well defined, and this picture should be seen only as a guide to formulating the basic principles of CFT which can then be developed into a mathematically consistent theory.

In two dimensions, it is useful to use the so-called complex coordinates $z = x^1 + ix^2$, $\bar{z} = x^1 - ix^2$. In CFT, there are local densities $\phi_j(z, \bar{z})$, called primary fields, whose correlation functions transform covariantly under conformal mappings $z \rightarrow z' = f(z)$:

$$\begin{aligned} &\langle \phi_1(z_1, \bar{z}_1) \phi_2(z_2, \bar{z}_2) \cdots \rangle \\ &= \prod_i f'(z_i)^{h_i} \bar{f}'(\bar{z}_i)^{\bar{h}_i} \langle \phi_1(z'_1, \bar{z}'_1) \phi_2(z'_2, \bar{z}'_2) \cdots \rangle \end{aligned} \quad [1]$$

where (h_j, \bar{h}_j) (usually real numbers, not complex conjugates of each other) are called the conformal weights of ϕ_j . These local fields can in general be normalized so that their two-point functions have the form

$$\langle \phi_j(z_j, \bar{z}_j) \phi_k(z_k, \bar{z}_k) \rangle = \delta_{jk} / (z_j - z_k)^{2h_j} (\bar{z}_j - \bar{z}_k)^{2\bar{h}_j} \quad [2]$$

They satisfy an algebra known as the operator product expansion (OPE)

$$\begin{aligned} &\phi_i(z_1, \bar{z}_1) \cdot \phi_j(z_2, \bar{z}_2) \\ &= \sum_k c_{ijk} (z_1 - z_2)^{-h_i - h_j + h_k} \\ &\quad \times (\bar{z}_1 - \bar{z}_2)^{-\bar{h}_i - \bar{h}_j + \bar{h}_k} \phi_k(z_1, \bar{z}_1) + \cdots \end{aligned} \quad [3]$$

which is supposed to be valid when inserted into higher-order correlation functions in the limit when $|z_1 - z_2|$ is much less than the separations of all the other points. The ellipses denote the contributions of other nonprimary scaling fields to be described below. The structure constants c_{ijk} , along with the conformal weights, characterize the particular CFT.

An essential role is played by the energy-momentum tensor, or, in Euclidean field theory language, the stress tensor $T^{\mu\nu}$. Heuristically, it is defined as the response of the partition function to a local change in the metric:

$$T^{\mu\nu}(x) = -(2\pi) \delta \ln Z / \delta g_{\mu\nu}(x) \quad [4]$$

(the factor of 2π is included so that similar factors disappear in later equations).

The symmetry of the theory under translations and rotations implies that $T^{\mu\nu}$ is conserved, $\partial_\mu T^{\mu\nu} = 0$, and symmetric. Scale invariance implies that it is also traceless $\Theta \equiv T^\mu_\mu = 0$. It should be noted that the vanishing of the trace of the stress tensor for a scale invariant classical field theory does

not usually survive when quantum corrections are taken into account: indeed, $\Theta \propto \beta(g)$, the renormalization group (RG) beta-function. A quantum field theory is thus only a CFT when this vanishes, that is, at an RG fixed point. In complex coordinates, the components $T_{z\bar{z}} = T_{\bar{z}z} = 4\Theta$ vanish, while the conservation equations read

$$\partial_{\bar{z}} T_{zz} = \partial_z T_{\bar{z}\bar{z}} = 0 \tag{5}$$

Thus, correlators of $T(z) \equiv T_{zz}$ are locally analytic (in fact, globally meromorphic) functions of z , while those of $\bar{T}(\bar{z}) \equiv T_{\bar{z}\bar{z}}$ are antianalytic. It is this property of analyticity which makes CFTs tractable in two dimensions.

Since an infinitesimal conformal transformation $z \rightarrow z + \alpha(z)$ induces a change in the metric, its effect on a correlation function of primary fields, given by [1], may also be expressed through an appropriate integral involving an insertion of the stress tensor. This leads to the conformal Ward identity:

$$\int_C \langle T(z) \prod_j \phi_j(z_j, \bar{z}_j) \rangle \alpha(z) dz = \sum_j (h_j \alpha'(z_j) + \alpha(z_j) (\partial/\partial z_j)) \langle \prod_j \phi_j(z_j, \bar{z}_j) \rangle \tag{6}$$

where C is a contour encircling all the points $\{z_j\}$. (A similar equation holds for the insertion of \bar{T} .) Using Cauchy's theorem, this determines the first few terms in the OPE of T with any primary density:

$$T(z) \cdot \phi_j(z_j, \bar{z}_j) = \frac{h_j}{(z - z_j)^2} \phi_j(z_j, \bar{z}_j) + \frac{1}{z - z_j} \partial_z \phi_j(z_j, \bar{z}_j) + O(1) \tag{7}$$

The other, regular, terms in the OPE generate new scaling fields, which are not in general primary, called descendants. One way of defining a density to be primary is by the condition that the most singular term in its OPE with T is a double pole.

The OPE of T with itself has the form

$$T(z) \cdot T(z_1) = \frac{c/2}{(z - z_1)^4} + \frac{2}{(z - z_1)^2} T(z_1) + \dots \tag{8}$$

The first term is present because $\langle T(z)T(z_1) \rangle$ is nonvanishing, and must take the form shown, with c being some number (which cannot be scaled to unity, since the normalization of T is fixed by its definition) which is a property of the CFT. It is known as the conformal anomaly number or the central charge. This term implies that T is not itself primary. In fact, under a finite conformal transformation $z \rightarrow z' = f(z)$,

$$T(z) \rightarrow f'(z)^2 T(z') + \frac{c}{12} \{z', z\} \tag{9}$$

where $\{z', z\} = (f'''f' - \frac{3}{2}f''^2)/f'^2$ is the Schwartzian derivative.

Virasoro Algebra

As with any quantum field theory, the local fields can be realized as linear operators acting on a Hilbert space. In ordinary QFT, it is customary to quantize on a constant-time hypersurface. The generator of infinitesimal time translations is the Hamiltonian \hat{H} , which itself is independent of which time slice is chosen, because of time translational symmetry. It is also given by the integral over the hypersurface of the time-time component of the stress tensor. In CFT, because of scale invariance, one may instead quantize on fixed circle of a given radius. The analog of the Hamiltonian is the dilatation operator \hat{D} , which generates scale transformations. Unlike \hat{H} , the spectrum of \hat{D} is usually discrete, even in an infinite system. It may also be expressed as an integral over the radial component of the stress tensor:

$$\begin{aligned} \hat{D} &= \frac{1}{2\pi} \int_0^{2\pi} r \hat{T}_{rr} r d\theta \\ &= \frac{1}{2\pi i} \int_C z \hat{T}(z) dz - \frac{1}{2\pi i} \int_C \bar{z} \hat{\bar{T}}(\bar{z}) d\bar{z} \\ &\equiv \hat{L}_0 + \hat{\bar{L}}_0 \end{aligned} \tag{10}$$

where, because of analyticity, C can be any contour encircling the origin.

This suggests that one define other operators

$$\hat{L}_n \equiv \frac{1}{2\pi} \int_C z^{n+1} \hat{T}(z) dz \tag{11}$$

and similarly the $\hat{\bar{L}}_n$. From the OPE [8] then follows the Virasoro algebra \mathcal{V} :

$$[\hat{L}_n, \hat{L}_m] = (n - m) \hat{L}_{n+m} + \frac{c}{12} n(n^2 - 1) \delta_{n+m,0} \tag{12}$$

with an isomorphic algebra $\bar{\mathcal{V}}$ generated by the $\hat{\bar{L}}_n$.

In radial quantization, there is a vacuum state $|0\rangle$. Acting on this with the operator corresponding to a scaling field gives a state $|\phi_j\rangle \equiv \hat{\phi}_j(0,0)|0\rangle$ which is an eigenstate of \hat{D} : in fact,

$$\hat{L}_0 |\phi_j\rangle = h_j |\phi_j\rangle, \quad \hat{\bar{L}}_0 |\phi_j\rangle = \bar{h}_j |\phi_j\rangle \tag{13}$$

From the OPE [7], one sees that $|\mathcal{L}_n \phi_j\rangle \propto \hat{L}_n |\phi_j\rangle$, and, if ϕ_j is primary, $\hat{L}_n |\phi_j\rangle = 0$ for all $n \geq 1$.

The states corresponding to a given primary field, and those generated by acting on these with all the \hat{L}_n with $n < 0$ an arbitrary number of times, form a

highest-weight representation of \mathcal{V} . However, this is not necessarily irreducible. There may be null vectors, which are linear combinations of states at a given level which are themselves annihilated by all the \hat{L}_n with $n > 0$. They exist whenever h takes a value from the Kac table:

$$h = h_{r,s} = \frac{(r(m+1) - sm)^2 - 1}{4m(m+1)} \quad [14]$$

with the central charge parametrized as $c = 1 - 6/(m(m+1))$, and r, s are non-negative integers. These null states should be projected out, giving an irreducible representation \mathcal{V}_b .

The full Hilbert space of the CFT is then

$$\mathcal{H} = \bigoplus_{b, \bar{b}} n_{b, \bar{b}} \mathcal{V}_b \otimes \bar{\mathcal{V}}_{\bar{b}} \quad [15]$$

where the non-negative integers $n_{b, \bar{b}}$ specify how many distinct primary fields of weights (b, \bar{b}) there are in the CFT.

The consistency of the OPE [3] with the existence of null vectors leads to the fusion algebra of the CFT. This applies separately to the holomorphic and antiholomorphic sectors, and determines how many copies of \mathcal{V}_c occur in the fusion of \mathcal{V}_a and \mathcal{V}_b :

$$\mathcal{V}_a \odot \mathcal{V}_b = \sum_c N_{ab}^c \mathcal{V}_c \quad [16]$$

where the N_{ab}^c are non-negative integers.

A particularly important subset of all CFTs consists of the minimal models. These have rational central charge $c = 1 - 6(p - q)^2/pq$, in which case the fusion algebra closes with a finite number of possible values $1 \leq r \leq q, 1 \leq s \leq p$ in the Kac formula [14]. For these models, the fusion algebra takes the form

$$\mathcal{V}_{r_1, s_1} \odot \mathcal{V}_{r_2, s_2} = \sum_{r=|r_1-r_2|}^{r_1+r_2-1'} \sum_{s=|s_1-s_2|}^{s_1+s_2-1'} \mathcal{V}_{r, s} \quad [17]$$

where the prime on the sums indicates that they are to be restricted to the allowed intervals of r and s .

There is an important theorem which states that the only unitary CFTs with $c < 1$ are the minimal models with $p/q = (m + 1)/m$, where m is an integer ≥ 3 .

Modular Invariance

The fusion algebra limits which values of (b, \bar{b}) might appear in a consistent CFT, but not which ones actually occur, that is, the values of the $n_{b, \bar{b}}$. This is answered by the requirement of modular invariance on the torus. First consider the theory on an infinitely long cylinder, of unit circumference.

This is related to the (punctured) plane by the conformal mapping $z \rightarrow (1/2\pi) \ln z \equiv t + ix$. The result is a QFT on the circle $0 \leq x < 1$, in imaginary time t . The generator of infinitesimal time translations is related to that for dilatations in the plane:

$$\begin{aligned} \hat{H} &= 2\pi \hat{D} - \frac{\pi c}{6} \\ &= 2\pi(\hat{L}_0 + \hat{\bar{L}}_0) - \frac{\pi c}{6} \end{aligned} \quad [18]$$

where the last term comes from the Schwartzian derivative in [9]. Similarly, the generator of translations in x , the total momentum operator, is $\hat{P} = 2\pi(\hat{L}_0 - \hat{\bar{L}}_0)$.

A general torus is, up to a scale transformation, a parallelogram with vertices $(0, 1, \tau, 1 + \tau)$ in the complex plane, with the opposite edges identified. We can make this by taking a cylinder of unit circumference and length $\text{Im} \tau$, twisting the ends by a relative amount $\text{Re} \tau$, and sewing them together. This means that the partition function of the CFT on the torus can be written as

$$\begin{aligned} Z(\tau, \bar{\tau}) &= \text{tr} e^{-(\text{Im} \tau) \hat{H} + i(\text{Re} \tau) \hat{P}} \\ &= \text{tr} q^{\hat{L}_0 - c/24} \bar{q}^{\hat{\bar{L}}_0 - c/24} \end{aligned} \quad [19]$$

using the above expressions for \hat{H} and \hat{P} and introducing $q \equiv e^{2\pi i \tau}$.

Through the decomposition [15] of \mathcal{H} , the trace sum can be written as

$$Z(\tau, \bar{\tau}) = \sum_{b, \bar{b}} n_{b, \bar{b}} \chi_b(q) \chi_{\bar{b}}(\bar{q}) \quad [20]$$

where

$$\chi_b(q) \equiv \text{tr}_{\mathcal{V}_b} q^{\hat{L}_0 - c/24} = \sum_N d_b(N) q^{b - (c/24) + N} \quad [21]$$

is the character of the representation of highest weight h , which counts the degeneracy $d_b(N)$ at level N . It is purely an algebraic property of the Virasoro algebra, and its explicit form is known in many cases.

All of this would be less interesting were it not for the observation that the parametrization of the torus through τ is not unique. In fact, the transformations $S: \tau \rightarrow -1/\tau$ and $T: \tau \rightarrow \tau + 1$ give the same torus (see Figure 1). Together, these

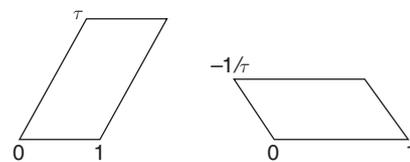


Figure 1 Two equivalent parametrizations of the same torus.

operations generate the modular group $SL(2, \mathbf{Z})$, and the partition function $Z(\tau, \bar{\tau})$ should be invariant under them. T -invariance is simply implemented by requiring that $h - \bar{h}$ is an integer, but the S -invariance of the right-hand side of [20] places highly nontrivial constraints on the $n_{h, \bar{h}}$. That this can be satisfied at all relies on the remarkable property of the characters that they transform linearly under S :

$$\chi_b(e^{-2\pi i/\tau}) = \sum_{b'} S_b^{b'} \chi_{b'}(e^{2\pi i\tau}) \quad [22]$$

This follows from applying the Poisson sum formula to the explicit expressions for the characters, which are related to Jacobi theta-functions. In many cases (e.g., the minimal models) this representation is finite dimensional, and the matrix S is symmetric and orthogonal. This means that one can immediately obtain a modular invariant partition function by forming the diagonal sum

$$Z = \sum_b \chi_b(q) \chi_b(\bar{q}) \quad [23]$$

so that $n_{b, \bar{b}} = \delta_{b\bar{b}}$. However, because of various symmetries of the characters, other modular invariants are possible: for the minimal models (and some others) these have been classified. Because of an analogy of the results with the classification of semisimple Lie algebras, the diagonal invariants are called the A-series.

Boundary CFT

In any field theory in a domain with a boundary, one needs to consider how to impose a set of consistent boundary conditions. Since CFT is formulated independently of a particular set of fundamental fields and a Lagrangian, this must be done in a more general manner. A natural requirement is that the off-diagonal component $T_{\parallel\perp}$ of the stress tensor parallel/perpendicular to the boundary should vanish. This is called the conformal boundary condition. If the boundary is parallel to the time axis, it implies that there is no momentum flow across the boundary. Moreover, it can be argued that, under the RG, any uniform boundary condition will flow into a conformally invariant one. For a given bulk CFT, however, there may be many possible distinct such boundary conditions, and it is one task of BCFT to classify these.

To begin with, take the domain to be the upper-half plane, so that the boundary is the real axis. The conformal boundary condition then implies that $T(z) = \bar{T}(\bar{z})$ when z is on the real axis. This has the immediate consequence that correlators of \bar{T} are those of T , analytically continued into the lower-

half plane. The conformal Ward identity, cf. [7], now reads

$$\begin{aligned} & \left\langle T(z) \prod_j \phi_j(z_j, \bar{z}_j) \right\rangle \\ &= \sum_j \left(\frac{h_j}{(z - z_j)^2} + \frac{1}{z - z_j} \partial_{z_j} \right. \\ & \quad \left. + \frac{\bar{h}_j}{(\bar{z} - \bar{z}_j)^2} + \frac{1}{\bar{z} - \bar{z}_j} \partial_{\bar{z}_j} \right) \left\langle \prod_j \phi_j(z_j, \bar{z}_j) \right\rangle \quad [24] \end{aligned}$$

In radial quantization, in order that the Hilbert spaces defined on different hypersurfaces be equivalent, one must choose semicircles centered on some point on the boundary, conventionally the origin. The dilatation operator is now

$$\hat{D} = \frac{1}{2\pi i} \int_S z \hat{T}(z) dz - \frac{1}{2\pi i} \int_S \bar{z} \hat{\bar{T}}(\bar{z}) d\bar{z} \quad [25]$$

where S is a semicircle. Using the conformal boundary condition, this can also be written as

$$\hat{D} = \hat{L}_0 = \frac{1}{2\pi i} \int_C z \hat{T}(z) dz \quad [26]$$

where C is a complete circle around the origin. As before, one may similarly define the \hat{L}_n , and they satisfy a Virasoro algebra.

Note that there is now only one Virasoro algebra. This is related to the fact that conformal mappings which preserve the real axis correspond to real analytic functions. The eigenstates of \hat{L}_0 correspond to boundary operators $\hat{\phi}_j(0)$ acting on the vacuum state $|0\rangle$. It is well known that in a renormalizable QFT operators at the boundary require a different renormalization from those in the bulk, and this will in general lead to a different set of conformal weights. It is one of the tasks of BCFT to determine these, for a given allowed boundary condition.

However, there is one feature unique to boundary CFT in two dimensions. Radial quantization also makes sense, leading to the same form [26] for the dilation operator, if the boundary conditions on the negative and positive real axes are different. As far as the structure of BCFT goes, correlation functions with this mixed boundary condition behave as though a local scaling field were inserted at the origin. This has led to the term ‘‘boundary condition changing (bcc) operator,’’ but it must be stressed that these are not local operators in the conventional sense.

The Annulus Partition Function

Just as consideration of the partition function on the torus illuminates the bulk operator content $n_{h, \bar{h}}$, it

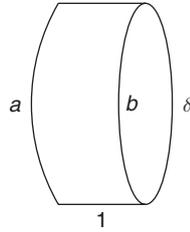


Figure 2 The annulus, with boundary conditions a and b on either boundary.

turns out that consistency on the annulus helps classify both the allowed boundary conditions, and the boundary operator content. To this end, consider a CFT in an annulus formed of a rectangle of unit width and height δ , with the top and bottom edges identified (see [Figure 2](#)). The boundary conditions on the left and right edges, labeled by a, b, \dots , may be different. The partition function with boundary conditions a and b on either edge is denoted by $Z_{ab}(\delta)$.

One way to compute this is by first considering the CFT on an infinitely long strip of unit width. This is conformally related to the upper-half plane (with an insertion of bcc operators at 0 and ∞ if $a \neq b$) by the mapping $z \rightarrow (1/\pi)\ln z$. The generator of infinitesimal translations along the strip is

$$\hat{H}_{ab} = \pi\hat{D} - \pi c/24 = \pi\hat{L}_0 - \pi c/24 \quad [27]$$

Thus, for the annulus,

$$Z_{ab}(\delta) = \text{tr} e^{-\delta\hat{H}_{ab}} = \text{tr} q^{\hat{L}_0 - \pi c/24} \quad [28]$$

with $q \equiv e^{-\pi\delta}$. As before, this can be decomposed into characters:

$$Z_{ab}(\delta) = \sum_b n_{ab}^b \chi_b(q) \quad [29]$$

but note that now the expression is linear. The non-negative integers n_{ab}^b give the operator content with the boundary conditions (ab): the lowest value of h with $n_{ab}^b > 0$ gives the conformal weight of the bcc operator, and the others give conformal weights of the other allowed primary fields which may also sit at this point.

On the other hand, the annulus partition function may be viewed, up to an overall rescaling, as the path integral for a CFT on a circle of unit circumference, being propagated for (imaginary) time δ^{-1} . From this point of view, the partition function is no longer a trace, but rather the matrix element of $e^{-\hat{H}/\delta}$ between boundary states:

$$Z_{ab}(\delta) = \langle a | e^{-\hat{H}/\delta} | b \rangle \quad [30]$$

Note that \hat{H} is the same Hamiltonian that appears in [\[18\]](#), and the boundary states lie in \mathcal{H} , [\[15\]](#).

How are these boundary states to be characterized? Using the transformation law [\[9\]](#) the conformal boundary condition applied to the circle implies that $L_n = \bar{L}_{-n}$. This means that any boundary state $|B\rangle$ lies in the subspace satisfying

$$\hat{L}_n |B\rangle = \hat{\bar{L}}_{-n} |B\rangle \quad [31]$$

Moreover, because of the decomposition [\[15\]](#) of \mathcal{H} , $|B\rangle$ is also some linear superposition of states from $\mathcal{V}_b \otimes \bar{\mathcal{V}}_{\bar{b}}$. This condition can therefore be applied in each subspace. Taking $n=0$ in [\[31\]](#) constrains $\bar{b}=b$. For simplicity, consider only the diagonal CFTs with $n_{b,\bar{b}} = \delta_{b,\bar{b}}$. It can then be shown that the solution of [\[31\]](#) is unique and has the following form. The subspace at level N of \mathcal{V}_b has dimension $d_b(N)$. Denote an orthonormal basis by $|b, N; j\rangle$, with $1 \leq j \leq d_b(N)$, and the same basis for $\bar{\mathcal{V}}_b$ by $|\bar{b}, N; j\rangle$. The solution to [\[31\]](#) in this subspace is then

$$|b\rangle \equiv \sum_{N=0}^{\infty} \sum_{j=1}^{d_b(N)} |b, N; j\rangle \otimes \overline{|\bar{b}, N; j\rangle} \quad [32]$$

These are called Ishibashi states. Matrix elements of the translation operator along the cylinder between them are simple:

$$\begin{aligned} \langle\langle b' | e^{-\hat{H}/\delta} | b \rangle\rangle &= \sum_{N'=0}^{\infty} \sum_{j'=1}^{d_{b'}(N')} \sum_{N=0}^{\infty} \sum_{j=1}^{d_b(N)} \langle b', N'; j' | \\ &\otimes \overline{\langle \bar{b}', N'; j' |} e^{-(2\pi/\delta)(\hat{L}_0 + \hat{\bar{L}}_0 - c/12)} \end{aligned} \quad [33]$$

$$\begin{aligned} &|b, N; j\rangle \otimes \overline{|\bar{b}, N; j\rangle} \\ &= \delta_{b'b} \sum_{N=0}^{\infty} \sum_{j=1}^{d_b(N)} e^{-(4\pi/\delta)(h+N-(c/24))} \end{aligned} \quad [34]$$

$$= \delta_{b'b} \chi_b(e^{-4\pi/\delta}) \quad [35]$$

Note that the characters which appear are related to those in [\[29\]](#) by the modular transformation S .

The physical boundary states satisfying [\[29\]](#), sometimes called the Cardy states, are linear combinations of the Ishibashi states:

$$|a\rangle = \sum_b \langle\langle b | a | b \rangle\rangle \quad [36]$$

Equating the two different expressions [\[29\]](#) and [\[30\]](#) for Z_{ab} , and using the modular transformation law

[22] and the linear independence of the characters gives the (equivalent) conditions:

$$n_{ab}^b = \sum_{b'} S_{b'}^b \langle a|b'\rangle \langle\langle b'|b\rangle\rangle \quad [37]$$

$$\langle a|b'\rangle \langle\langle b'|b\rangle\rangle = \sum_b S_b^{b'} n_{ab}^b \quad [38]$$

These are called the Cardy conditions. The requirements that the right-hand side of [37] should give a non-negative integer, and that the right-hand side of [38] should factorize in a and b , give highly nontrivial constraints on the allowed boundary states and their operator content.

For the diagonal CFTs considered here (and for the nondiagonal minimal models) a complete solution is possible. It can be shown that the elements S_0^b of S are all non-negative, so one may choose $\langle\langle b|\tilde{0}\rangle\rangle = (S_0^b)^{1/2}$. This defines a boundary state

$$|\tilde{0}\rangle \equiv \sum_b (S_0^b)^{1/2} |b\rangle\rangle \quad [39]$$

and a corresponding boundary condition such that $n_{00}^b = \delta_{b0}$. Then, for each $b' \neq 0$, one may define a boundary state

$$\langle\langle b|\tilde{h}'\rangle\rangle \equiv S_{b'}^b / (S_0^b)^{1/2} \quad [40]$$

From [37], this gives $n_{b'0}^b = \delta_{b'b}$. For each allowed b' in the torus partition function, there is therefore a boundary state $|\tilde{h}'\rangle\rangle$ satisfying the Cardy conditions. However, there is a further requirement:

$$n_{b'b''}^b = \frac{S_{b'}^b S_{b''}^b}{S_0^b} \quad [41]$$

should be a non-negative integer. Remarkably, this combination of elements of S occurs in the Verlinde formula, which follows from considering consistency of the CFT on the torus. This states that the right-hand side of [41] is equal to the fusion algebra coefficient $N_{b'b''}^b$. Since these are non-negative integers, the consistency of the above ansatz for the boundary states is consistent.

We conclude that, at least for the diagonal models, there is a bijection between the allowed primary fields in the bulk CFT and the allowed conformally invariant boundary conditions. For the minimal models, with a finite number of such primary fields, this correspondence has been followed through explicitly.

Example The simplest example is the diagonal $c = \frac{1}{2}$ unitary CFT corresponding to $m = 3$. The allowed values of the conformal weights are $h = 0, \frac{1}{2}, \frac{1}{16}$, and

$$S = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} \quad [42]$$

from which one finds the allowed boundary states

$$|\tilde{0}\rangle = \frac{1}{\sqrt{2}} |0\rangle\rangle + \frac{1}{\sqrt{2}} \left| \frac{1}{2} \right\rangle\rangle + \frac{1}{2^{1/4}} \left| \frac{1}{16} \right\rangle\rangle \quad [43]$$

$$\left| \frac{\tilde{1}}{2} \right\rangle\rangle = \frac{1}{\sqrt{2}} |0\rangle\rangle + \frac{1}{\sqrt{2}} \left| \frac{1}{2} \right\rangle\rangle - \frac{1}{2^{1/4}} \left| \frac{1}{16} \right\rangle\rangle \quad [44]$$

$$\left| \frac{\tilde{1}}{16} \right\rangle\rangle = |0\rangle\rangle - \left| \frac{1}{2} \right\rangle\rangle \quad [45]$$

The nontrivial part of the fusion algebra of this CFT is

$$\mathcal{V}_{\frac{1}{16}} \odot \mathcal{V}_{\frac{1}{16}} = \mathcal{V}_0 + \mathcal{V}_{\frac{1}{2}} \quad [46]$$

$$\mathcal{V}_{\frac{1}{16}} \odot \mathcal{V}_{\frac{1}{2}} = \mathcal{V}_{\frac{1}{16}} \quad [47]$$

$$\mathcal{V}_{\frac{1}{2}} \odot \mathcal{V}_{\frac{1}{2}} = \mathcal{V}_0 \quad [48]$$

from which can be read off the boundary operator content

$$n_b^b = 1 \quad n_{\frac{1}{16}\frac{1}{16}}^0 = n_{\frac{1}{16}\frac{1}{16}}^{\frac{1}{2}} = n_{\frac{1}{16}\frac{1}{16}}^{\frac{1}{16}} = n_{\frac{1}{16}\frac{1}{16}}^{\frac{1}{16}} = 1 \quad [49]$$

The $c = \frac{1}{2}$ CFT is known to describe the continuum limit of the critical Ising model, in which spins $s = \pm 1$ are localized on the sites of a regular lattice. The above boundary conditions may be interpreted as the continuum limit of the lattice boundary conditions $s = 1$, free and $s = -1$, respectively. Note there is a symmetry of the fusion rules which means that one could equally well have inverted the ordering of this correspondence.

Other Topics

Boundary Entropy

The partition function on annulus of length L and circumference β can be thought of as the quantum statistical mechanics partition function for a one-dimensional QFT in an interval of length L , at temperature β^{-1} . It is interesting to consider this in the thermodynamic limit when $\delta = L/\beta$ is large. In that case, only the ground state of \hat{H} contributes in [30], giving

$$Z_{ab}(L, \beta) \sim \langle a|0\rangle \langle 0|b\rangle e^{\pi c L / 6\beta} \quad [50]$$

from which the free energy $F_{ab} = -\beta^{-1} \ln Z_{ab}$ and the entropy $\mathcal{S}_{ab} = -\beta^2 (\partial F_{ab} / \partial \beta)$ can be obtained. The result is

$$\mathcal{S}_{ab} = (\pi c / 3\beta) L + s_a + s_b + o(1) \quad [51]$$

where the first term is the usual extensive contribution. The other two pieces $s_a \equiv \ln(\langle a|0\rangle)$ and $s_b \equiv \ln(\langle b|0\rangle)$ may be identified as the boundary entropy associated with the corresponding boundary states. A similar definition may be made in massive QFTs. It is an unproven but well-verified conjecture that the boundary entropy is a nonincreasing function along boundary RG flows, and is stationary only for conformal boundary states.

Bulk–Boundary OPE

The boundary Ward identity [24] has the implication that, from the point of view of the dependence of its correlators on z_j and \bar{z}_j , a primary field $\phi_j(z_j, \bar{z}_j)$ may be thought of as the product of two local fields which are holomorphic functions of z_j and \bar{z}_j , respectively. These will satisfy OPEs as $|z_j - \bar{z}_j| \rightarrow 0$, with the appearance of primary fields on the right-hand side being governed by the fusion rules. These fields are localized on the real axis: they are the boundary operators. There is therefore a kind of bulk–boundary OPE:

$$\phi_j(z_j, \bar{z}_j) = \sum_k d_{jk} (\text{Im } z_j)^{-h_j - \bar{h}_j + h_k} \phi_k^b(\text{Re } z_j) \quad [52]$$

where the sum on the right-hand side is, in principle, over all the boundary fields consistent with the boundary condition, and the coefficients d_{jk} are analogous to the OPE coefficients in the bulk. As before, they are nonvanishing only if allowed by the fusion algebra: a boundary field of conformal weight h_k is allowed only if $N_{h_j, \bar{h}_j}^{h_k} > 0$.

For example, in the $c = \frac{1}{2}$ CFT, the bulk operator with $h = \bar{h} = \frac{1}{16}$ goes over into the boundary operator with $h = 0$, or that with $h = \frac{1}{2}$, depending on the boundary condition. The bulk operator with $h = \bar{h} = \frac{1}{2}$, however, can only go over into the identity boundary operator with $h = 0$ (or a descendent thereof.)

The fusion rules also apply to the boundary operators themselves. The consistency of these with bulk–boundary and bulk–bulk fusion rules, as well as the modular properties of partition functions, was examined by Lewellen.

Extended Algebras

CFTs may contain other conserved currents apart from the stress tensor, which generate algebras (Kac–Moody, superconformal, W -algebras) which extend the Virasoro algebra. In BCFT, in addition to the conformal boundary condition, it is possible (but not necessary) to impose further boundary conditions relating the holomorphic and antiholomorphic parts of the other currents on the boundary. It is believed that all rational CFTs can be obtained from

Kac–Moody algebras via the coset construction. The classification of boundary conditions from this point of view is fruitful and also important for applications, but is beyond the scope of this article.

Stochastic Loewner Evolution

In recent years, there has emerged a deep connection between BCFT and conformally invariant measures on curves in the plane which start at a boundary of a domain. These arise naturally in the continuum limit of certain statistical mechanics models. The measure is constructed dynamically as the curve is extended, using a sequence of random conformal mappings called stochastic Loewner evolution (SLE). In CFT, the point where the curve begins can be viewed as the insertion of a boundary operator. The requirement that certain quantities should be conserved in mean under the stochastic process is then equivalent to this operator having a null state at level two. Many of the standard results of CFT correspond to an equivalent property of SLE.

Acknowledgments

This article was written while the author was a member of the Institute for Advanced Study. He thanks the School of Mathematics and the School of Natural Sciences for their hospitality. The work was supported by the Ellentuck Fund.

See also: Affine Quantum Groups; Eight Vertex and Hard Hexagon Models; Indefinite Metric; Operator Product Expansion in Quantum Field Theory; Quantum Phase Transitions; Stochastic Loewner Evolutions; String Field Theory; Superstring Theories; Symmetries in Quantum Field Theory: Algebraic Aspects; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras.

Further Reading

- Affleck I (1997) Boundary condition changing operators in conformal field theory and condensed matter physics. *Nuclear Physics B Proceedings Supplement* 58: 35.
- Cardy J (1984) Conformal invariance and surface critical behavior. *Nuclear Physics B* 240: 514–532.
- Cardy J (1989) Boundary conditions, fusion rules and the Verlinde formula. *Nuclear Physics B* 324: 581.
- di Francesco P, Mathieu P, and Senechal D (1999) *Conformal Field Theory*. New York: Springer.
- Kager W and Nienhuis B (2004) A guide to stochastic Loewner evolution and its applications. *Journal of Statistical Physics* 115: 1149.
- Lawler G (2005) *Conformally Invariant Processes in the Plane*. American Mathematical Society.
- Lewellen DC (1992) Sewing constraints for conformal field theories on surfaces with boundaries. *Nuclear Physics B* 372: 654.
- Petkova V and Zuber JB *Conformal Boundary Conditions and What They Teach Us*, Lectures given at the Summer School and Conference on Nonperturbative Quantum Field Theoretic

Methods and their Applications, August 2000, Budapest, Hungary, hep-th/0103007.
 Verlinde E (1988) Fusion rules and modular transformations in 2D conformal field theory. *Nuclear Physics B* 300: 360.

Werner W *Random Planar Curves and Schramm–Loewner Evolutions*, Springer Lecture Notes (to appear), math.PR/0303354.

Boundary Control Method and Inverse Problems of Wave Propagation

M I Belishev, Petersburg Department of Steklov Institute of Mathematics, St. Petersburg, Russia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Inverse problems are generally positioned as the problems of determination of a system (its structure, parameters, etc.) from its “input → output” correspondence.

The boundary-value inverse problems deal with systems which describe processes (wave, heat, electromagnetic ones, etc.) occurring in media occupying a spatial domain. The process is initiated by a boundary source (input) and is described by a solution of a certain partial differential equation in the domain. Certain additional information about the solution, which can be extracted from measurements on the boundary, plays the role of the output. The objective is to determine the parameters of the medium – in particular, the coefficients in the equation – from this information.

The boundary control (BC) method (Belishev 1986) is an approach to the boundary-value inverse problems based on their links with the control theory and system theory. The present article is a version of the BC method which solves the problem of reconstruction of a Riemannian manifold from its boundary spectral or dynamical data.

Forward Problems

Manifold

Let (Ω, d) be a smooth compact Riemannian manifold with the boundary Γ , $\dim \Omega \geq 2$; d is the distance determined by the metric tensor g . For $A \subset \Omega$ denote

$$\langle A \rangle^r := \{x \in \Omega \mid d(x, A) \leq r\}, \quad r \geq 0$$

the hypersurfaces $\Gamma^T := \{x \in \Omega \mid d(x, \Gamma) = T\}$, $T > 0$ are equidistant to Γ . In terms of the dynamics of the system, the value

$$T_* := \min\{T > 0 \mid \langle \Gamma \rangle^T = \Omega\} = \max_{\Omega} d(\cdot, \Gamma)$$

means the time needed for waves, moving from Γ with the unit speed, to fill Ω .

A point $x \in \Omega$ is said to belong to the set $c_0 \subset \Omega$ if x is connected with Γ via more than one shortest geodesic. The set $c := \bar{c}_0$ is called the separation set (cut locus) of Ω with respect to Γ . It is a closed set of zero volume. Let $\tau_*(\gamma)$ be the length of the geodesic emanating from $\gamma \in \Gamma$ orthogonally to Γ and connecting γ with c . The function $\tau_*(\cdot)$ is continuous on Γ .

For $x \in \Omega \setminus c$ the pair (γ, τ) , such that $\tau = d(x, \Gamma) = d(x, \gamma)$, constitutes the semigeodesic coordinates of x . The set of these coordinates

$$\Theta := \{(\gamma, \tau) \mid \gamma \in \Gamma, 0 \leq \tau < \tau_*(\gamma)\} \subset \Gamma \times [0, T_*]$$

is called the pattern of Ω . Pictorially, to get the pattern, one needs to slit Ω along c and then pull it on the cylinder $\Gamma \times [0, T_*]$. The part $\Theta^T := \Theta \cap (\Gamma \times [0, T])$ of the pattern consists of the semigeodesic coordinates of the points $x \in \langle \Gamma \rangle^T \setminus c$ (Figure 1).

Dynamical System

Propagation of waves in the manifold is described by a dynamical system α^T of the form

$$u_{tt} - \Delta_g u = b \quad \text{in } \Omega \times (0, T) \quad [1]$$

$$u|_{t=0} = u_t|_{t=0} = 0 \quad \text{in } \Omega \quad [2]$$

$$u = f \quad \text{on } \Gamma \times [0, T] \quad [3]$$

where Δ_g is the Beltrami–Laplace operator, $0 < T \leq \infty$, f and b are the boundary and volume sources (controls), $u = u^{f,b}(x, t)$ is the solution (wave).

Set $\mathcal{H} := L_2(\Omega)$; the spaces of the controls are

$$\mathcal{F}^T := L_2(\Gamma \times [0, T]), \quad \mathcal{G}^T := L_2([0, T]; \mathcal{H})$$

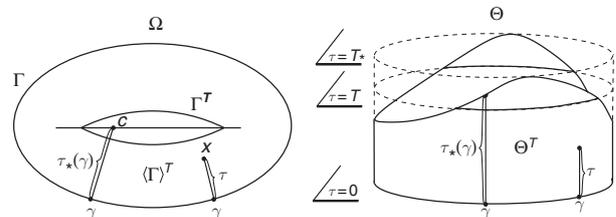


Figure 1 Manifold and pattern. (Data from Belishev (1997).)

The “input \mapsto state” map of the system α^T is realized by the control operator W^T :

$$\mathcal{F}^T \times \mathcal{G}^T \rightarrow \mathcal{H}, \quad W^T\{f, b\} := u^{f,b}(\cdot, T)$$

and its parts

$$\begin{aligned} W_{\text{bd}}^T : \mathcal{F}^T &\rightarrow \mathcal{H}, & W_{\text{vol}}^T : \mathcal{G}^T &\rightarrow \mathcal{H} \\ W_{\text{bd}}^T f &:= u^{f,0}(\cdot, T), & W_{\text{vol}}^T b &:= u^{0,b}(\cdot, T) \end{aligned}$$

In the case $f=0$ the evolution of the system is governed by the operator $L := -\Delta_g$ defined on the Sobolev class $H^2(\Omega) \cap H_0^1(\Omega)$ of functions vanishing on Γ , and the semigroup representation

$$\begin{aligned} u^{0,b}(\cdot, r) &= W_{\text{vol}}^r b \\ &= \int_0^r L^{-1/2} \sin[(r-t)L^{1/2}] b(\cdot, t) dt \end{aligned} \quad [4]$$

holds for all $r \geq 0$.

The “input \mapsto output” map is implemented by the response operator $R^T : \mathcal{F}^T \rightarrow \mathcal{F}^T$,

$$R^T f := \partial_\nu u^{f,0} \quad \text{on } \Gamma \times [0, T]$$

defined on controls $f \in H^1(\Gamma \times [0, T])$ vanishing on $\Gamma \times \{t=0\}$; here $\nu = \nu(\gamma)$ is the outward normal to Γ . The normal derivative $\partial_\nu u^{f,0}$ describes the forces appearing on Γ as a result of interaction of the wave with the boundary.

The map $C^T : \mathcal{F}^T \rightarrow \mathcal{F}^T$, $C^T := (W_{\text{bd}}^T)^* W_{\text{bd}}^T$, which is called the connecting operator, can be represented via the response operator of the system α^{2T} :

$$C^T = \frac{1}{2}(S^T)^* R^{2T} J^{2T} S^T \quad [5]$$

$S^T : \mathcal{F}^T \rightarrow \mathcal{F}^{2T}$ being the extension of controls from $\Gamma \times [0, T]$ onto $\Gamma \times [0, 2T]$ as odd functions of t with respect to $t=T$, and $J^{2T} : \mathcal{F}^{2T} \rightarrow \mathcal{F}^{2T}$ being the integration

$$(J^{2T} f)(\cdot, t) = \int_0^t f(\cdot, s) ds$$

Controllability

Open subsets $\sigma \subset \Gamma$ and $\omega \subset \Omega$ determine the subspaces

$$\begin{aligned} \mathcal{F}_\sigma^T &:= \{f \in \mathcal{F}^T \mid \text{supp } f \subset \bar{\sigma} \times [0, T]\} \\ \mathcal{G}_\omega^T &:= \{b \in \mathcal{G}^T \mid \text{supp } b \subset \bar{\omega} \times [0, T]\} \end{aligned}$$

of controls acting from σ and ω , respectively. In view of hyperbolicity of the problem [1]–[3], the relation

$$\text{supp } u^{f,b}(\cdot, t) \subset \langle \bar{\sigma} \rangle^t \cup \langle \bar{\omega} \rangle^t, \quad t \geq 0 \quad [6]$$

holds for $f \in \mathcal{F}_\sigma^T$ and $b \in \mathcal{G}_\omega^T$. This means that the waves propagate in Ω with the speed = 1.

The sets of waves

$$\mathcal{U}_\sigma^T := W_{\text{bd}}^T \mathcal{F}_\sigma^T, \quad \mathcal{U}_\omega^T := W_{\text{vol}}^T \mathcal{G}_\omega^T$$

are said to be reachable at time $t=T$ from σ and ω , respectively. Denoting

$$\mathcal{H}A := \{y \in \mathcal{H} \mid \text{supp } y \subset \bar{A}\}$$

by virtue of [6] one has the embeddings $\mathcal{U}_\sigma^T \subset \mathcal{H}\langle \bar{\sigma} \rangle^T$ and $\mathcal{U}_\omega^T \subset \mathcal{H}\langle \bar{\omega} \rangle^T$. The property of the system α^T that plays the key role in inverse problems is that these embeddings are dense:

$$\text{cl } \mathcal{U}_\sigma^T = \mathcal{H}\langle \bar{\sigma} \rangle^T, \quad \text{cl } \mathcal{U}_\omega^T = \mathcal{H}\langle \bar{\omega} \rangle^T \quad [7]$$

for any $T > 0$ (cl denotes the closure in \mathcal{H}).

In control theory, relations [7] are interpreted as an approximate controllability of the system in subdomains filled with waves; the name “BC method” is derived from the first one (boundary controllability). This property means that the sets of waves are rich enough: any function supported in the subdomain $\langle \bar{\sigma} \rangle^T$ reachable for waves excited on σ can be approximated with any precision in \mathcal{H} -norm by the wave $u^{f,0}(\cdot, T)$ due to appropriate choice of the control f acting from σ . The proof of [7] relies on the fundamental Holmgren–John–Tataru unique continuation theorem for the wave equation (Tataru 1993).

Laplacian on Waves

If $b=0$, so that the system is governed only by boundary controls, its trajectory $\{u^{f,0}(\cdot, t) \mid 0 \leq t \leq T\}$ does not leave the reachable set \mathcal{U}_Γ^T . In this case, the system possesses one more intrinsic operator L^T which acts in the subspace $\text{cl } \mathcal{U}_\Gamma^T$ and is introduced through its graph

$$\text{gr } L^T := \text{cl} \left\{ \{W_{\text{bd}}^T f, -W_{\text{bd}}^T f_{tt}\} \mid f \in C_0^\infty(\Gamma \times (0, T)) \right\} \quad [8]$$

(closure in $\mathcal{H} \times \mathcal{H}$). By virtue of the relation $L^T W_{\text{bd}}^T f = -\Delta_g W_{\text{bd}}^T f$ following from the wave equation [1] and [6], the operator L^T is interpreted as Laplacian on waves filling the subdomain $\langle \Gamma \rangle^T$.

In the case $T > T^*$, one has $\langle \Gamma \rangle^T = \Omega$, $\text{cl } \mathcal{U}_\Gamma^T = \mathcal{H}$, and L^T is a densely defined operator in \mathcal{H} , satisfying $L^T \subset L$. Using [7], one proves the equality $L^T = L$. This equality and representation [4] imply that

$$W_{\text{vol}}^r b = \int_0^r (L^T)^{-1/2} \sin[(r-t)(L^T)^{1/2}] b(\cdot, t) dt \quad [9]$$

for all $r \geq 0$ and any fixed $T > T^*$.

Spectral Problem

The Dirichlet homogeneous boundary-value problem is to find nontrivial solutions of the system

$$-\Delta_g \varphi = \lambda \varphi \quad \text{in } \Omega \quad [10]$$

$$\varphi = 0 \quad \text{on } \Gamma \quad [11]$$

This problem is equivalent to the spectral analysis of the operator L ; it has the discrete spectrum $\{\lambda_k\}_{k=1}^\infty, 0 < \lambda_1 < \lambda_2 \leq \dots, \lambda_k \rightarrow \infty$; the eigenfunctions $\{\varphi_k\}_{k=1}^\infty, L\varphi_k = \lambda_k \varphi_k$, form an orthonormal basis in \mathcal{H} .

Expanding the solutions of the problem (1)–(3) over the eigenfunctions of the problem [10], [11] one derives the spectral representation of waves:

$$u^{f,0}(\cdot, T) = W_{\text{bd}}^T f = \sum_{k=1}^\infty (f, s_k^T)_{\mathcal{F}^T} \varphi_k(\cdot) \quad [12]$$

where

$$s_k^T(\gamma, t) := \lambda_k^{-1/2} \sin[(T-t)\lambda_k^{1/2}] \partial_\nu \varphi_k(\gamma)$$

Thus, for a given control f , the Fourier coefficients of the wave $u^{f,0}$ are determined by the spectrum $\{\lambda_k\}_{k=1}^\infty$ and the derivatives $\{\partial_\nu \varphi_k\}_{k=1}^\infty$.

Inverse problems

General Setup

The set of pairs $\Sigma := \{\lambda_k; \partial_\nu \varphi_k\}_{k=1}^\infty$ associated with the problem [10], [11] is said to be the Dirichlet spectral data of the manifold (Ω, d) . The spectral (frequency domain) inverse problem is to recover the manifold from its spectral data.

Since the speed of wave propagation is unity, the response operator R^T contains the information not about the entire manifold but only about its part $(\Gamma)^{T/2}$. This fact is taken into account in the dynamical (time domain) inverse problem which aims to recover the manifold from the operator R^{2T} given for a fixed $T > T_*$.

If the manifolds (Ω', d') and (Ω'', d'') are isometric via an isometry $i: \Omega' \rightarrow \Omega''$, then, identifying the boundaries by $i(\gamma) \equiv \gamma$, one gets two manifolds with the common boundary $\Gamma = \partial\Omega' = \partial\Omega''$ which possess identical inverse data: $\Sigma' = \Sigma'', R'^{2T} = R''^{2T}$. Such manifolds are called equivalent: they are indistinguishable for the external observer extracting Σ or R^{2T} from the boundary measurements. Therefore, these data do not determine the manifold uniquely and both of the inverse problems need to be clarified. The precise formulation is given in the form of two questions:

1. Does the coincidence of the inverse data imply the equivalence of the manifolds?
2. Given the inverse data of an unknown manifold, how to construct a manifold possessing these data?

The BC method gives an affirmative answer to the first question and provides a procedure producing a representative of the class of equivalent manifolds from its inverse data. The method is based on the concepts of model and “coordinatization.”

Model

A pair consisting of an auxiliary Hilbert space $\tilde{\mathcal{H}}$ and an operator $\tilde{W}_{\text{bd}}^T: \mathcal{F}^T \rightarrow \tilde{\mathcal{H}}$ is said to be a model of the system α^T , if \tilde{W}_{bd}^T is determined by inverse data, and the map $U: \tilde{W}_{\text{bd}}^T f \mapsto W_{\text{bd}}^T f$ is an isometry from $\text{Ran } \tilde{W}_{\text{bd}}^T \subset \tilde{\mathcal{H}}$ onto $\text{Ran } W_{\text{bd}}^T \subset \mathcal{H}$. The model is an intermediate object in solving inverse problems. It plays the role of an auxiliary copy of the original dynamical system which an external observer can build from measurements on the boundary. While the genuine wave process inside Ω , initiated by a boundary control, remains inaccessible for direct measurements, its $\tilde{\mathcal{H}}$ -representation can be visualized by means of the model control operator \tilde{W}_{bd}^T . This is illustrated by the diagram on Figure 2, where the upper part is invisible for an external observer, whereas the lower part can be extracted from inverse data.

Each type of data determines a corresponding model. The spectral model is the pair

$$\tilde{\mathcal{H}} := l_2, \quad \tilde{W}_{\text{bd}}^T := \{(\cdot, s_k^T)_{\mathcal{F}^T}\}_{k=1}^\infty \quad [13]$$

(see [12]); the role of isometry U is played by the Fourier transform $F: \mathcal{H} \rightarrow \tilde{\mathcal{H}}, Fy := \{(y, \varphi)_{\mathcal{H}}\}_{k=1}^\infty$. By virtue of [4], the data Σ also determine the operator $\tilde{W}_{\text{vol}}^r: L_2([0, r]; \tilde{\mathcal{H}}) \rightarrow \tilde{\mathcal{H}}$,

$$\tilde{W}_{\text{vol}}^r = \int_0^r \tilde{L}^{-1/2} \sin[(r-t)(\tilde{L})^{1/2}] (\cdot)(t) dt, \quad r \geq 0 \quad [14]$$

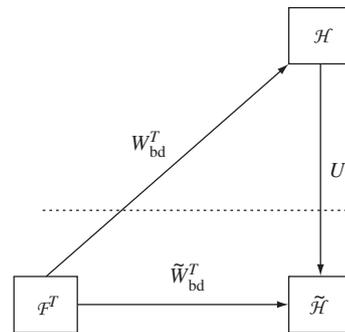


Figure 2 Model of a system. (Data from Belishev (1997).)

where $\tilde{L} := ULU^* = \text{diag}\{\lambda_k\}_{k=1}^\infty$. Thus, the spectral model allows one to see the Fourier images of invisible waves.

According to [5], the response operator R^{2T} determines the modulus of the control operator

$$|W_{\text{bd}}^T| = [(W_{\text{bd}}^T)^* W_{\text{bd}}^T]^{1/2} = (C^T)^{1/2}$$

which enters in the polar decomposition $W_{\text{bd}}^T = \Phi |W_{\text{bd}}^T|$. Along with it, the response operator determines the dynamical model

$$\tilde{\mathcal{H}} := \text{cl Ran}(C^T)^{1/2}, \quad \tilde{W}_{\text{bd}}^T := (C^T)^{1/2} \quad [15]$$

The correspondence “system \rightarrow model” is realized by the isometry $U = \Phi^* : W_{\text{bd}}^T f \mapsto |W_{\text{bd}}^T| f$. The operator $\tilde{L}^T := UL^T U^*$ dual to the Laplacian on waves, is determined by its graph

$$\begin{aligned} & \text{gr } \tilde{L}^T \\ & := \text{cl} \left\{ \{ \tilde{W}_{\text{bd}}^T f, -\tilde{W}_{\text{bd}}^T f_{tt} \} \mid f \in C_0^\infty(\Gamma \times (0, T)) \right\} \quad [16] \end{aligned}$$

(see [8]) and, therefore, \tilde{L}^T is also determined by R^{2T} . In the case $T > T_*$, the operator $\tilde{W}_{\text{vol}}^r : L_2([0, r]; \tilde{\mathcal{H}}) \rightarrow \tilde{\mathcal{H}}$ dual to W_{vol}^r , is represented in the form

$$\begin{aligned} \tilde{W}_{\text{vol}}^r &= \int_0^r (\tilde{L}^T)^{-1/2} \sin[(r-t)(\tilde{L}^T)^{1/2}] (\cdot)(t) dt, \\ r &\geq 0 \end{aligned} \quad [17]$$

in accordance with [9]. Thus, the dynamical model visualizes the Φ^* -images of the waves propagating inside Ω .

Wave Coordinatization

In a general sense, a coordinatization is a correspondence between points x of the studied set \mathcal{A} and elements \tilde{x} of another set $\tilde{\mathcal{A}}$ such that: (i) the elements of $\tilde{\mathcal{A}}$ are accessible and distinguishable; (ii) the map $x \mapsto \tilde{x}$ is a bijection; and (iii) relations between elements of \mathcal{A} determine those between points of $\tilde{\mathcal{A}}$ which are studied (H Weyl). Coordinatization enables one to study \mathcal{A} via operations with coordinates $\tilde{x} \in \tilde{\mathcal{A}}$.

The external observer investigating the manifold probes Ω with waves initiated by sources on Γ . The relevant coordinatization of Ω described below uses such waves and is implemented in three steps.

Step 1 (subdomains) Let $x(\gamma, \tau)$ be the end point of the geodesic of the length $\tau > 0$ emanating from $\gamma \in \Gamma$ in the direction $-\nu(\gamma)$, and let $\sigma_\gamma^\varepsilon \subset \Gamma$ be a small neighborhood shrinking to γ as $\varepsilon \rightarrow 0$. If $\tau \leq \tau_*(\gamma)$, then the family of subdomains

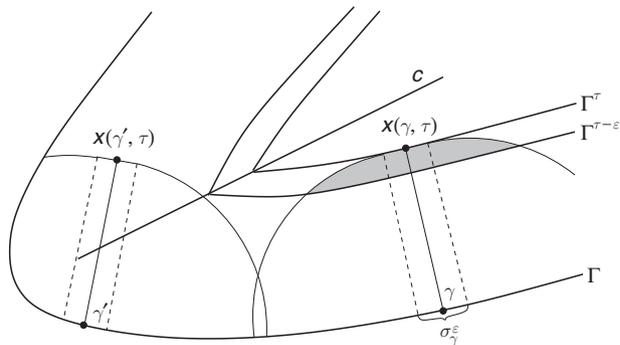


Figure 3 The subdomains.

$$\omega^\varepsilon(\gamma, \tau) := [\langle \Gamma \rangle^\tau \setminus \langle \Gamma \rangle^{\tau-\varepsilon}] \cap \langle \sigma_\gamma^\varepsilon \rangle^\tau$$

(shaded domain on Figure 3) shrinks to $x(\gamma, \tau)$; if $\tau > \tau_*(\gamma)$, then the family terminates: $\omega^\varepsilon(\gamma, \tau) = \emptyset$ as $\varepsilon < \varepsilon_0(\gamma)$ (the case $\gamma = \gamma'$ in Figure 3). Such behavior of subdomains implies that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \langle [\langle \Gamma \rangle^\tau \setminus \langle \Gamma \rangle^{\tau-\varepsilon}] \cap \langle \sigma_\gamma^\varepsilon \rangle^\tau \rangle^r \\ &= \begin{cases} \langle x(\gamma, \tau) \rangle^r, & \tau \leq \tau_*(\gamma) \\ \emptyset, & \tau > \tau_*(\gamma) \end{cases} \quad [18] \end{aligned}$$

Step 2 (wave subspaces) Pass from the subdomains to the corresponding subspaces $\mathcal{H}(\langle \Gamma \rangle^\tau)$, $\mathcal{H}(\langle \sigma_\gamma^\varepsilon \rangle^\tau)$, $\mathcal{H}(\langle \omega^\varepsilon(\gamma, \tau) \rangle^r)$, and represent them via reachable sets by [7]:

$$\begin{aligned} \mathcal{H}(\langle \Gamma \rangle^\tau) &= \text{cl } W_{\text{bd}}^r \mathcal{F}^\tau, & \mathcal{H}(\langle \sigma_\gamma^\varepsilon \rangle^\tau) &= \text{cl } W_{\text{bd}}^r \mathcal{F}_{\sigma_\gamma^\varepsilon}^\tau \\ \mathcal{H}(\langle \omega^\varepsilon(\gamma, \tau) \rangle^r) &= \text{cl } W_{\text{vol}}^r L_2([0, r]; \mathcal{H}(\langle \Gamma \rangle^\tau) \\ & \ominus \mathcal{H}(\langle \Gamma \rangle^{\tau-\varepsilon}) \cap \mathcal{H}(\langle \sigma_\gamma^\varepsilon \rangle^\tau)) \\ &= \text{cl } W_{\text{vol}}^r L_2([0, r]; [\text{cl } W_{\text{bd}}^r \mathcal{F}^\tau \\ & \ominus \text{cl } W_{\text{bd}}^{r-\varepsilon} \mathcal{F}^{r-\varepsilon}] \cap \text{cl } W_{\text{bd}}^r \mathcal{F}_{\sigma_\gamma^\varepsilon}^\tau) \end{aligned}$$

Define

$$\begin{aligned} \mathcal{W}_{(\gamma, \tau)}^r &:= \lim_{\varepsilon \rightarrow 0} \text{cl } W_{\text{vol}}^r L_2([0, r]; [\text{cl } W_{\text{bd}}^r \mathcal{F}^\tau \\ & \ominus \text{cl } W_{\text{bd}}^{r-\varepsilon} \mathcal{F}^{r-\varepsilon}] \cap \text{cl } W_{\text{bd}}^r \mathcal{F}_{\sigma_\gamma^\varepsilon}^\tau) \quad [19] \end{aligned}$$

$\mathcal{W}_{(\gamma, 0)}^r := \mathcal{W}_{(\gamma, +0)}^r$, $r \geq 0$ (the limits in the sense of the strong operator convergence of the projections in \mathcal{H} on the corresponding subspaces). By the definitions, one has $\mathcal{W}_{(\gamma, \tau)}^r = \lim_{\varepsilon \rightarrow 0} \mathcal{H}(\langle \omega^\varepsilon(\gamma, \tau) \rangle^r)$, whereas [18] leads to the equality

$$\mathcal{W}_{(\gamma, \tau)}^r = \begin{cases} \mathcal{H}(\langle x(\gamma, \tau) \rangle^r), & \tau \leq \tau_*(\gamma) \\ \{0\}, & \tau > \tau_*(\gamma) \end{cases} \quad [20]$$

for all $\gamma \in \Gamma, \tau \geq 0, r \geq 0$. As a result, since any $x \in \Omega$ can be represented as $x = x(\gamma, \tau)$, one attaches to every point of the manifold a family of expanding subspaces $\{\mathcal{W}_{(\gamma,\tau)}^r | r \geq 0\}$ built out of waves. As is seen from [20], the family is determined by the point x (not dependent on the representation $x = x(\gamma, \tau)$); the subspaces which it consists of coincide with $\mathcal{H}\langle x \rangle^r$.

Expressing the distance as

$$d(x', x'') = 2 \inf \{r > 0 | \mathcal{H}\langle x' \rangle^r \cap \mathcal{H}\langle x'' \rangle^r \neq \{0\}\}$$

in accordance with [20], one can represent

$$d(x', x'') = 2 \inf \{r > 0 | \mathcal{W}_{(\gamma',\tau')}^r \cap \mathcal{W}_{(\gamma'',\tau'')}^r \neq \{0\}\} \quad [21]$$

where $x' = x(\gamma', \tau'), x'' = x(\gamma'', \tau'')$, and hence find the distance via the above families.

Step 3 (wave copy) By varying $\gamma \in \Gamma, \tau \geq 0$, gather all nonzero families $\{\mathcal{W}_{(\gamma,\tau)}^r | r \geq 0\} =: \tilde{x}$ in the set $\tilde{\Omega} = \{\tilde{x}\}$. Redenoting $\mathcal{W}_{\tilde{x}}^r := \mathcal{W}_{(\gamma,\tau)}^r \in \tilde{x}$, endow the set with the distance

$$\tilde{d}(\tilde{x}', \tilde{x}'') := 2 \inf \{r > 0 | \mathcal{W}_{\tilde{x}'}^r \cap \mathcal{W}_{\tilde{x}''}^r \neq \{0\}\} \quad [22]$$

In view of [21], one has $d(x', x'') = \tilde{d}(\tilde{x}', \tilde{x}'')$, so that the metric space $(\tilde{\Omega}, \tilde{d})$ is an isometric copy of (Ω, d) by construction. Thus, the correspondence $x \mapsto \tilde{x}$ (“point \mapsto family”) is an isometry and satisfies the general principles (i)–(iii) of coordinatization.

The manifold $(\tilde{\Omega}, \tilde{d})$ is the end product of the wave coordinatization. It represents the original manifold as a collection of infinitesimal sources interacting with each other via the waves which they produce.

Solving Inverse Problems

The motivation for the above coordinatization is that the wave copy can be reproduced via any model. Namely, the external observer with the knowledge of Σ or $R^{2T}(T > T_*)$ can recover $(\tilde{\Omega}, \tilde{d})$ up to isometry by the following procedure:

1. Construct the model corresponding to the given inverse data and determine the operators $\tilde{W}_{\text{bd}}^T, 0 \leq T \leq T$ by [13], [15]; then determine \tilde{L}, \tilde{L}^T , and \tilde{W}_{vol}^T by [14] or [16], [17].
2. Replace on the right-hand side of [19] all operators W without tildes by the ones with tildes, and get the subspaces $\tilde{\mathcal{W}}_{(\gamma,\tau)}^r = U\mathcal{W}_{(\gamma,\tau)}^r, \gamma \in \Gamma, \tau \geq 0, r \geq 0$.
3. Gather all nonzero families $\{\tilde{\mathcal{W}}_{(\gamma,\tau)}^r | r \geq 0\} =: \hat{x}$ in the set $\hat{\Omega} = \{\hat{x}\}$ and redenote the subspaces as $\tilde{\mathcal{W}}_{\hat{x}}^r := \tilde{\mathcal{W}}_{(\gamma,\tau)}^r \in \hat{x}$; endow the set with the metric $\hat{d}(\hat{x}', \hat{x}'') := 2 \inf \{r > 0 | \tilde{\mathcal{W}}_{\hat{x}'}^r \cap \tilde{\mathcal{W}}_{\hat{x}''}^r \neq \{0\}\}$ (see [22]), and get a sample $(\hat{\Omega}, \hat{d})$ of the wave copy $(\tilde{\Omega}, \tilde{d})$.

This sample is isometric to the original (Ω, d) by construction. Identifying properly the boundaries $\partial\hat{\Omega}$ and Γ , one turns $(\hat{\Omega}, \hat{d})$ into a canonical representative of the class of equivalent manifolds possessing the given inverse data.

If the response operator R^{2T} is given for a fixed $T < T_*$, the above procedure produces the wave copy of the submanifold $(\langle \Gamma \rangle^T, d)$. This locality in time is an intrinsic feature and advantage of the BC method: longer time of observation on Γ increases the depth of penetration into Ω .

Amplitude Formula

Another variant of the BC method is based on geometrical optics formulas describing the propagation of singularities of the waves.

Let $y \in \mathcal{H}$, and let β be the density of the volume in semigeodesic coordinates: $dx = \beta d\Gamma d\tau$; the function

$$\tilde{y}(\gamma, \tau) := \begin{cases} \beta^{1/2}(\gamma, \tau) y(x(\gamma, \tau)), & (\gamma, \tau) \in \Theta \\ 0, & \text{otherwise} \end{cases}$$

defined on $\Gamma \times [0, T_*]$ is called the image of y . The amplitude formula represents the images of waves initiated by boundary controls in the form

$$u^{f,0}(\cdot, T)(\gamma, \tau) = \lim_{t \rightarrow T - \tau - 0} [(W_{\text{bd}}^T)^*(I - P^\tau)W_{\text{bd}}^T f](\gamma, t) \quad 0 < \tau < T$$

where I is the identity operator and P^τ is the projection in \mathcal{H} onto $\text{cl } W_{\text{bd}}^\tau \mathcal{F}^\tau$. The formula is derived by the ray method going back to J Hadamard, the derivation uses the controllability [7].

Any model determines the right-hand side of the last relation by the isometry: $(W_{\text{bd}}^T)^*(I - P^\tau)W_{\text{bd}}^T = (\tilde{W}_{\text{bd}}^T)^*(\tilde{I} - \tilde{P}^\tau)\tilde{W}_{\text{bd}}^T$, where $\tilde{W}_{\text{bd}}^T = U W_{\text{bd}}^T, \tilde{I}$ is the identity operator, and $\tilde{P}^\tau = U P^\tau U^*$ is the projection in $\tilde{\mathcal{H}}$ onto $\text{cl } \tilde{W}_{\text{bd}}^\tau \mathcal{F}^\tau$. This leads to the representation

$$u^{f,0}(\cdot, T)(\gamma, \tau) = \lim_{t \rightarrow T - \tau - 0} [(\tilde{W}_{\text{bd}}^T)^*(\tilde{I} - \tilde{P}^\tau)\tilde{W}_{\text{bd}}^T f](\gamma, t) \quad 0 < \tau < T \quad [23]$$

and makes the amplitude formula a useful tool for solving the inverse problems. The external observer can construct a model via inverse data and then visualize by [23] the wave images on the part Θ^T of the pattern (see Figure 1). The collection of images $u^{f,0}$ corresponding to all possible controls f is rich enough for recovering the tensor g on Θ^T (i.e., the metric tensor in semigeodesic coordinates) and turning the pattern into an isometric copy of the submanifold $(\langle \Gamma \rangle^T, d)$. This variant of the method is

more appropriate if one needs to recover unknown coefficients of the wave equation in Ω – it can be realized in terms of numerical algorithms.

Extensions of the Method

Electromagnetic waves are also well suited for coordinatization and for constructing the wave copy $(\tilde{\Omega}, \tilde{d})$. An appropriate version of the amplitude formula also exists for the system governed by the Maxwell equations (see [Further Reading](#)). At present (2004), the applicability of the BC method to three-dimensional inverse problems of elasticity theory is still an open question. The following hypothesis concerns the Lamé system: the wave coordinatization procedure (steps 1–3) using the elastic waves instead of the above $u^{f,0}$, gives rise to the copy of $\Omega \subset \mathbf{R}^3$ endowed with the metric $|dx|^2/c_p^2$ where $c_p = \sqrt{(\lambda + 2\mu)/\rho}$ is the speed of the pressure waves.

The concept of model is used for solving inverse problems for the heat and Schrödinger equations (Avdonin and Belishev, 1995–2004), as well as for the problem of boundary data continuation (Belishev 2001, Kurylev and Lassas 2002). A variant of the BC method allows one to recover not only the manifold but also the Schrödinger type operators on it and/or the dissipative term in the scalar wave equation (Kurylev and Lassas 1993–2003).

An appropriate version of the amplitude formula solves the inverse problem for one-dimensional two-velocity dynamical system which describes the waves consisting of two modes propagating with different speeds and interacting with each other (Belishev, Blagoveschenskii, Ivanov, 1997–2000).

One more variant of coordinatization going back to the first paper on the BC method, associates with points $x \in \Omega$ the Dirac measures δ_x ; then, their images $\tilde{\delta}_x$ are identified via suitable models. This variant solves inverse problems on graphs and the two-dimensional elliptic Calderon problem. The reader is referred to articles by the present author listed in [Further Reading](#).

Within the scope of the method, one derives some natural analogs of the classical Gelfand–Levitan–Krein–Marchenko equations (Belishev, 1987–2001). Also, an appropriate analog solves the kinematic inverse problem for a class of two-dimensional manifolds (Pestov 2004).

There exists an abstract version of the approach, embedding the BC method into the

framework of linear system theory (Belishev 2001). The method is also related to the problem of triangular factorization of operators (Belishev and Pushnitski 1996).

Numerical algorithms for solving two-dimensional spectral and dynamical inverse problems for the wave equation $\rho u_{tt} - \Delta u = 0$ which recover the variable density ρ have been developed and tested (Filippov, Gotlib, Ivanov, 1994–1999).

See also: Dynamical Systems and Thermodynamics; Geophysical Dynamics; Inverse Problem in Classical Mechanics.

Further Reading

- Belishev MI (1988) On an approach to multidimensional inverse problems for the wave equation. *Soviet Mathematics. Doklady* 36(3): 481–484.
- Belishev MI (1996) Canonical model of a dynamical system with boundary control in the inverse problem of heat conductivity. *St. Petersburg Mathematical Journal* 7(6): 869–890.
- Belishev MI (1997) Boundary control in reconstruction of manifolds and metrics. *Inverse Problems* 13(5): R1–R45.
- Belishev MI (2001) Dynamical systems with boundary control: models and characterization of inverse data. *Inverse Problems* 17: 659–682.
- Belishev MI (2002) How to see waves under the Earth surface (the BC-method for geophysicists). In: Kabanikhin SI and Romanov VG (eds.) *Ill-Posed and Inverse Problems*, pp. 67–84. Utrecht/Boston: VSP.
- Belishev MI (2003) The Calderon problem for two-dimensional manifolds by the BC-method. *SIAM Journal of Mathematical Analysis* 35(1): 172–182.
- Belishev MI (2004) Boundary spectral inverse problem on a class of graphs (trees) by the BC-method. *Inverse Problems* 20(3): 647–672.
- Belishev MI and Glasman AK (2001) Dynamical inverse problem for the Maxwell system: recovering the velocity in the regular zone (the BC-method). *St. Petersburg Mathematical Journal* 12(2): 279–319.
- Belishev MI and Gotlib VYu (1999) Dynamical variant of the BC-method: theory and numerical testing. *Journal of Inverse and Ill-Posed Problems* 7(3): 221–240.
- Belishev MI, Isakov VM, Pestov LN, and Sharafutdinov VA (2000) On reconstruction of metrics from external electromagnetic measurements. *Russian Academy of Sciences. Doklady. Mathematics* 61(3): 353–356.
- Belishev MI and Ivanov SA (2002) Characterization of data of dynamical inverse problem for two-velocity system. *Journal of Mathematical Sciences* 109(5): 1814–1834.
- Belishev MI and Lasiecka I (2002) The dynamical Lamé system: regularity of solutions, boundary controllability and boundary data continuation. *ESAIM COCV* 8: 143–167.
- Katchalov A, Kurylev Y, and Lassas M (2001) Inverse Boundary Spectral Problems. *Chapman and Hall/CRC Monographs and Surveys in Pure and Applied Mathematics*, vol. 123. Boca Raton, FL: Chapman and Hall/CRC.

Boundary-Value Problems for Integrable Equations

B Pelloni, University of Reading, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Integrable equations are a special class of nonlinear equations arising in the modeling of a wide variety of physical phenomena. It has been argued that integrable PDEs are in a certain, specific sense “universal” models for physical phenomena involving weak nonlinearity. Indeed, integrable equations are obtained by a procedure involving rescaling and an asymptotic expansion from very large classes of nonlinear evolution equations, which preserves integrability while retaining in the limit weakly nonlinear effects. For this reason, integrable equations are a very important class of PDEs. Important examples are the nonlinear Schrödinger (NLS) equation

$$iq_t + q_{xx} - 2\lambda|q|^2q = 0, \quad \lambda = \pm 1 \quad [1]$$

the Korteweg–deVries (KdV) equation

$$q_t + q_x \pm q_{xxx} + 6qq_x = 0 \quad [2]$$

the modified KdV (mKdV) equation

$$q_t \pm q_{xxx} \mp 6\lambda q^2 q_x = 0, \quad \lambda = \pm 1 \quad [3]$$

and the sine-Gordon (SG) equation in light-cone or laboratory coordinates

$$q_{xt} + \sin q = 0 \quad \text{or} \quad q_{tt} - q_{xx} + \sin q = 0 \quad [4]$$

A general method for solving the initial-value problem for integrable equations in one space dimension was discovered in 1967, when in a pioneering and much celebrated work ([Gardner *et al.* 1967](#)), the initial-value problems for KdV with decaying initial condition was completely solved. Soon afterwards, it was understood that this method, now known as the “inverse scattering transform,” is of more general applicability. Indeed, it can be applied to those nonlinear equations that can be written as the compatibility condition of a pair of linear eigenvalue equations. The method of solution for the Cauchy problem essentially relies on the possibility of expressing the equation through this pair, now called a Lax pair after the work of [Lax \(1968\)](#), who first clarified the connection. [Zakharov and Shabat \(1972\)](#) constructed such a pair for the NLS equation, and in subsequent years the Lax pairs associated with all important integrable equations in one and two spatial variables were constructed. These include the NLS, sG, mKdV,

Davey–Stewartson I and II, and Kamdotsev–Petviashvili I and II equations.

There is no universally accepted definition of an integrable PDE, but on account of the above results, the existence of a Lax pair can be taken as the defining property of such equations. In the course of the 1970s, the inverse scattering transform was applied to solve the initial-value (Cauchy) problem for many integrable equations. In principle, there is no obstruction to solving analytically the initial-value problem by the inverse scattering transform as soon as a Lax pair is constructed for the equation, and appropriate decaying initial conditions are prescribed. The solution is then characterized in terms of a certain integral equation. This approach is equivalent to associating with the initial-value problem a classical problem in complex analysis, namely a matrix Riemann–Hilbert problem, defined in the complex spectral space. This point of view is currently taken by many authors as it provides a unifying and very flexible framework for the analysis.

After the success of the inverse scattering transform in solving the Cauchy problem, it was natural to attempt to generalize the approach to boundary-value problems. To describe the difficulties involved in this generalization, consider the case of evolution equations in one space and one time dimensions. The independent variables can be denoted by (x, t) , with $t > 0$ representing time. While the initial-value problem is posed on the full real line, hence for $x \in (-\infty, \infty)$, the simplest boundary-value problem is posed on a half-line, for $x \in (0, \infty)$. In addition to initial conditions for initial time $t=0$, it is necessary to prescribe conditions at the boundary $x=0$. The number of conditions that must be prescribed to obtain a problem which admits a unique solution depends on the particular equation, but for evolution equation it is roughly equal to half the number of x -derivatives involved in the equation. For example, for the NLS equation, a well-posed problem is defined as soon as one boundary condition at $x=0$ is prescribed; hence a typical boundary-value problem for this equation is obtained, for example, when $q(x, 0) = q_0(x)$ and $q(0, t) = g_0(t)$ are prescribed and compatible, so that $q_0(0) = g_0(0)$. It follows that, while $q_{xx}(0, t)$ can be computed from the equation, $q_x(0, t)$ is not immediately known. An even more difficult situation arises for the KdV equation [2] (with the + sign), for which a well-posed problem is again defined as soon as one boundary condition is prescribed, so that there are two unknown boundary values.

Because of this simple fact, a straightforward application of the ideas of the inverse scattering transform immediately encounters one crucial difficulty. This transform method yields an integral representation of the solution which involves not only the given boundary conditions $f(t)$, but also the other “unknown” boundary values – in our example for the NLS equation, the function $q_x(0, t)$. The problem of characterizing these unknown boundary values has impeded progress in this direction for over thirty years.

On account of their physical significance, various boundary-value problems for the KdV equation have been considered, and classical PDE techniques (not specific to integrable models) have been used to establish existence and uniqueness results (Bona *et al.* 2001, Colin and Ghidaglia 2001, Colliander and Kenig 2001). These approaches, and in particular the approach of Colliander and Kenig, are quite general and possibly of wide applicability, and give global existence results in wide functional classes. However, they do not rely on integrability properties. Indeed, none of these results use the integrable structure of the equation in any fundamental or systematic way. However, the fact that these equations are integrable on the full line implies very special properties that should be exploited in the analysis and it is natural to try to generalize the inverse scattering transform approach.

Such a generalization is sometimes directly possible. For example, it has been used for studying the problem on the half-line for the hyperbolic version of the sG equation [4a] which does not involve unknown boundary values (Fokas 2000, Pelloni). It has also been used to study some specific boundary-value problems for the NLS equation, for example, for homogeneous Dirichlet or Neumann conditions, when it is possible to use even or odd extensions of the problem to the full line (Ablowitz and Segur 1974), or more recently in Degasperis *et al.* (2001). In the latter case, however, the unknown boundary values are characterized through an integral Fredholm equation, which does not admit a unique solution. Some special cases of boundary-value problems for the KdV equation (Adler *et al.* 1997, Habibullin 1999) and elliptic sG (Sklyanin 1987) have also been studied via the inverse scattering transform. However all the examples considered are nongeneric, and it has recently been shown (Fokas, *in press*) that the boundary conditions chosen fall in the special class of the so-called “linearizable” boundary conditions, for which the problem can be solved as if it were posed on the full line. One cannot hope to use similar methods to solve the problem with generic boundary conditions.

Recently, Fokas (2000) introduced a general methodology to extend the ideas of the inverse scattering transform to boundary-value problems. This methodology provides the tools to analyze boundary-value problems for integrable equations to a considerable degree of generality. We note as a side remark that linear PDEs are trivially integrable, in the sense of admitting a Lax pair (in this case the Lax pair can be found algorithmically, while the construction of the Lax pair associated with a nonlinear equation is by no means trivial). As a consequence of this remark, the extension of the inverse scattering transform also provides a method for solving boundary-value problems for a large variety of linear PDEs of mathematical physics.

What follows is a general description of the approach of Fokas, considering, for the sake of concreteness, the case of an integrable PDE in the two variables (x, t) which vary in the domain D (typically, for an evolution problem $D = (0, \infty) \times (0, T)$). We assume that $q(x, t)$ denotes the unique solution of a boundary-value problem posed for such an equation.

The method consists of the following steps.

1. Write the PDE as the compatibility condition of a Lax pair. This is a pair of linear ODEs for the function $\mu = \mu(x, t, k)$ involving the solution $q(x, t)$ of the PDE, the derivatives of this solution, and a complex parameter k , called the spectral parameter. This can be done algorithmically for linear PDEs, and in this case $\mu(x, t, k)$ is a scalar function. For nonlinear integrable PDEs, $\mu(x, t, k)$ is in general a matrix-valued function.
2. (a) The equivalence of the PDE with a Lax pair can be reformulated in the language of differential forms, and in this language it is easier to describe the methodology in general. Assume then that $\Omega(x, t, k)$ is a differential 1-form expressed in terms of a function $q(x, t)$ and its derivatives, and of a complex variable k , and one which is characterized by the property that $d\Omega = 0$ if and only if $q(x, t)$ satisfies the given PDE. The closure of the form Ω yields the two important consequences 2(a) and 2(b) below.
 - (a) Since the domain D under consideration is simply connected, the closed form Ω is also exact; hence, it is possible to find the particular, 0-form $\mu(x, t, k)$, solving $d\mu = \Omega$. In particular, $\mu(x, t, k)$ can be chosen to be sectionally bounded with respect to k by solving either a Riemann–Hilbert problem or a d -bar problem in the complex spectral k plane, and the solution $\mu(x, t, k)$ is then expressed in terms of certain “spectral functions” depending on all the boundary values

of the solution $q(x, t)$ of the PDE. The function $q(x, t)$ can then be expressed in terms of $\mu(x, t, k)$. (b) The integral of Ω along the boundary of the domain D vanishes. This yields an integral constraint between all boundary values of the solution of the PDE, which becomes an algebraic constraint for the spectral functions. The resulting algebraic identity is called the “global relation.”

3. The last step is the analysis the k -invariance properties of the global relation. This analysis yields the characterization of the spectral functions in terms only of the given boundary conditions.

The crucial and most difficult step in the solution process is the characterization described above. The analysis required depends on the type of problem under consideration. For nonlinear integrable evolution PDEs posed on the half-line $x > 0$, in general the characterization mentioned in step (3) involves solving a system of nonlinear Volterra integral equations. This is an important difference from the case of the Cauchy problem, where the solution is given by a single integral equation where all the terms are explicitly known.

The method outlined above has been applied successfully to solve a variety of boundary-value problems for linear and integrable nonlinear PDEs. For concreteness, here the focus is on the important case of integrable evolution PDEs in one space, which illustrates clearly the generalities of this method.

Integrable Evolution Equations in One Space Dimension

The crucial property of integrable PDEs which is used in the inverse scattering transform approach to solve the initial-value problem is the fact that they can be written as the compatibility of a Lax pair. Many integrable evolution equations of physical significance (such as NLS, KdV, sG, and mKdV) admit a Lax pair of the form

$$\begin{aligned} \mu_x + if_1(k)\sigma_3\mu &= Q(x, t, k)\mu \\ \mu_t + if_2(k)\sigma_3\mu &= \tilde{Q}(x, t, k)\mu \end{aligned} \tag{5}$$

where $\mu(x, t, k)$ is a 2×2 matrix, $\sigma_3 = \text{diag}(1, -1)$, $f_i(k), i = 1, 2$, are analytic functions of the complex parameter k , and Q, \tilde{Q} are analytic functions of k , of the function $q(x, t)$ (and of its complex conjugate $\overline{q(x, t)}$ for complex-valued problems) and of its derivatives. For example, the NLS equation [1] is equivalent to the compatibility condition of the pair

$$\begin{aligned} \mu_x + ik\sigma_3\mu &= Q\mu, \quad Q = \begin{pmatrix} 0 & q \\ \lambda\bar{q} & 0 \end{pmatrix} \\ \mu_t + 2ik^2\sigma_3\mu &= (2kQ - iQ_x\sigma_3 - i\lambda|q|^2\sigma_3)\mu \end{aligned} \tag{6}$$

The first step towards a systematic new approach to solving boundary-value problem was the work of Fokas and Its, who associated the boundary-value problem for NLS on the half-line to a single Riemann–Hilbert problem determined by both equations in the Lax pair. The jump determining this Riemann–Hilbert problem has an explicit exponential dependence on both x and t . This differs from the classical inverse scattering approach, in which the x -part of the Lax pair is used to determine an x -transform with t -dependent scattering data, and the t -part of the Lax pair is then exploited to find the time evolution of these data. The work of Fokas and Its led to the understanding that both equations in the Lax pair [6] must be considered in order to construct a spectral transform appropriate to solve boundary-value problems. Fokas (2000) reviews his systematic way to solve these problems by performing the simultaneous spectral analysis of both equations in the Lax pair. The transform thus obtained, which is a nonlinearization of the Fourier transform, precisely generalizes the inverse scattering transform.

This simultaneous analysis also leads naturally to the identification of the “global relation” which holds between initial and boundary data, and which plays an essential role in deriving an expression for the solution of the problem which does not involve unknown boundary values.

The Riemann–Hilbert problem with explicit (x, t) dependence, the global relation, and the invariance properties of the latter with respect to the spectral parameter are the fundamental ingredients of this systematic approach to solve boundary-value problems for integrable equations.

The steps involved in this method are summarized in the introduction. While steps (1) and (2) can be described generally, and, once the Lax pair is identified, can be performed algorithmically (at least under the assumption that the solution of the PDE exists), the last step is the most difficult part of the analysis, and it needs to be considered separately for each given problem. However, it is this step that yields the effective characterization of the solution.

The results obtained for the particular case of eqn [1] are reviewed in detail in the next section, as they provide an important example, which can be generalized without any conceptual difficulty to eqns [2]–[4].

The NLS Equation

As already mentioned, the initial-value problem for NLS was solved, for decaying initial condition, by Zakharov and Shabat, and studied in depth by many others. However, by the mid-1990s only a handful of papers had been written on the solution of the boundary-value problem posed on the half-line, all on a specific example or aspect of the problem, or attempts at solving the problem using general PDE techniques.

For this equation, the approach of Fokas yields the following results. Let the complex-valued function $q(x, t)$ satisfy the NLS equation [1], for $x > 0$ and $t > 0$, for prescribed one initial and one boundary conditions. For the sake of concreteness, we select the specific initial and boundary conditions

$$\begin{aligned} q(x, 0) &= q_0(x) \in \mathcal{S}(\mathbb{R}^+) \\ q(0, t) &= g_0(t) \in \mathcal{S}(\mathbb{R}^+) \\ q_0(0) &= g_0(0) \end{aligned} \tag{7}$$

where \mathcal{S} denotes the space of Schwartz functions (similar results hold for different choices of boundary conditions, and less restrictive function classes).

The solution of this initial boundary-value (IBV) problem can be constructed as follows (Fokas 2000, 2002; in press):

- Given $q_0(x)$ construct the spectral functions $\{a(k), b(k)\}$. These functions are defined by

$$a(k) = \phi_2(0, k), \quad b(k) = \phi_1(0, k)$$

where the vector $\phi(x, k)$ with components $\phi_1(x, k)$ and $\phi_2(x, k)$ is the following solution of the x -problem of the associated Lax pair evaluated at $t = 0$:

$$\begin{aligned} \phi_x + ik\sigma_3\phi &= Q(x, 0, k)\phi, \quad 0 < x < \infty, \text{Im } k \geq 0 \\ \phi(x, k) &= e^{ikx} \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} + o(1) \right) \text{ as } x \rightarrow \infty \\ Q(x, 0, k) &= \begin{pmatrix} 0 & q_0(x) \\ \lambda \bar{q}_0(x) & 0 \end{pmatrix} \end{aligned}$$

(σ_3 and $Q(x, t, k)$ are defined after eqns [5] and [6], respectively).

- Given $q_0(x)$ and $g_0(t)$ characterize $g_1(t)$ by the requirement that the spectral functions $\{A(t, k), B(t, k)\}$ satisfy the global relation

$$\begin{aligned} B(t, k) - R(k)A(t, k) &= e^{4ik^2t} \frac{c(t, k)}{a(k)} \\ R(k) &= \frac{b(k)}{a(k)}, \quad t \in [0, T], k \in \bar{D} \end{aligned} \tag{8}$$

where D denotes the first quadrant of the complex k -plane:

$$D = \{k | \text{Re } k > 0, \text{Im } k > 0\}$$

\bar{D} denotes the closure of D , and $c(t, k)$ is a function of k analytic in D and of order $O(1/k)$ as $k \rightarrow \infty$. The spectral functions are defined by

$$\begin{aligned} A(t, k) &= e^{2ik^2t} \overline{\Phi_2(t, \bar{k})}, \\ B(t, k) &= -e^{2ik^2t} \Phi_1(t, k) \end{aligned} \tag{9}$$

where the vector $\Phi(t, k)$ with components Φ_1 and Φ_2 is the following solution of the t -problem of the associated Lax pair evaluated at $x = 0$:

$$\begin{aligned} \Phi_t + 2ik^2\sigma_3\Phi &= \tilde{Q}(0, t, k)\Phi \\ 0 < t < T, \quad k \in \mathbb{C} \\ \Phi(0, k) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \tilde{Q}(0, t, k) &= \begin{pmatrix} -|g_0(t)|^2 & 2kg_0(t) + i\lambda\lambda g_1(t) \\ 2k\bar{g}_0(t) - i\lambda\bar{g}_1(t) & |g_0(t)|^2 \end{pmatrix} \end{aligned} \tag{10}$$

- Given $a(k), b(k)$ and $A(k), B(k)$, define a 2×2 matrix Riemann–Hilbert problem. This problem has the distinctive feature that its jump has explicit (x, t) dependence in the exponential form of $\exp\{ikx + 2ik^2t\}$. Determine $q(x, t)$ in terms of the solution of this Riemann–Hilbert problem by using the fact that these functions are related by the Lax pair. Then the function $q(x, t)$ solves the IBV problem [1]–[7] with $q(x, 0) = q_0(x), q(0, t) = g_0(t)$, and $q'_x(0, t) = g_1(t)$.

The above construction can be summarized in the following theorem (Fokas 2002):

Theorem 1 Consider the boundary-value problem for the NLS equation [1] determined by the conditions [7]. Let $a(k), b(k)$ be given by [8], and suppose that there exists a function $g_1(t)$ such that if $A(k), B(k)$ are defined by [9], then the global relation [8] holds.

Let $M(x, t, k)$ be the solution of the 2×2 Riemann–Hilbert problem with jump on the real and imaginary axes given by

- $M_-(x, t, k) = M_+(x, t, k)J(x, t, k)$ with $M = M_-$ in the second and fourth quadrants of \mathbb{C} , $M = M_+$ in the first and third quadrants of \mathbb{C} , and $J(x, t, k)$ is defined in terms of a, b, A, B and the exponential $e^{ikx - 2ik^2t}$.
 - $M = I + O(1/k)$ as $k \rightarrow \infty$ and has appropriate residue conditions if there are poles
- Then $M(x, t, k)$ exists and is unique, and

$$q(x, t) = 2i \lim_{k \rightarrow \infty} (kM(x, t, k))_{12}$$

The result above relies on characterizing the unknown boundary value $g_1(t)$ *a priori* by requiring that the global relation hold. Recently, substantial progress has been made in this direction in the case of integrable nonlinear evolution equations, in particular of NLS. Namely Fokas (in press) contains an effective description of the map assigning to each given $q(x, 0) = q_0(x)$ and $g_0(t) = q(0, t)$ a unique value for $q_x(0, t)$ (called the Dirichlet to Neumann map) for the NLS, as well as for a version of the Korteweg–deVries and sG equations. We state below the relevant theorem for the case of the NLS equation.

Theorem 2 *Let $q(x, t)$ satisfy the NLS equation on the half-line $0 < x < \infty, t > 0$ with the initial and boundary conditions [7]. Then $g_1(t) := q_x(0, t)$ is given by*

$$g_1(t) = \frac{g_0(t)}{\pi} \int_{\partial D} e^{-2ik^2t} (\Phi_2(t, k) - \Phi_2(t, -k)) dk \\ + \frac{4i}{\pi} \int_{\partial D} e^{-2ik^2t} k R(k) \overline{\Phi_2(t, \bar{k})} dk \\ + \frac{2i}{\pi} \int_{\partial D} e^{-2ik^2t} (k[\Phi_1(t, k) - \Phi_1(t, -k)] + ig_0(t)) dk$$

with $\Phi = (\Phi_1, \Phi_2)^T$ given by the solution of [10]. The Neumann datum $g_1(t)$ is unique and exists globally in t .

This result yields a rigorous proof of the global existence of the solution of boundary-value problems on the half-line for the NLS equation. Therefore, the assumption in Theorem 1 that a suitable function $g_1(t)$ exists can be dropped.

Generalizations and Summary of Results

Results analogous to the ones presented in the previous section can be phrased exclusively in terms of integral equations rather than in terms of Riemann–Hilbert problems, as done for example in Khruslov and Kotlyarov (2003). This is the point of view of the school of Gelfand and Marchenko, and in this setting the functions Φ are given in the so-called Gelfand–Levitan–Marchenko representation. Results on boundary-value problems for the NLS equation using this representation have been obtained only under additional assumptions on the unknown part of the boundary values. It was only after the idea that the x - and t -parts of the spectral equations should be treated simultaneously that this approach yielded complete results. However, the Gelfand–Levitan–Marchenko representation yields a crucial simplification for deriving the explicit form of the Dirichlet to Neumann map and proving Theorem 2. This

representation has now been derived for all equations [1]–[3], see Fokas (in press).

The analysis of the invariance properties of the global relation with respect to k also yields the characterization of all the boundary conditions for which the transform obtained to represent the solution linearizes. For these boundary conditions, called linearizable, the solution can be represented as effectively as for the Cauchy problem. For example, the linearizable boundary conditions for the NLS equation are given by any boundary values that satisfy

$$g_0(t) \overline{g_1(\bar{t})} - \overline{g_0(\bar{t})} g_1(t) = 0$$

An example of boundary condition satisfying this constraint, encompassing also Dirichlet and Neumann homogeneous conditions, is $q(0, t) - \chi q_x(0, t) = 0$, with χ a non-negative constant.

As mentioned at the beginning of the previous section, the approach described in general can be used to obtain results similar to those given for the NLS equation for many other integrable evolution equations, in particular, mKdV (Boutet de Monvel *et al.* 2004), sG, and KdV (Fokas 2002). The results obtained are essentially the same as for NLS, starting from the general form [5] of the Lax pair, and include the derivation of the solution representation, the complete characterization of linearizable boundary conditions, and the analysis of the Dirichlet to Neumann map.

The approach above can also be used for studying boundary-value problems posed on finite domains, for $x \in [0, 1]$. This has been done for a model for transient simulated Raman scattering (Fokas and Menyuk 1999), for the sG equation in light-cone coordinates (Pelloni, in press), and for the NLS equation (Fokas and Its 2004). In this case also the method yields a representation of the solution which is suitable for asymptotic analysis. In this respect, the question of soliton generation from boundary data is of some importance, and has been recently considered by various authors (Fokas and Menyuk 1999, Boutet de Monvel and Kotlyarov 2003, Pelloni in press, Boutet de Monvel *et al.* 2004). The results are however still considered case by case, and there is no general framework for this problem identified yet. For problem on the half-line, solitons may be generated but not necessarily in correspondence to the singularities that generate soliton for the full line problem, even when the same singularities are present. For problems posed on finite domains, in some specific cases at least for the simulated Raman scattering, and the sG equations, it appears that the dominant asymptotic behavior is given by a similarity solution.

In conclusion, the extension of the inverse scattering transform given by Fokas provides the tool for analyzing boundary-value problems specific to nonlinear integrable equations. This tool relies, in an essential way, on the integrability structure of the problem, and yields a full characterization of the solution as well as uniqueness and existence results. The solution representation thus obtained is not always fully explicit, but it is always suitable for asymptotic analysis using standard techniques such as the recent nonlinearization of the classical steepest descent method.

See also: $\bar{\partial}$ Approach to Integrable Systems; Integrable Discrete Systems; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Nonlinear Schrödinger Equations; Riemann–Hilbert Methods in Integrable Systems; Separation of Variables for Differential Equations; Sine-Gordon Equation.

Further Reading

- Ablowitz MJ and Segur HJ (1974) The inverse scattering transform: semi-infinite interval. *Journal of Mathematical Physics* 16: 1054.
- Adler VE, Gurel B, Gurses M, and Habibullin IT (1997) *Journal of Physics A* 30: 3505.
- Bona J, Sun S, and Zhang BY (2001) A non-homogeneous boundary value problem for the Korteweg–deVries equation. *Transactions of the American Mathematical Society* 354: 427–490.
- Boutet de Monvel A, Fokas AS, and Shepelsky D (2004) The modified KdV equation on the half-line. *Journal of the Institute of Mathematics of Jussieu* 3: 139–164.
- Boutet de Monvel A and Kotlyarov VP (2003) Generation of asymptotic solitons of the nonlinear Schrödinger equation by boundary data. *Journal of Mathematical Physics* 44: 3185–3215.
- Colin T and Ghidaglia J-M (2001) An initial-boundary value problem for the Korteweg–deVries equation posed on a finite interval. *Advanced Differential Equations* 6(12): 1463–1492.
- Colliander JE and Kenig CE (2001) The generalized Korteweg–deVries equation on the half line (<http://arxiv.org/abs/math.AP/0111294>).
- Degasperis A, Manakov S, and Santini PM (2001) The nonlinear Schrödinger equation on the half line. *JETP Letters* 74(10): 481–485.
- Fokas AS (2000) On the integrability of linear and nonlinear PDEs. *Journal of Mathematical Physics* 41: 4188.
- Fokas AS (2002) Integrable nonlinear evolution equations on the half line. *Communications in Mathematical Physics* 230: 1–39.
- Fokas AS (2005) A generalised Dirichlet to Neumann map for certain nonlinear evolution PDEs. *Communications on Pure and Applied Mathematics* 58: 639–670.
- Fokas AS and Its AR (2004) The nonlinear Schrödinger equation on the interval. *Journal of Physics A: Mathematical and General* 37: 6091–6114.
- Fokas AS and Menyuk CR (1999) Integrability and self-similarity in transient stimulated Raman scattering. *Journal of Nonlinear Science* 9: 1–31.
- Gardner GS, Greene JM, Kruskal MD, and Miura RM (1967) Method for solving the Korteweg–de Vries equation. *Physical Review Letters* 19: 1095.
- Habibullin IT (1999) KdV equation on a half-line with the zero boundary condition. *Theoretical and Mathematical Fizika* 119: 397.
- Khruslov E and Kotlyarov VP (2003) Generation of asymptotic solitons in an integrable model of stimulated Raman scattering by periodic boundary data. *Mat. Fiz. Anal. Geom.* 10(3): 366–384.
- Lax PD (1968) Integrals of nonlinear equations of evolution and solitary waves. *Communications in Pure and Applied Mathematics* 21: 467–490.
- Pelloni B (2005) The asymptotic behaviour of the solution of boundary value problems for the Sine–Gordon equation on a finite interval. *Journal of Nonlinear Mathematical Physics* 12: 518–529.
- Sklyanin EK (1987) Boundary conditions for integrable equations. *Functional Analysis and its Applications* 21: 86–87.
- Zakharov VE and Shabat AB (1972) An exact theory of two-dimensional self-focusing and one-dimensional automodulation of waves in a nonlinear medium. *Soviet Physics – JEPT* 34: 62–78.

Braided and Modular Tensor Categories

V Lyubashenko, Institute of Mathematics, Kyiv, Ukraine

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Tensor or monoidal categories are encountered in various branches of modern mathematical physics. First examples came without mentioning the name of a monoidal category as categories of modules over a group or a Lie algebra. The operation of a monoidal product in this case is the usual tensor product $X \otimes_{\mathbb{C}} Y$ of modules (representations) X and Y . These categories are symmetric: the modules $X \otimes Y$ and $Y \otimes X$ are

isomorphic; moreover, the permutation isomorphism (the twist) $c: X \otimes Y \rightarrow Y \otimes X$, $x \otimes y \rightarrow y \otimes x$, is involutive, $c^2 = \text{id}_{X \otimes Y}$. Next examples of monoidal categories were given by categories of representations of supergroups or Lie superalgebras. They are also symmetric: now the symmetry (Koszul’s rule) $c: X \otimes Y \rightarrow Y \otimes X$, $x \otimes y \rightarrow (-1)^{\deg x \cdot \deg y} y \otimes x$, is the twist with a sign, which depends on the degree (or parity) $\deg x$ of elements $x \in X$.

The development of the theory of exactly solvable models in statistical mechanics led Drinfeld (1987) to the notion of quantum groups – Hopf algebras H with additional structures (quasitriangular Hopf algebras). H -Modules also form a monoidal category; however, it is not symmetric, but only braided.

The coherence theorem of Mac Lane (1963) states that any monoidal category \mathcal{C} is equivalent to a strictly monoidal category, in which $X \otimes (Y \otimes Z) = (X \otimes Y) \otimes Z$, $1 \otimes X = X = X \otimes 1$, and the isomorphisms a, l, r are identity isomorphisms. Thus, in theoretical constructions, one may ignore the associativity isomorphism. It is not always so in practice. For instance, working with quasi-Hopf algebras related with the Knizhnik–Zamolodchikov equation one prefers to keep the original category, which is (a deformation of) the category of modules over a Lie algebra, rather than to replace it with a strict monoidal category, that is not a category of modules any more.

Definition 3 A rigid category \mathcal{C} is a monoidal category in which, to every object $X \in \mathcal{C}$, dual objects X^\vee and ${}^\vee X \in \mathcal{C}$ are assigned together with morphisms of evaluation and coevaluation

$$\begin{aligned} \text{ev}_X &: X \otimes X^\vee \rightarrow 1 = X \cup X^\vee \\ \text{ev}'_X &: {}^\vee X \otimes X \rightarrow 1 = {}^\vee X \cup X \\ \text{coev}_X &: 1 \rightarrow X^\vee \otimes X = X^\vee \cap X \\ \text{coev}'_X &: 1 \rightarrow X \otimes {}^\vee X = X \cap {}^\vee X \end{aligned}$$

The evaluations and coevaluations are chosen such that the compositions

$$\begin{aligned} X &\xrightarrow{r^{-1}} X \otimes 1 \xrightarrow{1 \otimes \text{coev}} X \otimes (X^\vee \otimes X) \xrightarrow{a} (X \otimes X^\vee) \otimes X \xrightarrow{\text{ev} \otimes 1} 1 \otimes X \xrightarrow{l} X \\ X &\xrightarrow{l^{-1}} 1 \otimes X \xrightarrow{\text{coev} \otimes 1} (X \otimes {}^\vee X) \otimes X \xrightarrow{a^{-1}} X \otimes ({}^\vee X \otimes X) \xrightarrow{1 \otimes \text{ev}'} X \otimes 1 \xrightarrow{r} X \\ X^\vee &\xrightarrow{l^{-1}} 1 \otimes X^\vee \xrightarrow{\text{coev} \otimes 1} (X^\vee \otimes X) \otimes X^\vee \xrightarrow{a^{-1}} X^\vee \otimes (X \otimes X^\vee) \xrightarrow{1 \otimes \text{ev}} X^\vee \otimes 1 \xrightarrow{r} X^\vee \\ {}^\vee X &\xrightarrow{r^{-1}} {}^\vee X \otimes 1 \xrightarrow{1 \otimes \text{coev}'} {}^\vee X \otimes (X \otimes {}^\vee X) \xrightarrow{a} ({}^\vee X \otimes X) \otimes {}^\vee X \xrightarrow{\text{ev}' \otimes 1} 1 \otimes {}^\vee X \xrightarrow{l} {}^\vee X \end{aligned}$$

are all identity morphisms.

In a rigid monoidal category \mathcal{C} , there is a pairing

$$\begin{aligned} (X \otimes Y) \otimes (Y^\vee \otimes X^\vee) &\xrightarrow{\sim} (X \otimes (Y \otimes Y^\vee)) \\ &\otimes X^\vee \xrightarrow{\text{X} \otimes \text{ev} \otimes X^\vee} (X \otimes 1) \otimes X^\vee \xrightarrow{r \otimes X^\vee} X \otimes X^\vee \xrightarrow{\text{ev}} 1 \end{aligned}$$

which induces an isomorphism $j_{+X,Y} : Y^\vee \otimes X^\vee \rightarrow (X \otimes Y)^\vee$, such that the above pairing coincides with

$$(X \otimes Y) \otimes (Y^\vee \otimes X^\vee) \xrightarrow{1 \otimes j_+} (X \otimes Y) \otimes (X \otimes Y)^\vee \xrightarrow{\text{ev}} 1$$

The equation

$$\begin{aligned} \text{coev}_{X \otimes Y} &= \left(1 \xrightarrow{\text{coev}_Y} Y^\vee \otimes Y \simeq Y^\vee \otimes 1 \otimes Y \right. \\ &\quad \left. \xrightarrow{1 \otimes \text{coev}_X \otimes 1} Y^\vee \otimes X^\vee \otimes X \otimes Y \right. \\ &\quad \left. \xrightarrow{j_+ \otimes 1} (X \otimes Y)^\vee \otimes (X \otimes Y) \right) \end{aligned}$$

also holds. Similarly, there is an isomorphism $j_{-X,Y} : {}^\vee Y \otimes {}^\vee X \rightarrow {}^\vee(X \otimes Y)$.

Morphisms constructed from braidings and (co-)evaluations are often described by tangles. The

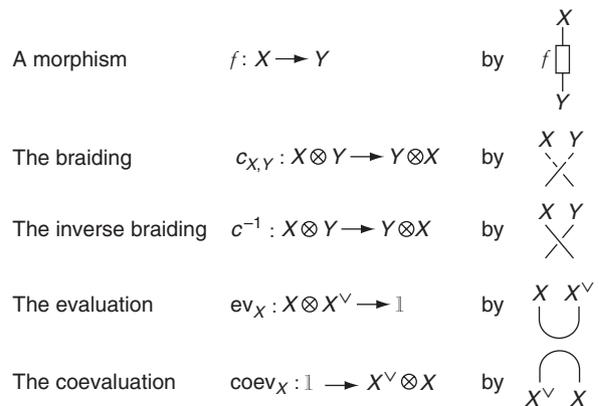
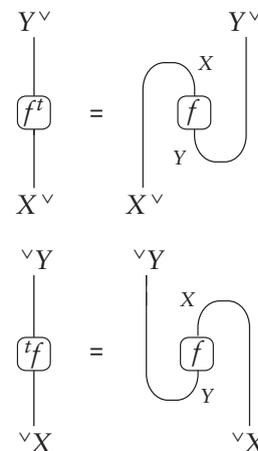


Figure 1 Conventions for notation of morphisms from tangles.

conventions are listed in **Figure 1**. The suggested assignment of morphisms in \mathcal{C} to elementary pictures extends to a unique functor Φ from the category of \mathcal{C} -colored tangles to the category \mathcal{C} itself. With the above interpretation, these tangles need not be oriented. We shall use the same notation for framed tangles, and the framing will be within the plane.

The maps $\text{Ob } \mathcal{C} \rightarrow \text{Ob } \mathcal{C}, X \mapsto X^\vee$, and $X \mapsto {}^\vee X$ extend to contravariant self-equivalences $\mathcal{C} \rightarrow \mathcal{C}$, $f \mapsto f^t$, and $f \mapsto {}^t f$. For given f , the morphisms f^t and ${}^t f$ can be defined, respectively, by the following pictures using the assignment from **Figure 1**:



We have a monoidal self-equivalence of \mathcal{C} ,

$$\begin{aligned} (-^{\vee\vee}, j_2) &: (\mathcal{C}, \otimes, 1) \rightarrow (\mathcal{C}, \otimes, 1), X \mapsto X^{\vee\vee}, f \mapsto f^{\vee\vee} \\ j_{2X,Y} &= \left(X^{\vee\vee} \otimes Y^{\vee\vee} \xrightarrow{j_+^t} (Y^\vee \otimes X^\vee)^\vee \xrightarrow{j_+^{\vee\vee}} (X \otimes Y)^{\vee\vee} \right) \quad [1] \end{aligned}$$

It is not always true that the two duals X^\vee and ${}^\vee X$ are isomorphic. However, there are canonical isomorphisms

$$X \rightarrow {}^\vee(X^\vee), \quad X \rightarrow ({}^\vee X)^\vee$$

We may replace the category \mathcal{C} with an equivalent one, such that the above isomorphisms become identity morphisms, and the functors $-^\vee$ and ${}^\vee-$ are inverse to each other. We shall assume this to simplify notations. Finally, we denote the iterated duals by $X^{(n\vee)} = X^{\vee \cdots \vee}$ (n times) and $X^{(-n\vee)} = {}^{\vee \cdots \vee}X$ (n times) for $n \geq 0$.

Braided Categories

Here we review the definitions of the braiding isomorphism and further derived isomorphisms. Several basic relations between them are listed. Two important classes of examples of braided categories are given by the categories of modules over quasitriangular Hopf algebras and the categories of tangles.

Definition 4 A braided category (\mathcal{C}, c) is a monoidal category \mathcal{C} equipped with a functorial isomorphism $c = c_{X,Y} : X \otimes Y \rightarrow Y \otimes X$ – the braiding, or the commutativity isomorphism – such that the two hexagons commute,

$$\begin{array}{ccc} X \otimes (Y \otimes Z) \xrightarrow{1 \otimes c^{\pm 1}} X \otimes (Z \otimes Y) \xrightarrow{a} (X \otimes Z) \otimes Y & & \\ a \downarrow & & \downarrow c^{\pm 1} \otimes 1 \\ (X \otimes Y) \otimes Z \xrightarrow{c^{\pm 1}} Z \otimes (X \otimes Y) \xrightarrow{a} (Z \otimes X) \otimes Y & & \end{array}$$

(one for c and one for c^{-1}).

The graphical notation for the braiding and its inverse is

$$c = (c_{X,Y} : X \otimes Y \rightarrow Y \otimes X) = \begin{array}{c} X & & Y \\ & \searrow & / \\ & & \text{---} \\ & / & \searrow \\ Y & & X \end{array}$$

$$c^{-1} = \begin{array}{c} X & & Y \\ & / & \searrow \\ & & \text{---} \\ & \searrow & / \\ Y & & X \end{array}$$

In a rigid braided category, we can define functorial isomorphisms using again the conventions from Figure 1:

$$u_1^2 = \begin{array}{c} X \\ | \\ \bigcirc \\ / \backslash \\ X^{\vee\vee} \end{array}, \quad u_{-1}^2 = \begin{array}{c} X \\ | \\ \bigcirc \\ / \backslash \\ X^{\vee\vee} \end{array}$$

$$u_{-1}^{-2} = \begin{array}{c} X \\ | \\ \bigcirc \\ / \backslash \\ {}^{\vee\vee}X \end{array}, \quad u_1^{-2} = \begin{array}{c} X \\ | \\ \bigcirc \\ / \backslash \\ {}^{\vee\vee}X \end{array}$$

These are isomorphisms of monoidal functors (see [1])

$$u_1^2 : (\text{Id}, c^{-2}) \rightarrow (-^{\vee\vee}, j_2)$$

$$u_{-1}^2 : (\text{Id}, c^2) \rightarrow (-^{\vee\vee}, j_2)$$

In particular, this implies the commutativity of the diagram

$$\begin{array}{ccc} X \otimes Y & \xrightarrow{c^{-2}} & X \otimes Y \\ u_1^2 \otimes u_1^2 \downarrow & & \downarrow u_1^2 \\ X^{\vee\vee} \otimes Y^{\vee\vee} & \xrightarrow{j_2} & (X \otimes Y)^{\vee\vee} \end{array}$$

The square of the monoidal functor $(-^{\vee\vee}, j_2)$ is

$$(-^{\vee\vee\vee\vee}, j_4) : (\mathcal{C}, \otimes, 1) \rightarrow (\mathcal{C}, \otimes, 1),$$

$$X \mapsto X^{\vee\vee\vee\vee}, \quad f \mapsto f^{\text{tttt}}$$

where

$$j_{4X,Y} = \left(X^{\vee\vee\vee\vee} \otimes Y^{\vee\vee\vee\vee} \xrightarrow{j_2} (X^{\vee\vee} \otimes Y^{\vee\vee})^{\vee\vee} \xrightarrow{j_2} (X \otimes Y)^{\vee\vee\vee\vee} \right)$$

The natural isomorphism $u_0^4 = u_{-1}^2 \circ u_1^2$ is, in fact, an isomorphism of monoidal functors $u_0^4 : (\text{Id}, \text{id}) \rightarrow (-^{\vee\vee\vee\vee}, j_4)$.

Ribbon Categories

Now we define balancing and recall some properties of balanced (ribbon) categories.

Definition 5 Let \mathcal{C} be a rigid braided category. A balancing $\beta_X : X \rightarrow X^{\vee\vee}$ is an isomorphism of monoidal functors $\beta : (\text{Id}, \text{id}, \text{id}) \rightarrow (-^{\vee\vee}, j_2, d_2)$ such that $\beta^2 = u_0^4$ and $\beta_X^t = \beta_{X^{\vee\vee}}^{-1} : X^{\vee\vee\vee\vee} \rightarrow X^{\vee}$. The category \mathcal{C} equipped with a balancing is called balanced.

We also use the notation $u_0^2 = \beta$. In any balanced category, there exists a canonical ribbon twist v . A ribbon twist $v = v_X : X \rightarrow X, v : \text{Id} \rightarrow \text{Id}$ is a self-adjoint ($v_{X^\vee} = v_X^t$) automorphism of the identity functor such that $c^2 = (v_X^{-1} \otimes v_Y^{-1}) \circ v_{X \otimes Y}$. It can be determined from the equations

$$u_0^2 = u_1^2 \circ v^{-1} = u_{-1}^2 \circ v : X \rightarrow X^{\vee\vee}$$

$$\beta^{-1} = u_0^{-2} = u_1^{-2} \circ v^{-1} = u_{-1}^{-2} \circ v : X \rightarrow {}^{\vee\vee}X$$

In particular, its square is given by the canonical isomorphism $v^2 = u_1^{-2} \circ u_1^2$. Conversely, in any rigid braided category with a ribbon twist (called ribbon category) there exists a canonical balancing u_0^2 given by the above formulas. Thus, ribbon categories and balanced categories are synonyms.

In the case of $X = 1$, we have $v_1 = \text{id}_1$.

The following result can be used to simplify notations:

Proposition 1 *For any ribbon category \mathcal{C} there exists a ribbon category \mathcal{D} equivalent to \mathcal{C} such that in it*

- (i) $1^\vee = 1$;
- (ii) for any object X we have ${}^\vee X = X^\vee, X^{\vee\vee} = X$, and $\beta_X = \text{id}_X : X \rightarrow X^{\vee\vee} = X$.
- (iii) for any object X we have $\text{ev}_X = \text{ev}'_{X^\vee} : X \otimes X^\vee \rightarrow 1$, and $\text{coev}_X = \text{coev}'_{X^\vee} : 1 \rightarrow X^\vee \otimes X$.

In the category $\mathcal{C} = H\text{-mod}$, where H is a ribbon Hopf algebra, the equation $X^\vee = {}^\vee X$ is not necessarily satisfied. Nevertheless, X^\vee is canonically isomorphic to ${}^\vee X$. The same holds in any ribbon category. We identify these objects via $\beta = u_0^2 : {}^\vee X \rightarrow X^\vee$. This allows us to use the right dual objects in place of the left ones. In that role, the right duals are equipped with the left evaluation and coevaluation, called flipped evaluation and coevaluation, respectively:

$$\begin{aligned} \tilde{\text{ev}} &: X^\vee \otimes X \xrightarrow{X^\vee \otimes \beta} X^\vee \otimes X^{\vee\vee} \xrightarrow{\text{ev}} 1 \\ \widetilde{\text{coev}} &: 1 \xrightarrow{\text{coev}} X^{\vee\vee} \otimes X^\vee \xrightarrow{\beta^{-1} \otimes X^\vee} X \otimes X^\vee \end{aligned}$$

They are often denoted simply ev and coev and should be replaced by $\tilde{\text{ev}}$ and $\widetilde{\text{coev}}$ in applications. In the context of Hopf algebra, β is given by the action of a group-like element introduced by Drinfeld.

Hopf Algebras in Braided Categories

Let \mathcal{C} be a braided monoidal category. A Hopf algebra H in \mathcal{C} is an object $H \in \text{Ob } \mathcal{C}$ together with an associative multiplication $m : H \otimes H \rightarrow H$ and an associative comultiplication $\Delta : H \rightarrow H \otimes H$, obeying the bialgebra axiom

$$\begin{aligned} & \left(H \otimes H \xrightarrow{m} H \xrightarrow{\Delta} H \otimes H \right) \\ &= \left(H \otimes H \xrightarrow{\Delta \otimes \Delta} H \otimes H \otimes H \otimes H \right. \\ & \quad \left. \xrightarrow{H \otimes c \otimes H} H \otimes H \otimes H \otimes H \right. \\ & \quad \left. \xrightarrow{m \otimes m} H \otimes H \right) \end{aligned}$$

Moreover, H has a unit $\eta : 1 \rightarrow H$, a counit $\varepsilon : H \rightarrow 1$, an antipode $\gamma : H \rightarrow H$, and the inverse antipode $\gamma^{-1} : H \rightarrow H$. The defining relations for these are the same as in the classical case. Notice, in particular, that the unit is also a morphism. Associativity of multiplication, as well as coassociativity of comultiplication, is formulated with the use of associativity isomorphism (in the nonstrict case).

Hopf algebras in braided categories have also been called braided groups. Their basic properties

are very similar to those of usual Hopf algebras, for example, the antipode is antimultiplicative with respect to the braiding (see, e.g., Majid (1993)). For Hopf algebras in rigid braided categories, there exist integrals in a sense very much similar to the case of ordinary finite-dimensional Hopf algebras, as shown by Bespalov *et al.* (2000).

Modular Categories

Assume that a braided rigid monoidal category \mathcal{C} is equivalent as a category (with monoidal structure ignored) to the category of finite-dimensional modules over a finite-dimensional algebra. In particular, \mathcal{C} is abelian. Then there exists an object F in \mathcal{C} , equipped with a morphism $i_X : X \otimes X^\vee \rightarrow F$ for each $X \in \text{Ob } \mathcal{C}$, such that the diagram

$$\begin{array}{ccc} X \otimes Y^\vee & \xrightarrow{f \otimes Y^\vee} & Y \otimes Y^\vee \\ X \otimes f^t \downarrow & & \downarrow i_Y \\ X \otimes X^\vee & \xrightarrow{i_X} & F \end{array}$$

is commutative for all morphisms $f : X \rightarrow Y$ of \mathcal{C} , and, moreover, F is universal between objects with such properties. Here $f^t : Y^\vee \rightarrow X^\vee$ is the transpose of a morphism $f : X \rightarrow Y$. In other words, F is a direct limit, called the coend and denoted as $F = \int^{Z \in \mathcal{C}} Z \otimes Z^\vee$. It can also be defined via an exact sequence

$$\bigoplus_{f: X \rightarrow Y \in \mathcal{C}} X \otimes Y^\vee \xrightarrow{f \otimes Y^\vee - X \otimes f^t} \bigoplus_{Z \in \mathcal{C}} Z \otimes Z^\vee \xrightarrow{\oplus i_Z} F \rightarrow 0$$

It turns out that the coend F is a Hopf algebra in the braided category \mathcal{C} , when it is equipped with the following operations. The comultiplication in F is uniquely determined by the equation

$$\begin{aligned} & \left(X \otimes X^\vee \xrightarrow{i_X} F \xrightarrow{\Delta} F \otimes F \right) \\ &= \left(X \otimes X^\vee = X \otimes 1 \otimes X^\vee \right. \\ & \quad \left. \xrightarrow{X \otimes \text{coev} \otimes X^\vee} X \otimes X^\vee \otimes X \otimes X^\vee \right. \\ & \quad \left. \xrightarrow{i_X \otimes i_X} F \otimes F \right) \end{aligned}$$

The counit in F is determined by the equation

$$\left(X \otimes X^\vee \xrightarrow{i_X} F \xrightarrow{\varepsilon} 1 \right) = \left(X \otimes X^\vee \xrightarrow{\text{ev}} 1 \right)$$

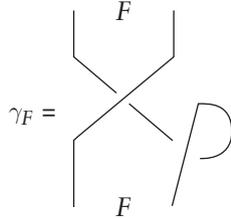
The multiplication $m : F \otimes F \rightarrow F$ is defined by the following diagram:

$$m = \begin{array}{c} X \\ | \\ X \end{array} \begin{array}{c} X^\vee \\ \diagdown \\ Y \\ \diagup \\ Y^\vee \\ | \\ Y \end{array} \begin{array}{c} Y \\ \diagdown \\ Y^\vee \\ | \\ Y^\vee \\ \diagup \\ X^\vee \\ | \\ X \end{array} \quad \text{and} \quad \begin{array}{ccc} X \otimes X^\vee \otimes (Y \otimes Y^\vee) & \xrightarrow{i_X \otimes i_Y} & F \otimes F \\ X \otimes c \downarrow & & \exists \downarrow m \\ X \otimes Y \otimes (X \otimes Y)^\vee & \xrightarrow{i_X \otimes Y} & F \end{array}$$

The unit is given by the morphism

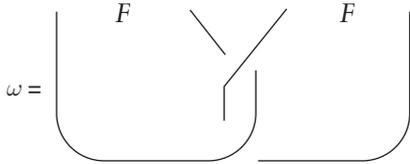
$$\eta : 1 = 1 \otimes 1^\vee \xrightarrow{i_1} F$$

The diagram corresponding to the antipode $\gamma_F : F \rightarrow F$ is given by



The structure of the coend F as a Hopf algebra can also be found directly from its universal property, as in Majid (1993).

There is a pairing of Hopf algebras $\omega : F \otimes F \rightarrow 1$ in \mathcal{C} :



It induces a homomorphism of Hopf algebras $F \rightarrow F^\vee$.

Definition 6 A ribbon category \mathcal{C} , equivalent as a category to the category of finite-dimensional modules over a finite-dimensional algebra, is called modular if the pairing ω is nondegenerate, that is, the induced morphism $F \rightarrow F^\vee$ is invertible.

Examples of nonsemisimple modular categories include $\mathcal{C} = H\text{-mod}$, where $H = u_q(\mathfrak{g})$ is a finite-dimensional algebra, quotient of the quantum universal enveloping algebra $U_q(\mathfrak{g})$, and q is a root of unity of odd degree. In these examples, the coalgebra F identifies with the dual Hopf algebra H^* , but the multiplication in F differs from that of H^* . Explicit formula for the multiplication in F uses the R -matrix for H (see, e.g., Majid (1993)). A definition of modularity for another type of categories (not necessarily abelian) was given by Turaev (1994).

When the category \mathcal{C} is modular, the integrals for the Hopf algebra F have especially simple properties. The integral element in F is two sided. It is a morphism $\mu : 1 \rightarrow F$ such that

$$\begin{aligned} & \left(F = F \otimes 1 \xrightarrow{1 \otimes \mu} F \otimes F \xrightarrow{m} F \right) \\ & = \left(F \xrightarrow{\varepsilon} 1 \xrightarrow{\mu} F \right) \\ & = \left(F = 1 \otimes F \xrightarrow{\mu \otimes 1} F \otimes F \xrightarrow{m} F \right) \end{aligned}$$

and μ is universal between morphisms with such property. By duality, the integral functional $\lambda : F \rightarrow 1$ is also two sided. It satisfies

$$\begin{aligned} & \left(F \xrightarrow{\Delta} F \otimes F \xrightarrow{1 \otimes \lambda} F \otimes 1 = F \right) \\ & = \left(F \xrightarrow{\lambda} 1 \xrightarrow{\eta} F \right) \\ & = \left(F \xrightarrow{\Delta} F \otimes F \xrightarrow{\lambda \otimes 1} 1 \otimes F = F \right) \end{aligned}$$

and is universal between morphisms with such property. The integral element and the integral functional are unique up to a multiplication by an element of $\text{Aut}_{\mathcal{C}} 1$.

Semisimple Abelian Modular Categories

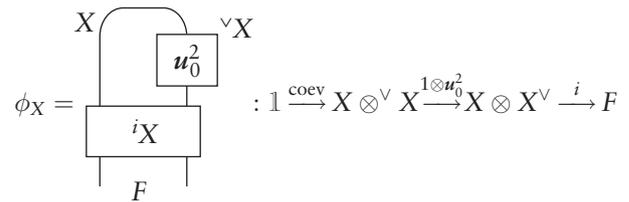
Reshetikhin and Turaev proposed to construct invariants of 3-manifolds via quantum groups. More precisely, they use certain abelian semisimple ribbon categories obtained from quantum groups at roots of unity as trace quotients. One can forget about the origin of these categories and work simply with semisimple modular categories. We shall describe them as input data for the modular functor construction.

Let \mathcal{C} be a \mathbb{C} -linear abelian semisimple modular ribbon category. Assume that the number of isomorphism classes of simple objects is finite. Assume also that 1 is simple and for each simple object X the endomorphism algebra $\text{End } X = \mathbb{C}$. We denote by $\mathcal{S} = \{X_i\}_i$ the list of (representatives of isomorphism classes of) all simple objects.

Under these assumptions, many formulas simplify. The coend $F \in \mathcal{C}$ takes the form

$$F = \bigoplus_{X \in \mathcal{S}} X \otimes X^\vee \in \mathcal{C}$$

Any morphism $1 \rightarrow F$ is a \mathbb{C} -linear combination of the standard morphisms for $X \in \mathcal{S}$,



The morphisms ϕ_X form a basis of the commutative algebra $\text{Inv } F = \text{Hom}_{\mathcal{C}}(1, F)$. The Grothendieck ring of the category \mathcal{C} determines the multiplication law in $\text{Inv } F$ via the algebra isomorphism $\mathbb{C} \otimes_{\mathbb{Z}} K_0(\mathcal{C}) \rightarrow \text{Inv } F, [X] \mapsto \phi_X$.

Any morphism $F \rightarrow 1$ can be represented as a linear combination of the morphisms

$$\psi_X : F \xrightarrow{\text{pr}_X} X \otimes X^\vee \xrightarrow{\text{ev}_X} 1$$

where $X \in \mathcal{S}$. The functional $\psi_1 : F \rightarrow \mathbb{1}$ satisfies the properties of a two-sided integral λ of the braided Hopf algebra F .

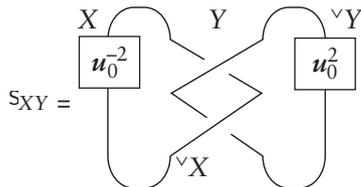
The Verlinde Formula

The number

$$\dim_q(X) = \begin{array}{c} X^\vee \\ \text{---} \\ \boxed{u_0^2} \\ \text{---} \\ X \end{array} : \mathbb{1} \xrightarrow{\text{coev}} X^\vee \otimes X \xrightarrow{1 \otimes u_0^2} X^\vee \otimes X^{\vee\vee} \xrightarrow{\text{ev}} \mathbb{1}$$

is called the dimension of an object $X \in \text{Ob } \mathcal{C}$. (The index q reminds us that this number coincides with the q -dimension in the case $\mathcal{C} = U_q(\mathfrak{g})\text{-mod.}$) We have $\dim_q(X^\vee) = \dim_q(X)$.

Definition 7 Introduce a biadditive function of two variables $s : \text{Ob } \mathcal{C} \times \text{Ob } \mathcal{C} \rightarrow \mathbb{C}$ on the class of objects of \mathcal{C} :



In particular, its restriction to \mathcal{S} is a matrix $s|_{\mathcal{S}} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{C}$, denoted again by $s = (s_{XY})_{X, Y \in \mathcal{S}}$ by abuse of notation; here X and Y run over simple objects.

Notice that $s_{XY} = s_{YX}$, so the matrix s is symmetric. Let us consider the \mathbb{C} -algebra $\text{Inv } F = \text{Hom}_{\mathcal{C}}(\mathbb{1}, F)$. It has the basis $\phi_X, X \in \mathcal{S}$; hence, it is n -dimensional, where $n = \text{Card } \mathcal{S}$. The form ω on F induces a bilinear form

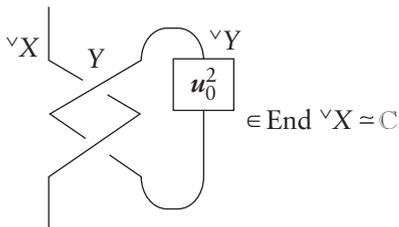
$$\omega' : \text{Inv } F \times \text{Inv } F \xrightarrow{\otimes} \text{Hom}(\mathbb{1}, F \otimes F) \xrightarrow{\text{Hom}(\mathbb{1}, \omega)} \mathbb{1}$$

The matrix (s_{XY}) is the matrix of the form ω' in the basis (ϕ_X) .

Lemma 1 (The Verlinde formula) *For any simple $X \in \mathcal{S}$ and any objects Y and Z of \mathcal{C} , we have*

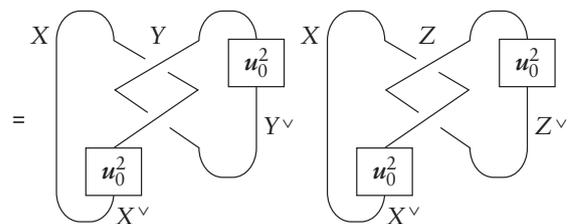
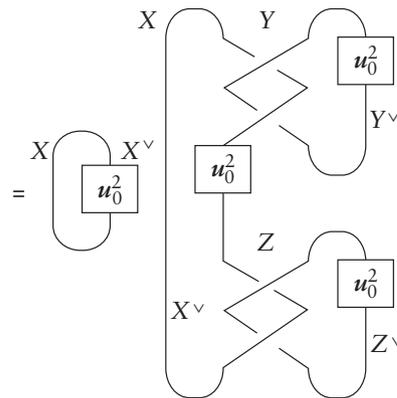
$$s_{X\mathbb{1}} = \dim_q(X), \quad s_{X\mathbb{1}}s_{X, Y \otimes Z} = s_{XY}s_{XZ} \quad [2]$$

Proof The first formula is straightforward. Since



is a number, we can move it from the second factor to the first in the following computation:

$s_{X\mathbb{1}}s_{X, Y \otimes Z}$



$$= s_{XY}s_{XZ}$$

This proves the second formula. □

Proposition 2 (Criterion of modularity) *In the above assumption of semisimplicity, the following conditions are equivalent:*

- (i) \mathcal{C} is modular (ω is nondegenerate);
- (ii) the matrix $(s_{XY})_{X, Y \in \mathcal{S}}$ is nondegenerate;
- (iii) for any $X \in \mathcal{S}$ its dimension $\dim_q X$ does not vanish, and there exist numbers $\mu'_Y, Y \in \mathcal{S}$, such that for all $X \in \mathcal{S}$ we have $\sum_{Y \in \mathcal{S}} s_{XY} \mu'_Y = \delta_{X\mathbb{1}}$; and
- (iv) for each simple $X \neq \mathbb{1}$ we have $\sum_{Y \in \mathcal{S}} s_{XY} \dim_q Y = 0$ and $\dim_q X \neq 0$.

The easy implication (ii) \implies (iii) can be deduced from the Verlinde formula. If the dimension $\dim_q(X) = s_{X\mathbb{1}}$ of a simple object X vanishes, then $s_{XY}^2 = 0$ for all $Y \in \text{Ob } \mathcal{C}$. This contradicts to the assumption of nondegeneracy of (s_{XY}) .

Let us determine the coefficients μ_Y of the integral element

$$\mu = \sum_{Y \in \mathcal{S}} \mu_Y \phi_Y : \mathbb{1} \rightarrow F$$

of the Hopf algebra F . It also has a two-sided integral-functional $\lambda : F \rightarrow \mathbb{1}$. The corresponding endomorphism is

$$\tilde{\lambda}_Z = \left(Z \xrightarrow{\delta_Z} F \otimes Z \xrightarrow{\lambda \otimes Z} \mathbb{1} \otimes Z = Z \right)$$

for an arbitrary object Z of \mathcal{C} , where δ_Z is the natural coaction. The equation

follows from the properties of the two-sided integral λ of the Hopf algebra F . Due to uniqueness of integrals, λ is proportional to ψ_1 . In eqn [3], X and Y vary over \mathcal{S} . The right-hand side is the identity morphism if $X=Y$, and vanishes otherwise. Substituting the definition of ϕ_Y , we rewrite the equation as follows:

For $X=1$, we get

$$\mu_Y \cdot \tilde{\lambda}_Y = \delta_{1Y} \cdot \text{id}_Y : Y \rightarrow Y \tag{5}$$

If $Y \neq 1$, then $\tilde{\lambda}_Y = 0$. So [5] tells essentially that

$$\mu_1 \cdot \tilde{\lambda}_1 = \text{id}_1 : 1 \rightarrow 1 \tag{6}$$

Now return to [4] with $X=Y$. If we compose that equation with $\text{coev} : 1 \rightarrow Y^v \otimes Y$, we obtain

Multiplying both sides of [7] with μ_1 , we find

$$\mu_Y = \mu_1 \cdot \text{dim}_q(Y)$$

The normalization is fixed by eqn [6], which we can write as

Hence,

$$(\mu_1)^2 = \left(\sum_{Y \in \mathcal{S}} (\text{dim}_q(Y))^2 \right)^{-1} \tag{8}$$

So, we find μ_1 , unique up to a sign.

Conjugation Properties

From the Verlinde formula [2], we conclude that the commutative \mathbb{C} -algebra $\text{Inv } F$ possesses homomorphisms

$$\begin{aligned} \chi_X : \text{Inv } F &\rightarrow \mathbb{C} \\ \phi_Y &\mapsto (\text{dim}_q(X))^{-1} s_{XY} = s_{XY}/s_{X1} \end{aligned}$$

The matrix \mathbf{s} is invertible, so that its columns cannot be proportional. Hence, all χ_X are different characters. Their number is $n = \text{Card } \mathcal{S} = \text{dim}_{\mathbb{C}} F$; hence, there is an isomorphism of \mathbb{C} -algebras

$$\begin{aligned} \chi : \text{Inv } F &\rightarrow \mathbb{C} \times \dots \times \mathbb{C} = \mathbb{C}^n \\ \phi &\mapsto (\chi_1(\phi), \dots, \chi_n(\phi)) \end{aligned}$$

Now we show that the dimensions $\text{dim}_q(Y)$ are real numbers, so that μ_1 is also a real number. One can introduce in $\text{Inv } F$ an antilinear involution,

$$-* : \text{Inv } F \rightarrow \text{Inv } F, \quad (\phi_X)^* = \phi_{X^v}$$

and a scalar (Hermitian) product

$$(\phi_X | \phi_Y) = \delta_{XY}, \quad X, Y \in \mathcal{S}$$

Then $\text{Inv } F$ becomes a finite-dimensional commutative Hilbert algebra. Indeed,

$$\begin{aligned} (\phi_X \phi_Y | \phi_Z) &= \text{dim Hom}(X \otimes Y, Z) \\ &= \text{dim Hom}(X, Y^v \otimes Z) = (\phi_X | \phi_Y^* \phi_Z) \end{aligned}$$

From the theory of finite-dimensional commutative Hilbert algebras, we know that idempotents in the algebra $\text{Inv } F$ are self-adjoint (only in that case the scalar product can be positive definite). Hence, χ is a $*$ -morphism, that is, $\chi_X(\phi^*) = \overline{\chi_X(\phi)}$. Therefore,

$s_{XY^\vee}/s_{X1} = \overline{s_{XY}}/\overline{s_{X1}}$. In the particular case of $X = 1$, we obtain

$$\dim_q(Y) = \dim_q(Y^\vee) = s_{1Y^\vee} = \overline{s_{1Y}} = \overline{\dim_q(Y)}$$

since $s_{11} = 1$. This proves that for any $Y \in \mathcal{C}$ its dimension $\dim_q(Y)$ is a real number.

It is natural to take for μ_1 the positive root of the right-hand side of [8]. Positiveness fixes μ_1 uniquely.

Examples of Semisimple Modular Categories

In their original paper, Reshetikhin and Turaev (1991) use as algebraic input data the representation theory of the quantum deformation $U = U_q(\mathfrak{sl}_2)$ of the Lie algebra $\mathfrak{sl}(2, \mathbb{C})$, where q is a root of unity. They construct the invariant as a trace over U -equivariant morphisms, and prove the necessary modularity condition concerning the nondegeneracy of the braided pairing.

The general picture is drawn by Turaev (1994), where 3-manifold invariants and TQFTs are constructed from semisimple modular categories. He shows how to obtain the latter as quotients of certain subcategories of representations of a modular Hopf algebra by the ideal of trace-negligible morphisms.

Finkelberg (1996), based on results of Gelfand and Kazhdan, establishes (via the theory of Kazhdan and Lusztig) an equivalence between two modular categories. The first is the semisimple category \mathcal{C} of integrable modules over an affine Lie algebra $\hat{\mathfrak{g}}$ of positive integer level k . The second is a certain subquotient of the category of $U_q(\mathfrak{g})$ -modules for $q = \exp(\pi i m^{-1}/(k + h^\vee))$, where $m \in \{1, 2, 3\}$ and h^\vee is the dual Coxeter number of \mathfrak{g} . Huang and Lepowsky (1999) describe the rigid braided structure of \mathcal{C} using vertex operators. Bakalov and Kirillov (2001) use geometrical constructions to make \mathcal{C} into a modular category, associated with the Wess–Zumino–Witten (WZW) model. They construct the corresponding WZW modular functor.

Modular Functor and TQFT

Modular categories give rise to a modular functor and a TQFT. The meanings of those differ from author to author, but the common features are the following. Such a TQFT is a functor from the category whose objects are smooth surfaces with additional structures and morphisms are three-dimensional manifolds with additional structures to the category of vector spaces. A modular functor is the restriction of such TQFT to the subcategory whose morphisms are homeomorphisms of surfaces. One of

the constructions due to Kerler and Lyubashenko (2001) takes a nonsemisimple modular category as an input and assigns to it a double TQFT functor, that is, a functor between double categories. The target is the 2-category of abelian categories.

See also: Axiomatic Approach to Topological Quantum Field Theory; Hopf Algebras and q -Deformation Quantum Groups; The Jones Polynomial; Knot Invariants and Quantum Gravity; Quantum 3-Manifold Invariants; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Topological Quantum Field Theory: Overview; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; von Neumann Algebras: Subfactor Theory.

Further Reading

- Bakalov B and Kirillov A Jr. (2001) *Lectures on Tensor Categories and Modular Functors*, University Lecture Series, vol. 21. Providence, RI: American Mathematical Society.
- Bespalov Y, Kerler T, Lyubashenko VV, and Turaev VG (2000) Integrals for braided Hopf algebras. *Journal of Pure and Applied Algebra* 148(2): 113–164 (arXiv:math.QA/9709020).
- Drinfeld VG (1987) Quantum groups. In: Gleason A (ed.) *Proceedings of the International Congress of Mathematicians (Berkeley, 1986)*, vol. 1, pp. 798–820. Providence, RI: American Mathematical Society.
- Drinfeld VG (1989a) Quasi-Hopf algebras. *Algebra i Analiz* 1(6): 114–148.
- Drinfeld VG (1989b) Quasi-Hopf algebras and Knizhnik–Zamolodchikov equations. In: *Problems of Modern Quantum Field Theory*, pp. 1–13. Berlin–New York: Springer.
- Finkelberg M (1996) An equivalence of fusion categories. *Geometric and Functional Analysis* 6(2): 249–267.
- Huang Y-Z and Lepowsky J (1999) Intertwining operator algebras and vertex tensor categories for affine Lie algebras. *Duke Mathematical Journal* 99(1): 113–134 (arXiv:q-alg/9706028) (arXiv:q-alg/9706028).
- Joyal A and Street RH (1991) Tortile Yang–Baxter operators in tensor categories. *Journal of Pure and Applied Algebra* 71: 43–51.
- Kerler T and Lyubashenko VV (2001) *Non-Semisimple Topological Quantum Field Theories for 3-Manifolds with Corners*, Lecture Notes in Mathematics, vol. 1765, vi+379 pp. Heidelberg: Springer.
- Mac Lane S (1971) *Categories for the Working Mathematician*, GTM, vol. 5. New York: Springer.
- Majid S (1993) Braided groups. *Journal of Pure and Applied Algebra* 86(2): 187–221.
- Majid S (1995) *Foundations of Quantum Group Theory*. Cambridge: Cambridge University Press.
- Moore G and Seiberg N (1989) Classical and quantum conformal field theory. *Communications in Mathematical Physics* 123: 177–254.
- Reshetikhin NY and Turaev VG (1991) Invariants of 3-manifolds via link polynomials and quantum groups. *Inventiones Mathematicae* 103(3): 547–597.
- Turaev VG (1994) Quantum Invariants of Knots and 3-Manifolds, *de Gruyter Stud. Math*, vol. 18. Berlin–New York: Walter de Gruyter.

Brane Construction of Gauge Theories

S L Cacciatori, Università di Milano, Milan, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Branes appear in string theories and M-theory as extended objects which contain some nonperturbative information about the theory, and, apart from gravity, they can couple with gauge fields.

At low energies, M-theory can be approximated with an 11-dimensional $N=1$ supergravity, which in fact is unique and contains a graviton field (the metric $g_{\mu\nu}$), a spin $3/2$ field ψ (the gravitino) and a gauge field consisting of a 3-form potential field c . The gauge field, whose field strength is a 4-form $G = dc$, can then couple electrically with two-dimensional extended objects, called M2 membranes. Moving in spacetime, an M2 membrane describes a three-dimensional world volume W_3 so that its coupling to the gauge field is

$$S_2 = k \int_{W_3} c \quad [1]$$

k representing the charge.

With c we can associate a dual field \tilde{c} such that $d\tilde{c} = *G$. It is a 6-form and can then electrically couple with a five-dimensional object, the M5 membrane. However, as c is the true field, we say that M5 couples magnetically with c .

In superstring theories, which however are related to M-theory by a dualities web, there are many more objects to be considered. In particular, we will consider type II strings, which at low energies are described by ten-dimensional $N=2$ supergravity theories. They contain a Neveu-Schwarz sector consisting of a graviton $g_{\mu\nu}$, a 2-form potential $B_{\mu\nu}$, and a scalar field ϕ , the dilaton. The content of the Ramond-Ramond fields depends on the chirality of the supercharges.

Type IIA strings are nonchiral (their left and right supercharges having opposite chiralities) and contain only odd-dimensional p -form potentials $A^{(p)}$, with $p = 1, 3, 5, 7, 9$.

Type IIB strings are chiral and contain only even-dimensional p -form potentials $A^{(p)}$, with $p = 0, 2, 4, 6, 8$.

Proceeding as before, we see that a $(p+1)$ -form potential can couple electrically with a p -dimensional object and magnetically with a $(6-p)$ -dimensional object. Such objects in fact exist in type II strings: the Dp branes are p -dimensional extended objects, with $p = 0, 2, 4, 6, 8$ for IIA strings and $p = -1, 1, 3, 5, 7, 9$ for IIB strings. In particular, D0 and D1 branes are

called D-particles and D-strings respectively, whereas $D(-1)$ branes are instantons, that is, points in spacetime. Concretely, D-branes are extended regions in spacetime where the endpoints of open strings are constrained to live. Mathematically, they are defined imposing Dirichlet conditions (whence the “D” of D-brane) on the ends of the string, along certain spatial directions. Excitation of these string states gives rise to the dynamic of the brane. They correspond to a ten-dimensional $U(1)$ gauge field, whose components, which are tangent to the brane world volume, give rise to a gauge field in $p+1$ dimensions, whereas the orthogonal components generate deformations of the brane shape. Moreover, if n parallel p -branes overlap, the gauge theory on the world volume is enhanced to a $U(n)$ gauge theory. Closed strings can generate gravitational interactions responsible for wrappings of the brane. However, in the cases when gravitational interaction is negligible, we can use this mechanism to construct $(p+1)$ -dimensional gauge theories, as we will see.

Before explaining how the construction works let us remember that there are two other interesting objects which often appear. In fact, we have not yet considered the Neveu-Schwarz B -field: this field can couple electrically with a one-dimensional object and magnetically with a five-dimensional object. These are the usual string (also called a fundamental or F-string) and a five-dimensional membrane called NS5 brane.

We will see how supersymmetric gauge theory configurations can be realized geometrically, considering more or less simple configurations of branes. We will also show that quantum corrections, be they exact or perturbative, can be described in this geometrical fashion. To be explicit, we will work with four-dimensional gauge theories, but it is clear that similar constructions can be done in different dimensions.

Gauge Groups on the Branes

A deeper understanding of how D-branes and related world-volume gauge theories work requires the introduction of dualities, but a quite simple heuristic argument can be given, giving up some rigor in favor of intuition.

To set our ideas, let us think of an open string moving in a nearly flat (but ten-dimensional) spacetime. Its trajectory will describe a two-dimensional surface having a boundary traced by the ends of the string (Figure 1). The string can then be described by a map from a two-dimensional surface Σ , having a

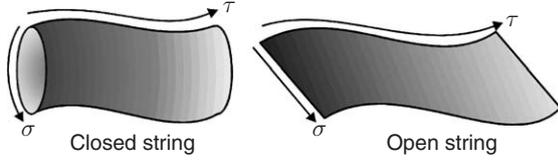


Figure 1 Strings moving in spacetime.

boundary $\gamma = \partial\Sigma$, to spacetime, say $X^\mu(\sigma, \tau)$ with $\mu = 0, 1, \dots, 9$. Here we chose on Σ local coordinates $\sigma^\alpha = (\sigma, \tau)$, where $\sigma \in [0, \pi]$ is a spacelike coordinate and τ is a timelike one. Then $\sigma = 0, \pi$ individuate the ends of the string and are identified for the closed string. Now, on a given background, the string evolution is usually described as a two-dimensional (supersymmetric) conformal field theory for the fields $X^\mu(\sigma, \tau)$. The action for the bosonic part is the same for both type IIA and IIB strings, and reads

$$S[X] = \frac{1}{4\pi\alpha'} \int_{\Sigma} d^2\sigma \sqrt{-h} h^{\alpha\beta} g_{\mu\nu}(X) \frac{\partial X^\mu}{\partial \sigma^\alpha} \frac{\partial X^\nu}{\partial \sigma^\beta} + \frac{1}{4\pi\alpha'} \int_{\Sigma} B_{\mu\nu}(X) \frac{\partial X^\mu}{\partial \sigma^\alpha} \frac{\partial X^\nu}{\partial \sigma^\beta} d\sigma^\alpha \wedge d\sigma^\beta \quad [2]$$

where $g_{\mu\nu}$ and B are the metric and a 2-form potential field for the given spacetime background, and $h_{\alpha\beta}$ is a metric for Σ . In general, we must also add a scalar field $\phi(X)$, but it will not play any role here. Using conformal invariance, we can reduce $h_{\alpha\beta}$ to the flat metric. Also consider a flat background $g_{\mu\nu}(X) = \eta_{\mu\nu}$ and concentrate for a moment on the B -field.

Conceived as a 2-form field over the spacetime, the potential field B is a gauge field: its field strength 3-form $H = dB$ is unchanged under a shift

$$B \longrightarrow B + dA \quad [3]$$

generated by the 1-form field $A(X)$. Here A should be a totally unphysical field. However, note that if one considers open strings, the action for the B -field, and then the full action is shifted by a boundary term

$$S[X] \longrightarrow S[X] + \frac{1}{4\pi\alpha'} \int_{\gamma} A_\mu(X) \frac{\partial X^\mu}{\partial \sigma^\alpha} d\sigma^\alpha \quad [4]$$

The boundary γ just describes the timelike world lines of the ends of the string. Thus, the ends of the string carry a U(1) charge and, even though the B -field vanishes, we can have the open-string action

$$S[X] = \frac{1}{4\pi\alpha'} \int_{\Sigma} \partial_\alpha X^\mu \partial^\alpha X_\mu d^2\sigma + \int_{\gamma} A_\mu(X) \partial_\alpha X^\mu d\sigma^\alpha \quad [5]$$

Here we conventionally rescaled the A field to normalize the action. To define the equation of motion, however, we must also specify boundary conditions for $X^\mu(\sigma, \tau)$ on γ . Let us choose Neumann conditions for $\mu = 0, 1, \dots, p$ and Dirichlet conditions for the remaining directions

$$\partial_\sigma X^a(\gamma) = 0, \quad a = 0, \dots, p \quad [6]$$

$$\partial_\sigma X^i(\gamma) = 0, \quad i = p + 1, \dots, 9 \quad [7]$$

This means that the extrema of the string are bound on a $(p + 1)$ -dimensional region (including time): the Dp brane. If for Σ we consider the full strip $(\sigma, \tau) = [0, \pi] \times \mathbb{R}$ then the U(1) action reduces to

$$S_A[X] = \int_{-\infty}^{\infty} A_a \partial_\tau X^a(\pi, \tau) - \int_{-\infty}^{\infty} A_a \partial_\tau X^a(0, \tau) \quad [8]$$

Thus, only the components of A_a tangent to the brane interact with the ends of the strings. What about the normal components A_i ?

To understand its meaning, let us proceed to compute the mean momentum transferred by the string, as it would be rigid. Imitating the Hamilton–Jacobi procedures for particles, let us consider the action up to a fixed time, say $\tau = 0$, so that $\Sigma = [0, \pi] \times [-\infty, 0]$. It is then a function of the position $X^\mu(\sigma, 0)$ of the string at the instant $\tau = 0$. To compute the momentum, we must vary the action by changing the position by a constant shift $\delta X^\mu(\sigma) = \Delta_0^\mu$. The variation will then contain some boundary terms which, for reasons of consistency, we must make vanish.

Before doing such a computation, let us make some further comments. It is plausible to assume that the two ends of the string could be charged for different U(1) fields. To the states of the open string we can in fact add two discrete labels $I, J = 1, \dots, n$, for some integer n , called Chan–Paton factors, and referring, respectively, to the two ends of the string. We will indicate the ends of the string as $X^\mu(0, \tau; I)$ and $X^\mu(\pi, \tau; J)$ when we need to specify the states. If the string is in the excited state (I, J) , then $X(0, \tau; I)$ can couple with the field A^I and $X(\pi, \tau; J)$ with $A^{(J)}$. For simplicity, we will now assume that these fields are constant. Note however that $A^{(I)}$ must be intended as a function of $X(0, \tau)$ only, and similarly for $A^{(J)}$. Also to realize the variation we can vary $X^\mu(\sigma, \tau)$ by a function $\delta X^\mu(\sigma, \tau) = \Delta^\mu(\tau)$ strictly picked to Δ_0^μ at $\tau = 0$ so that essentially

$$\partial_\tau \Delta^\mu(\tau) = \Delta_0^\mu \delta(\tau) \quad [9]$$

where $\delta(\tau)$ is the Dirac delta function.

Using the chosen boundary conditions, the variation of the full action contains the boundary terms

$$\begin{aligned} \delta S_{\text{bound}} &= (A_i^{(J)} - A_i^{(I)}) \int_{-\infty}^0 \partial_\tau \Delta^i(\tau) d\tau \\ &\quad + \frac{1}{2\pi\alpha'} \int_0^\pi \Delta^i \partial_\sigma X_i(\sigma, 0) d\sigma \\ &= \frac{\Delta^i}{2\pi\alpha'} \left[X_i(\pi, 0) - X_i(0, 0) \right. \\ &\quad \left. + 2\pi\alpha' (A_i^{(J)} - A_i^{(I)}) \right] \end{aligned} \quad [10]$$

Imposing the condition of its vanishing gives the physical interpretation for the normal components of the U(1) fields

$$X_i(\pi, 0) - X_i(0, 0) = -2\pi\alpha' (A_i^{(J)} - A_i^{(I)}) \quad [11]$$

This means that, up to a constant shift, the fields $A_i^{(K)}$ measure the positions of the ends of the strings in the transverse directions! (Figure 2). Equivalently, we can say that the string ends on two different Dp branes, parallel but displaced in the transverse directions by a quantity $-2\pi\alpha' (A_i^{(J)} - A_i^{(I)})$. We are thus also able to interpret the Chan–Paton factors. They mean that the string is living in a background of n parallel branes, stretched between the I th and the J th brane. On every brane, a U(1) gauge group lives so that the full gauge group is $U(1)^n$. However, when k of the branes overlap, the corresponding set of states become indistinguishable, so that the gauge group can be enhanced to a $U(k)$ group. In conclusion, n overlapping parallel Dp branes carry a $(p + 1)$ -dimensional $U(n)$ gauge theory which breaks in $U(k_i)$ block factors if the branes separate in stacks of k_i overlapping branes.

We can say a little bit more about this. If the string excited states represent gauge degree of freedom, they must become massive to break gauge symmetry when the branes separate. To see this, let us conclude by computing the mean momentum carried by the string. After elimination of the

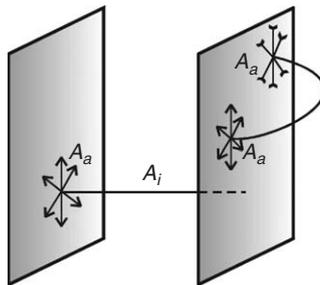


Figure 2 Tangential components of A_a appear as gauge modes. Normal components A_i appear as shift modes.

boundary terms, the total variation of the action due to the shift $\delta X^\mu(\sigma, 0) = \Delta^\mu$ becomes

$$\begin{aligned} \delta S &= \frac{1}{2\pi\alpha'} \int_\Sigma \partial_\tau \Delta^\mu \partial_\tau X_\mu d\sigma^2 \\ &= \frac{\Delta^\mu}{2\pi\alpha'} \int_0^\pi \partial_\tau X_\mu(\sigma, 0) d\sigma \end{aligned} \quad [12]$$

The resulting momentum is

$$P_\mu = \frac{1}{2\pi\alpha'} \int_0^\pi \partial_\tau X_\mu(\sigma, 0) d\sigma$$

On the bulk, the fields X^μ satisfy the standard wave equation in two dimensions, so that the general solution is the sum of a left-moving and a right-moving part, $X^\mu(\sigma, \tau) = X_L^\mu(\tau + \sigma) + X_R^\mu(\tau - \sigma)$. Imposing the boundary conditions, one finds

$$\begin{aligned} X^a(\sigma, \tau) &= X_L^a(\tau + \sigma) + X_L^a(\tau - \sigma) \\ &\quad + 2\pi\alpha' p^a \tau + X_0^a \end{aligned} \quad [13]$$

$$\begin{aligned} X^i(\sigma, \tau) &= X_L^i(\tau + \sigma) - X_L^i(\tau - \sigma) \\ &\quad + 2\alpha' (A^{(J)i} - A^{(I)i}) \sigma + X_0^i \end{aligned} \quad [14]$$

Here X_0^μ and p^a are integration constants and $X_L^i(\tau + \pi) - X_L^i(\tau - \pi) = 0$. A direct computation then shows that $P^a = p^a$ and $P^i = 0$, which is also what intuition suggests: the string can freely move along the branes but is fixed between them in the orthogonal directions. However, if it is stretched between two separated branes (i.e., if $I \neq J$), there is another contribution to the energy. In fact the factor $T := 1/(2\pi\alpha')$ represents the string tension, so that if Δ is its minimal length, its minimal contribution to the energy will be $\delta E = T\Delta$. This energy must equally contribute to the spectrum of the excited modes, the gauge field bosons. Here in fact, is where T -duality comes into play, but we will not discuss it.

The conclusion is that the spectrum corresponding to the stretched string must satisfy the condition $E \geq T\Delta$, which is as if the string states acquired a mass $T\Delta$, that is,

$$m^2 = \sum_{i=p+1}^9 (A^{(J)i} - A^{(I)i})^2 \quad [15]$$

This gives us a geometric tool to construct $(p + 1)$ -dimensional gauge theories: on n coincident Dp branes there exists a $U(n)$ gauge theory which can be broken separating the branes and thus giving a mass to the gauge bosons. Such a mass is proportional to the distance between the branes (Figure 3).

Before continuing with some examples, let us make two comments. First, the theory obtained in this way is a supersymmetric one, because the

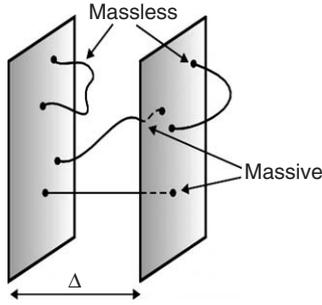


Figure 3 Stretched strings acquire a mass.

Dirichlet conditions allow the action of supersymmetric transformations of the form $\epsilon_L Q_L + \epsilon_R Q_R$, where Q_L and Q_R are the fermionic left and right supercharge operators and ϵ_L, ϵ_R are spinors satisfying the brane projection condition $\epsilon_L = \pm \Gamma^0 \Gamma^1 \cdots \Gamma^p \epsilon_R$. Here Γ^μ are the ten-dimensional Dirac matrices and one refers to “antibranes” for the negative sign.

Second, the gauge group can be converted into an $SO(n)$ or an $Sp(n/2)$ (for even n), adding an orientifold plane parallel to the branes. The orientifold plane acts on the orthogonal spacetime directions with a \mathbb{Z}_2 -action

$$X^i \sim -X^i \quad [16]$$

if $X^i = 0$ is the position of the orientifold. It further acts on the string world sheet as $\sigma \sim \pi - \sigma$ making it an unoriented string. The effect is to project out some states from the spectra, thus reducing the gauge group.

Geometric Engineering of Gauge Theories from Branes

To illustrate how brane construction of gauge theories works, we will consider a particular configuration of branes (Witten 1997).

We would like to obtain a four-dimensional $U(n)$ gauge theory. A possibility could be to take n D3 branes in a type IIB string background. However, such a model would contain too many supersymmetries: in ten dimensions, supersymmetries are generated by two 16-dimensional chiral spinors ϵ_L, ϵ_R ($\Gamma^0 \cdots \Gamma^9 \epsilon_{L,R} = \epsilon_{L,R}$). From the four-dimensional point of view, each of them represents four four-dimensional spinors giving an $N = 8$ supersymmetric theory. The projection condition, due to the branes, reduces the number of supersymmetries to four. Supersymmetry not being manifest in nature, it is desirable to have fewer supersymmetric gauge theories at hand. Because different brane projection conditions can further reduce supersymmetry, we

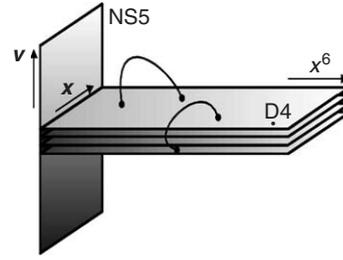


Figure 4 D4 branes ending on an NS5 brane. Gauge degrees of freedom are frozen in four dimensions.

can try to consider the coexistence of more kinds of branes.

One way to do this is to consider n parallel 4-branes ending on an NS5 brane in type IIA string theory (Figure 4), and then analyze the gauge theory restricted to the four-dimensional intersection (here the theory is nonchiral as $\Gamma^0 \cdots \Gamma^9 \epsilon_{L/R} = \pm \epsilon_{L/R}$). What kind of branes can end on other kind of branes can be established, starting from the fact that strings can end on a brane, and using the dualities tool (Giveon and Kutasov 1999).

Let us fix some conventions. We will indicate with $\mathbf{x} = (x^0, x^1, x^2, x^3) \in \mathbb{R}^4$ the coordinates on the intersection, so that $(\mathbf{x}; \mathbf{v}) = (\mathbf{x}; x^4, x^5) \in \mathbb{R}^6$ define the NS5 brane, and (\mathbf{x}, x^6) , with $x^6 \in [0, \infty)$, the 4-branes. Also v_I will indicate the position of the I th 4-brane on the 5-brane, and $\mathbf{y} = (x^7, x^8, x^9)$ will collect the remaining coordinates. Finally, we will indicate the product of Γ -matrices, corresponding to given directions, indicizing a simple Γ with the respective coordinates. For example $\Gamma^\nu = \Gamma^4 \Gamma^5$. With these conventions, the brane projection conditions for D4 and NS5 branes, respectively, read

$$\epsilon_L = \Gamma^x \Gamma^6 \epsilon_R \quad [17]$$

$$\epsilon_L = \Gamma^x \Gamma^\nu \epsilon_L, \quad \epsilon_R = \Gamma^x \Gamma^\nu \epsilon_R \quad [18]$$

These projections reduce supersymmetry to $N = 2$. After a short manipulation and using for example antichirality of ϵ_R , it is easy to see that the first condition can be substituted by

$$\epsilon_L = \Gamma^x \Gamma^y \epsilon_R \quad [19]$$

In other words, we could add a number of 6-branes in the (\mathbf{x}, \mathbf{y}) directions, without further reducing supersymmetry. We will consider this possibility later.

On the D4 branes there is an eventually broken $U(n)$ gauge theory. Here the vector fields A_μ , $\mu = 0, 1, 2, 3, 6$, and the scalar fields v_I and y live. The last ones are set to zero by the Dirichlet conditions, whereas v_I measure the fluctuations of the D3 brane positions over NS5. The $O(2)$ group

of rotations of the (x^4, x^5) coordinates acts on them, which can be broken by an expectation value $\langle v_I \rangle \neq 0$. The $SO(3)$ rotations of (x^6, x^7, x^8) (under which v_I are singlets) do not influence the projection conditions and can then be identified with the R-symmetry group $SU(2)_R$. It could be broken by a nonvanishing expectation value $\langle y \rangle \neq 0$, but as we said it cannot happen in the actual configuration. This highlights an unbroken supersymmetric Coulomb branch.

What is the physics as seen by an observer living on the four-dimensional spacetime x ? The components A_α , $\alpha = 0, 1, 2, 3$, of the vector fields transform as vectors with respect to the four-dimensional Lorentz group $SO(1, 3)$. They satisfy Neumann boundary conditions on $x^6 = 0$ and then survive as $U(n)$ gauge vector fields. The A_6 component behaves as a scalar with respect to $SO(1, 3)$ but is eliminated by a Dirichlet condition in $x^6 = 0$. The v scalar field will be responsible for the eventual breaking of the gauge group.

This seems to be quite a good scenario but actually the situation is unsatisfactory. If a 4-brane extends to the interval $[0, L]$ in the x^6 direction, the effective action for the gauge fields goes like this:

$$\begin{aligned} & \frac{1}{g_{D_4}^2} \int_0^L dx^6 \int_{\mathbb{R}^4} d^4x \text{tr} F_{\mu\nu} F^{\mu\nu} \\ & \approx \frac{L}{g_{D_4}^2} \int_{\mathbb{R}^4} d^4x \text{tr} F_{\alpha\beta} F^{\alpha\beta} \end{aligned} \quad [20]$$

where $\alpha, \beta = 0, 1, 2, 3$. Thus, the gauge coupling in four dimensions appears to be $g_4 = (g_{D_4})/\sqrt{L}$. In our case, where L goes to infinity, the gauge coupling vanishes and the gauge degrees of freedom are frozen. Moreover, an argument similar to the one made for the stretched strings shows that the energy of the D4 brane is very high and makes the mechanism of gauge group breaking difficult. The same is true for the NS5 brane, which also turns out to be extremely massive and does not participate in the dynamics. But this is what we want.

To solve the problem and restore gauge dynamics in four dimensions, one must consider a stack of 4-branes of finite length in the x^6 direction. This can be achieved placing in $x^6 = L$ a second NS5 brane parallel to the first one and in the same point in y (Figure 5). In this way, the D4 branes can stretch between the NS5 branes. If L is little enough, the gauge dynamics is restored also requiring a small value for g_{D_4} , to ensure the gravitational coupling (and the couplings with the Kaluza–Klein and NS5 modes) to be negligible. However, L must be bigger than the δX^6 fluctuations in order to avoid quantum corrections.

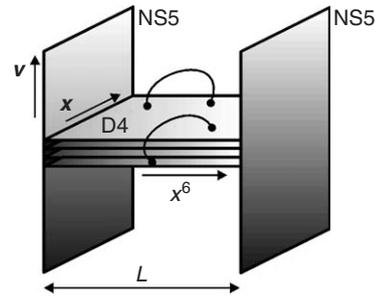


Figure 5 $N=2$ four-dimensional super Yang–Mills theory, with $U(n)$ gauge group.

What we just obtained is an $N=2$ supersymmetric classical $U(n)$ gauge theory in four dimensions, without matter, and in the Coulomb branch. Before considering quantization, let us briefly discuss some possible generalizations. For example, matter can be realized attaching to the left-hand side NS5 brane, new D4 branes parallel to the previous ones, but extended in the x^6 direction from $-\infty$ to 0 (Figure 6). Considering strings stretched between long and short branes, we obtain states whose half-gauge action, associated with the end connected to the long brane, is frozen. The corresponding states thus appear in the fundamental representation and can be interpreted as matter states.

To consider the Higgs branch, one should be able to break supersymmetry giving an expectation value to y . As mentioned above, in the actual configuration this cannot happen because y is set to 0 by Dirichlet conditions. Fortunately, as we said, one can add 6-branes in the (x, y) directions. If we insert such branes to stop the long D4 branes in a large but finite value of x^6 , say $x^6 = -M$ with $M \gg L$, then long branes have Neumann conditions in the y directions. Thus, fluctuations of the long branes can give an expectation value to y , breaking supersymmetry and subsequently the Higgs branch can be tuned, shifting 4-branes stretched between 6-branes (Figure 7).

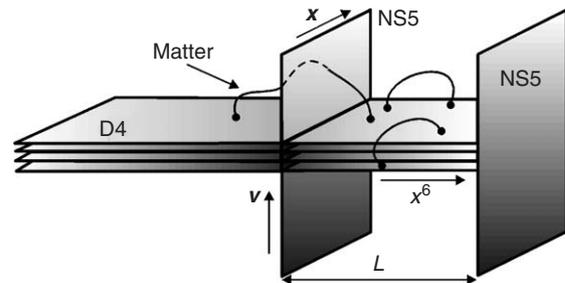


Figure 6 Adding matter.

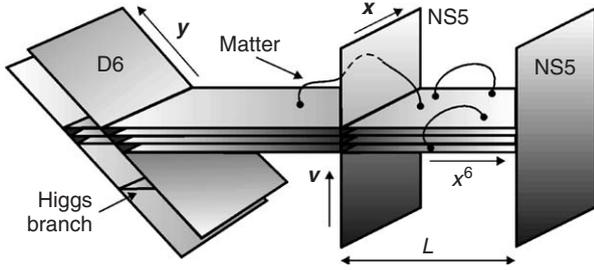


Figure 7 Permitting Higgs phases.

The details require some careful inspection, but we shall stop our analysis here (Giveon and Kutasov 1999).

More general gauge configurations can be realized by adding more parallel NS5 branes, and thus obtaining product groups. Adding orientifold planes, one can change gauge groups as explained in the previous section (Figure 8).

Finally, we can take a further step towards more physical models, constructing $N = 1$ gauge theories. For example, this can be achieved from the previous $N = 2$ model, rotating the second NS5 brane from the (x, v) position, to the (x, w) position, where $w = (x^8, x^9)$ (Figure 9). Then a new brane projection condition appears ($\epsilon_L = \Gamma^x \Gamma^w \epsilon_R$), breaking supersymmetry down to $N = 1$.

In this case, one could also obtain chiral matter, adding, for example, orientifold planes.

Quantum Corrections from M-Theory

Up to this point we have considered classical gauge configurations. Quantum corrections could be computed switching on brane fluctuations. However, it is an amusing fact that working with M-theory one can obtain exact quantum results. As an example, let us sketch how the exact Seiberg–Witten solution can be obtained for the $N = 2$ model described in the previous section, in the simplest case without matter.

The full web of dualities suggests the existence of a unique unifying theory called M-theory. At low energies, M-theory appears as the strong-coupling limit of type IIA strings. In such a limit, D0 branes become the dominant objects and the corresponding states can be interpreted as Kaluza–Klein modes coming from an eleventh dimension x^{10} compactified on a circle S^1 (Figure 10).

Thus, M-theory manifests itself as an 11-dimensional supergravity. In particular, it can be shown that there can be only a unique 11-dimensional supergravity. As said, here the nonperturbative objects are two- or five-dimensional membranes.

From the M-theory point of view, the D4 branes considered in our model appear as M5 membranes wrapped on the eleventh direction S^1 (Figure 11). Because quantum corrections are no longer negligible, we can no longer think of these branes as stretched in the x^6 direction, but v must also be considered. Thus, the M5 membranes will describe, in $\mathbb{R}^{10} \times S^1$, a region $\mathbb{R}^4 \times S$, where \mathbb{R}^4 are the x coordinates, and S is a Riemann surface immersed in $\mathcal{Q} \times S^1$, \mathcal{Q} being spanned by the (v, x^6) coordinates. In fact, supersymmetry constrains the surface to be a holomorphic curve, so that to describe it, it is convenient to collect $v = (x^4, x^5)$ and (x^6, x^{10}) into complex coordinates $v = x^4 + ix^5$ and $s = x^6 + ix^{10}$.

To compute quantum fluctuations, let us note that the end of a D4 brane over an NS5 brane is free to move along the v directions. A fully free end of a brane would satisfy a free wave equation. However, as x^6 is constrained in all directions but the v ones, it will simply satisfy a Laplace equation in two dimensions: $\Delta_v X^6 = 0$. Let us solve it, for a fixed NS5 brane. It will be (at least for large values of v)

$$x^6(v) = k \sum_{i=1}^{n_L} \log |v - v_{L_i}^{(\alpha)}| - k \sum_{i=1}^{n_R} \log |v - v_{R_i}^{(\alpha)}| \quad [21]$$

where n_L is the number of D4 branes ending on the left-hand side of the NS5 brane, in the positions $v_{L_i}^{(\alpha)}$, and similar for the R index, which refers to

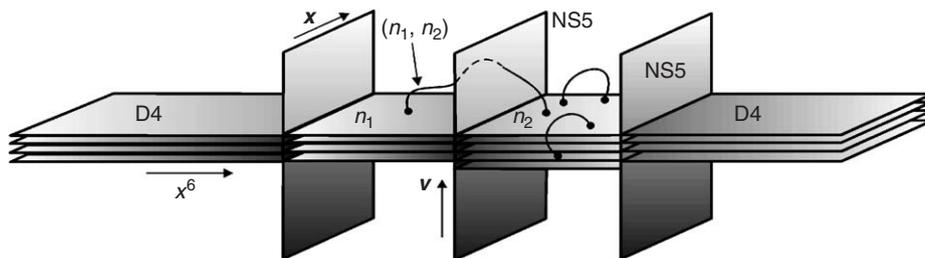


Figure 8 $N = 2$ four-dimensional super Yang–Mills theory with $U(n_1) \times U(n_2)$ gauge group and matter. Strings crossing the central NS5 brane give matter in the (n_1, n_2) representation.

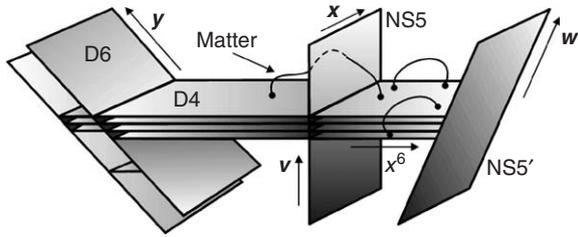


Figure 9 Going down to $N = 1$ supersymmetry.

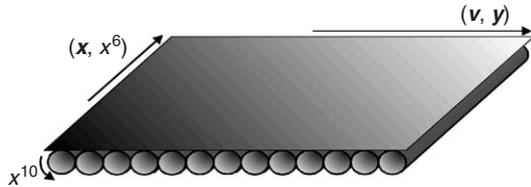


Figure 10 In M-theory one can think as if at any ten-dimensional spacetime point, there is attached an S^1 circle of ray R_{10} .

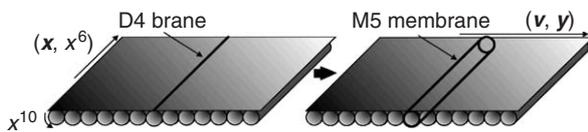


Figure 11 D4 branes become M5 membranes in M-theory.

the right-hand side. Here (α) refers to the α th NS5 brane, and k is an integration constant.

Because x^6 is the real part of a holomorphic field, whose imaginary part is compactified on a circle of ray R_{10} , we then find

$$s(v) = R_{10} \sum_{i=1}^{n_L} \log(v - v_{L_i}^{(\alpha)}) - R_{10} \sum_{i=1}^{n_R} \log(v - v_{R_i}^{(\alpha)}) \quad [22]$$

This describes the quantum fluctuations of the NS5 brane as seen in M-theory. In particular, because of the imaginary part of s , the ends of the D4 branes appear as vortices on the NS5 brane. In place of s , it is now convenient to introduce a new field $t := \exp(-s/R_{10})$ so that

$$t(v) = \frac{\prod_{i=1}^{n_R} (v - v_{R_i}^{(\alpha)})}{\prod_{i=1}^{n_L} (v - v_{L_i}^{(\alpha)})} \quad [23]$$

Before continuing, let us look a bit again at the classical limit. In this case, a fixed value of v will correspond to the position of a D4 brane, whereas a fixed value of s will correspond to the fixed position of an NS5 brane. The classical configuration is then

$$(s - s^{(1)})(s - s^{(2)}) \prod_{i=1}^n (v - v_i) = 0 \quad [24]$$

Here $s^{(\alpha)}$ are the positions of the NS5 branes, and the positions v_i of the D4 branes coincide for both the NS5 branes. Also, for large values of v , one has $t^{(1)} \approx v^n$ and $t^{(2)} \approx v^{-n}$.

Quantum mechanically, the configuration is determined in terms of v and t by the holomorphic curve S , which can be described as an algebraic curve $F(v, t) = 0$, generalizing the classical configuration. As there are two NS5 branes and n D4 branes, F must be a polynomial of degree 2 in t ,

$$F(v, t) = A_2(v)t^2 + A_1(v)t + A_0(v) \quad [25]$$

where A_a , $a = 1, 2, 3$, are all polynomials of degree n . Note that values of v such that A_1 vanishes give the solution $t = 0$, which corresponds to sending the right-hand side NS5 brane to ∞ . Similarly, $A_2 = 0$ sends the other NS5 brane to $-\infty$. To avoid these undesirable configurations, we can set $A_0 = A_2 = 1$. For A_1 , we can take the most general choice, up to an eventual shift in v , giving the quantum configuration

$$t^2 + [v^n + a_{n-2}v^{n-2} + \dots + a_1v + a_0]t + 1 = 0 \quad [26]$$

This realizes a quantum-mechanical correspondence between the M5 membrane configurations described by the given polynomials, and the $N = 2$ super Yang–Mills vacua. But this is also the claimed Seiberg–Witten curve. In particular, M-theory gives a concrete physical meaning for the support Riemann surfaces of the Seiberg–Witten solutions.

To conclude, let us make some further comments. It is clear how the construction can be extended for involving more configurations, for example, with more NS5 branes, or adding matter.

Also, we have seen that the geometrical picture which branes give of gauge theories extends at the quantum level.

A similar construction can be made for the $N = 1$ model, which also permits a full geometrical proof of the Seiberg duality at both classical and quantum levels.

Finally, we should note that there are also other methods, which work in spacetimes where extra dimensions are compactified. There, the branes wrap around certain singular loci which contain information about gauge symmetries (Lerche 1997).

See also: AdS/CFT Correspondence; Compactification of Superstring Theory; Gauge Theories from Strings; Noncommutative Geometry from Strings; Seiberg–Witten Theory; Supergravity; Superstring Theories; Supersymmetric Particle Models.

Further Reading

Giveon A and Kutasov D (1999) Brane dynamics and gauge theory. *Reviews of Modern Physics* 71: 983.
 Johnson CV (2003) *D-Branes*. Cambridge: Cambridge University Press.
 Lerche W (1997) Introduction to Seiberg–Witten Theory and Its Stringy Origin. *Nucl. Phys. Proc. Suppl. B* 55: 83.

Polchinski J (1998) *String Theory. Vol. 1: An Introduction to the Bosonic String*. Cambridge: Cambridge University Press.
 Polchinski J (2004) *String Theory. Vol. 2: Superstring Theory and Beyond*. Cambridge: Cambridge University Press.
 Witten E (1997) Solutions of four-dimensional field theories via M-theory. *Nuclear Physics B* 500: 3.
 Zwiebach B (2004) *A First Course in String Theory*. Cambridge: Cambridge University Press.

Brane Worlds

R Maartens, Portsmouth University, Portsmouth, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

At high enough energies, Einstein’s classical theory of general relativity breaks down, and will be superseded by a quantum gravity theory. The singularities predicted by general relativity in gravitational collapse and in the hot big bang origin of the universe are thought to be artifacts of the classical nature of Einstein’s theory, which will be removed by a quantum theory of gravity. Developing a quantum theory of gravity and a unified theory of all the forces and particles of nature are the two main goals of current work in fundamental physics. The problem is that general relativity and quantum field theory cannot simply be molded together. There is as yet no generally accepted (pre-)quantum gravity theory.

The quest for a quantum gravity theory has a long and thus far not very successful history. Many different lines of attack have been developed, each having a different way of dealing with the classical singularities that arise from point particles and smooth spacetime geometry. String theory does away with zero-dimensional point particles, and particles are modeled as different states of new fundamental objects, the one-dimensional strings. It turns out, however, that there is a price to pay – the number of spacetime dimensions must be greater than four for a consistent theory. When fermions are included, which leads to superstring theory, the required number of dimensions is ten – one time and nine space dimensions.

There are in fact five distinct (1+9)-dimensional superstring theories. In the mid-1990s, duality transformations were discovered that relate these superstring theories to each other and to the (1+10)-dimensional supergravity theory. This led to the conjecture that all of these theories arise as different limits of a single theory, which has come to be known as M theory. It was also discovered that

extended objects of higher dimension than strings play a fundamental role in the theory. These objects are known as “branes” (from membranes), and the relation between them and strings leads to a new picture of how gravity and matter may be connected in the universe. Roughly speaking, open strings describe the particles of the nongravitational sector, and their ends are attached to branes, while closed strings, which describe the graviton and associated particles of the gravitational sector, can move freely in all dimensions.

Thus, the observable universe could be a (1+3)-surface – a “brane,” embedded in a (1+3+d)-dimensional spacetime – the “bulk,” with standard-model particles and fields trapped on the brane, while gravity is free to access the bulk. Brane-world models offer a phenomenological way to test some of the novel predictions and corrections to general relativity that are implied by M theory.

Higher-Dimensional Gravity

Brane worlds can be seen as reviving the original higher-dimensional ideas of Kaluza and Klein in the 1920s, but in a new context of quantum gravity. An important consequence of extra dimensions is that the four-dimensional Planck scale $M_p \equiv M_{(4)} = 1.2 \times 10^{19}$ GeV is no longer the fundamental energy scale of gravity. The fundamental scale is instead $M_{(4+d)}$. This can be seen from the modification of the gravitational potential. For an Einstein–Hilbert gravitational action,

$$S_{\text{gravity}} = \frac{1}{2\kappa_{(4+d)}^2} \int d^4x d^d y \sqrt{-^{(4+d)}g} \times \left[^{(4+d)}R - 2\Lambda_{(4+d)} \right] \quad [1]$$

we have the higher-dimensional Einstein field equations,

$$\begin{aligned} ^{(4+d)}G_{AB} &\equiv ^{(4+d)}R_{AB} - \frac{1}{2} ^{(4+d)}R ^{(4+d)}g_{AB} \\ &= -\Lambda_{(4+d)} ^{(4+d)}g_{AB} + \kappa_{(4+d)}^2 ^{(4+d)}T_{AB} \quad [2] \end{aligned}$$

where $x^A = (x^a, y^1, \dots, y^d)$ and $\kappa_{(4+d)}^2$ is the gravitational coupling constant given by

$$\kappa_{(4+d)}^2 = 8\pi G_{(4+d)} = \frac{8\pi}{M_{(4+d)}^{2+d}} \quad [3]$$

The static weak field limit of the field equations leads to the $(4+d)$ -dimensional Poisson equation, whose solution is the gravitational potential

$$V(r) \propto \frac{\kappa_{(4+d)}^2}{r^{1+d}} \quad [4]$$

In the simplest scenario, we can assume a toroidal configuration for the d extra dimensions, with each compactified on the same length scale L . Then on scales $r \lesssim L$, the potential is $(4+d)$ -dimensional, $V \sim r^{-(1+d)}$. By contrast, on scales large relative to L , where the extra dimensions do not contribute to variations in the potential, V behaves like a four-dimensional potential, $V \sim L^{-d} r^{-1}$. This means that the usual Planck scale becomes an effective coupling constant, describing gravity on scales much larger than the extra dimensions, and related to the fundamental scale via the volume of the extra dimensions:

$$M_p^2 \sim M_{(4+d)}^{2+d} L^d \quad [5]$$

Large Extra Dimensions

If the extra-dimensional volume is significantly above the Planck scale, then the true fundamental scale $M_{(4+d)}$ can be much less than the effective scale M_p ,

$$L^d \gg M_p^{-d} \Rightarrow M_{(4+d)} \ll M_p \quad [6]$$

In this case, we understand the weakness of gravity as due to the fact that it “spreads” into extra dimensions, and only a part of it is felt in four dimensions.

A lower limit on $M_{(4+d)}$ is given by null results in table-top experiments to test for deviations from Newton’s law in four dimensions, $V \propto r^{-1}$. These experiments currently probe submillimeter scales, and find no detectable deviation, so that

$$\begin{aligned} L &\lesssim 10^{-1} \text{ mm} \sim (10^{-15} \text{ TeV})^{-1} \\ \Rightarrow M_{(4+d)} &\gtrsim 10^{(32-15d)/(d+2)} \text{ TeV} \end{aligned} \quad [7]$$

Stronger bounds can be derived from null results in particle accelerators in some brane-world models, or from constraints imposed by observations of supernovae or of light-element abundance.

Brane worlds, arising in the framework of string theory, thus incorporate the possibility that the

fundamental scale is much less than the Planck scale felt in four dimensions. This emerges by virtue of the large size of the extra dimensions. It is not necessary for all extra dimensions to be of equal size for this mechanism to operate. There are string theory solutions (Horava–Witten solutions) with two $(1+9)$ -branes located at the boundaries of the bulk, at the endpoints of an S^1/Z_2 orbifold, that is, a circle folded on itself across a diameter. The orbifold extra dimension is the large one, whereas the other six extra dimensions on the branes are compactified on a very small scale, close to the fundamental scale, and their effect on the dynamics is felt through “moduli” fields, that is, five-dimensional scalar fields.

These solutions can be thought of as effectively five dimensional, with an extra dimension that can be large relative to the fundamental scale. They provide the basis for the Randall–Sundrum 1 (RS1) phenomenological models of five-dimensional gravity. The single-brane Randall–Sundrum 2 (RS2) models with infinite extra dimension arise when the orbifold radius tends to infinity. The RS models are not the only phenomenological realizations of M theory ideas. They were preceded by the brane-world models of Arkani-Hamed, Dimopoulos, and Dvali (ADD), which put forward the idea that a large volume for the compact extra dimensions would lower the effective Planck scale $M_{(4+d)}$. If $M_{(4+d)}$ is close to the electroweak scale, M_{ew} , then this would address the long-standing “hierarchy” problem, that is, why there is such a large gap between $M_{\text{ew}} \sim 1 \text{ TeV}$ and $M_p \sim 10^{16} \text{ TeV}$.

In the ADD models, more than one extra dimension is required for agreement with experiments, and there is “democracy” among the equivalent extra dimensions, which, in addition, are flat. By contrast, the RS models have a “preferred” extra dimension, with other extra dimensions treated as ignorable (i.e., stabilized except at energies near the fundamental scale). Furthermore, this extra dimension is curved or “warped” rather than flat: the bulk is a portion of anti-de Sitter (AdS_5) spacetime. The RS branes are Z_2 -symmetric (mirror symmetry), and have a tension, which serves to counter the influence on the brane of the negative bulk cosmological constant. This also means that the self-gravity of the branes is incorporated in the RS models. The novel feature of the RS models compared to previous higher-dimensional models is that the observable three dimensions are protected from the large extra dimension (at low energies) by curvature (warping), rather than straightforward compactification.

The RS brane worlds provide phenomenological models that reflect at least some of the features of

M theory, and that bring exciting new geometric and particle physics ideas into play. The RS2 models also provide a framework for exploring holographic ideas that have emerged in M theory. Roughly speaking, holography suggests that higher-dimensional dynamics may be determined from a knowledge of the fields on a lower-dimensional boundary. The AdS/CFT correspondence is an example in which the classical dynamics of the higher-dimensional AdS gravitational field are equivalent to the quantum dynamics of a conformal field theory (CFT) on the boundary.

Kaluza–Klein Modes

The dilution of gravity via extra dimensions not only weakens gravity, it also broadens the range of graviton modes felt on the brane. The graviton is more than just the four-dimensional massless mode of four-dimensional gravity – other modes, with an effective mass on the brane, arise from the fact that the graviton is a $(4+d)$ -dimensional massless particle. These extra modes on the brane are known as Kaluza–Klein (KK) modes of the graviton.

For simplicity, consider a flat brane with one flat extra dimension, compactified through the identification $y \leftrightarrow y + 2\pi nL$, where $n = 0, 1, 2, \dots$. The perturbative five-dimensional graviton is defined via

$${}^{(5)}\eta_{AB} \rightarrow {}^{(5)}\eta_{AB} + h_{AB} \quad [8]$$

where ${}^{(5)}\eta_{AB}$ is the five-dimensional Minkowski metric and h_{AB} is a small transverse traceless perturbation. Its amplitude can be Fourier expanded as

$$h(x^a, y) = \sum_n e^{iny/L} h_n(x^a) \quad [9]$$

where h_n are the amplitudes of the KK modes, that is, the effective four-dimensional modes of the five-dimensional graviton. To see that these KK modes are massive from the brane viewpoint, we start from the five-dimensional wave equation that the massless five-dimensional field h satisfies (in a suitable gauge):

$${}^{(5)}\square h = 0 \Rightarrow \square h + \partial_y^2 h = 0 \quad [10]$$

It follows that the KK modes satisfy a four-dimensional Klein–Gordon equation with an effective four-dimensional mass, m_n :

$$\square h_n = m_n^2 h_n, \quad m_n = \frac{n}{L} \quad [11]$$

The massless mode, h_0 , is the usual four-dimensional graviton mode. But there is a tower of massive modes, $L^{-1}, 2L^{-1}, \dots$, which imprint the effect of the five-dimensional gravitational field on the four-dimensional brane. Compactness of the extra dimension leads to discreteness of the spectrum. For an infinite extra dimension, $L \rightarrow \infty$, the separation between the modes disappears and the tower forms a continuous spectrum.

Randall–Sundrum Brane Worlds

RS brane worlds do not rely on compactification to localize gravity at the brane, but on the curvature of the bulk. What prevents gravity from “leaking” into the extra dimension at low energies is a negative bulk cosmological constant,

$$\Lambda_{(5)} = -\frac{6}{\ell^2} = -6\mu^2 \quad [12]$$

where ℓ is the curvature radius of AdS_5 and μ is the corresponding energy scale. The bulk cosmological constant with its repulsive gravity effect acts to “squeeze” the gravitational potential closer to the brane. We can see this clearly in Gaussian normal coordinates $x^A = (x^\mu, y)$ based on the brane at $y=0$, for which the metric takes the form

$${}^{(5)}ds^2 = dy^2 + e^{-2|y|/\ell} \eta_{\mu\nu} dx^\mu dx^\nu \quad [13]$$

with $\eta_{\mu\nu}$ the Minkowski metric. The exponential warp factor reflects the confining role of the bulk cosmological constant. The Z_2 -symmetry about the brane at $y=0$ is incorporated via the $|y|$ term. In the bulk, this metric is a solution of the five-dimensional Einstein equations,

$${}^{(5)}G_{AB} = -\Lambda_{(5)} {}^{(5)}g_{AB} \quad [14]$$

that is, ${}^{(5)}T_{AB} = 0$ in eqn [2]. The brane is a flat Minkowski spacetime, $g_{AB}(x^\mu, 0) = \eta_{\mu\nu} \delta^\mu_A \delta^\nu_B$, with self-gravity in the form of brane tension.

The two RS models are distinguished as follows:

RS1 There are two branes in RS1, at $y=0$ and $y=L$, with Z_2 -symmetry identifications

$$y \leftrightarrow -y, \quad y + L \leftrightarrow L - y \quad [15]$$

The branes have equal and opposite tensions, $\pm\lambda$, where

$$\lambda = \frac{3}{4\pi} \frac{M_p^2}{\ell^2} \quad [16]$$

The positive-tension “TeV” brane has fundamental scale $M_{(5)} \sim 1 \text{ TeV}$. Because of the exponential

warping factor, the effective scale on the negative tension “Planck” brane at $y=L$ is M_p . On the positive tension brane,

$$M_p^2 = M_{(5)}^3 \ell \left[1 - e^{-2L/\ell} \right] \quad [17]$$

So RS1 gives a new approach to the hierarchy problem. Because of the finite separation between the branes, the KK spectrum is discrete.

RS2 In RS2, there is only one, positive-tension, brane. This may be thought of as arising from sending the negative tension brane off to infinity, $L \rightarrow \infty$. Then the energy scales are related via

$$M_{(5)}^3 = \frac{M_p^2}{\ell} \quad [18]$$

On the RS2 brane, the negative $\Lambda_{(5)}$ is offset by the positive brane tension λ . The fine-tuning in eqn [16] ensures that there is zero effective cosmological constant on the brane, so that the brane has the induced geometry of Minkowski spacetime. To see how gravity is localized at low energies, we consider the five-dimensional graviton perturbations of the metric:

$$\begin{aligned} {}^{(5)}g_{AB} &\rightarrow {}^{(5)}g_{AB} + h_{AB} \\ h_{Ay} &= 0 = h^\mu{}_\mu = \partial_\nu h^{\mu\nu} \end{aligned} \quad [19]$$

We split the amplitude h into three-dimensional Fourier modes, and the linearized five-dimensional Einstein equations lead to the wave equation ($y > 0$)

$$e^{2y/\ell} \left[\ddot{h} + k^2 h \right] = h'' - \frac{4}{\ell} h' \quad [20]$$

Separability means we can write

$$h(t, y) = \sum_m \varphi_m(t) h_m(y) \quad [21]$$

and the wave equation reduces to

$$\ddot{\varphi}_m + (m^2 + k^2) \varphi_m = 0 \quad [22]$$

$$h_m'' - \frac{4}{\ell} h_m' + e^{2y/\ell} h_m = 0 \quad [23]$$

The zero-mode solution is

$$\varphi_0(t) = A_{0+} e^{+ikt} + A_{0-} e^{-ikt} \quad [24]$$

$$h_0(y) = B_0 + C_0 e^{4y/\ell} \quad [25]$$

and the massive KK mode ($m > 0$) solutions are

$$\begin{aligned} \varphi_m(t) &= A_{m+} \exp\left(+i\sqrt{m^2 + k^2} t\right) \\ &+ A_{m-} \exp\left(-i\sqrt{m^2 + k^2} t\right) \end{aligned} \quad [26]$$

$$\begin{aligned} h_m(y) &= e^{2y/\ell} \left[B_m J_2\left(m\ell e^{y/\ell}\right) \right. \\ &\left. + C_m Y_2\left(m\ell e^{y/\ell}\right) \right] \end{aligned} \quad [27]$$

where J_2, Y_2 are Bessel functions.

The boundary condition for the perturbations is $h'(t, 0) = 0$, which implies

$$C_0 = 0, \quad C_m = -\frac{J_1(m\ell)}{Y_1(m\ell)} B_m \quad [28]$$

In the RS1 model, we have a further boundary condition, $h'(t, L) = 0$, which leads to a discrete eigenspectrum, namely the masses m that satisfy

$$J_1\left(m\ell e^{L/\ell}\right) Y_1(m\ell) - Y_1\left(m\ell e^{L/\ell}\right) J_1(m\ell) = 0 \quad [29]$$

The zero mode is normalizable, since

$$\left| \int_0^\infty B_0 e^{-2y/\ell} dy \right| < \infty \quad [30]$$

Its contribution to the gravitational potential $V = (1/2)h_{00}$ gives the four-dimensional result, $V \propto r^{-1}$. The contribution of the massive KK modes sums to a correction of the four-dimensional potential. For $r \ll \ell$, one obtains

$$V(r) \approx \frac{GM}{r} \left(1 + \frac{\ell}{r} \right) \approx \frac{GM\ell}{r^2} \quad [31]$$

which simply reflects the fact that the potential becomes truly five dimensional on small scales. For $r \gg \ell$,

$$V(r) \approx \frac{GM}{r} \left(1 + \frac{2\ell^2}{3r^2} \right) \quad [32]$$

which gives the small correction to four-dimensional gravity at low energies from extra-dimensional effects.

Cosmological Brane Worlds

The RS models contain vacuum (Minkowski) branes. In order to pursue brane-world ideas in cosmology, we need to generalize the RS models to incorporate cosmological branes with matter and radiation on them. The effective field equations on the brane are the vehicle for brane-bound observers to interpret cosmological dynamics. They arise from projecting the five-dimensional field equations onto the brane, via the Gauss–Codazzi equations. These equations involve also the extrinsic curvature $K_{\mu\nu}$ of the brane, which determines how the brane is imbedded in the bulk.

The stress-energy on the brane (tension, matter, radiation) means that there is a jump in $K_{\mu\nu}$ across

the brane. More precisely, the junction conditions across the brane are

$$g_{\mu\nu}^+ - g_{\mu\nu}^- = 0 \quad [33]$$

$$K_{\mu\nu}^+ - K_{\mu\nu}^- = -\kappa_{(5)}^2 \left[T_{\mu\nu}^{\text{brane}} - \frac{1}{3} T^{\text{brane}} g_{\mu\nu} \right] \quad [34]$$

where

$$T_{\mu\nu}^{\text{brane}} = T_{\mu\nu} - \lambda g_{\mu\nu} \quad [35]$$

is the total energy–momentum tensor on the brane and $T^{\text{brane}} = g^{\mu\nu} T_{\mu\nu}^{\text{brane}}$. The Z_2 -symmetry means that when approaching the brane from one side and going through it, one emerges into a bulk that looks the same, but with the normal reversed. This implies that

$$K_{\mu\nu}^- = -K_{\mu\nu}^+ \quad [36]$$

so that we can use the junction condition (eqn [34]) to determine the extrinsic curvature:

$$K_{\mu\nu} = -\frac{1}{2} \kappa_{(5)}^2 \left[T_{\mu\nu} + \frac{1}{3} (\lambda - T) g_{\mu\nu} \right] \quad [37]$$

where $T = T^\mu{}_\mu$, we have dropped the (+) and we evaluate quantities on the brane by taking the limit $y \rightarrow +0$.

Together with the Gauss–Codazzi equations, eqn [37] leads to the induced field equations on the brane:

$$G_{\mu\nu} = -\Lambda g_{\mu\nu} + \kappa^2 T_{\mu\nu} + 6 \frac{\kappa^2}{z\lambda} \mathcal{S}_{\mu\nu} - \mathcal{E}_{\mu\nu} \quad [38]$$

where

$$\kappa^2 \equiv \kappa_{(4)}^2 = \frac{1}{6} \lambda \kappa_{(5)}^4 \quad [39]$$

$$\Lambda \equiv \Lambda_{(4)} = \frac{1}{2} [\Lambda_{(5)} + \kappa^2 \lambda] \quad [40]$$

$$\begin{aligned} \mathcal{S}_{\mu\nu} = & \frac{1}{12} T T_{\mu\nu} - \frac{1}{4} T_{\mu\alpha} T^{\alpha\nu} \\ & + \frac{1}{24} g_{\mu\nu} [3 T_{\alpha\beta} T^{\alpha\beta} - T^2] \end{aligned} \quad [41]$$

and

$$\mathcal{E}_{\mu\nu} = {}^{(5)}C_{ACBD} n^C n^D g_\mu{}^A g_\nu{}^B \quad [42]$$

where n^A is the unit normal to the brane and ${}^{(5)}C_{ACBD}$ is the Weyl tensor in the bulk.

The induced field equations [38] show two key modifications to the standard four-dimensional Einstein field equations arising from extra-dimensional effects.

- $\mathcal{S}_{\mu\nu} \sim (T_{\mu\nu})^2$ is the high-energy correction term, which is negligible for $\rho \ll \lambda$, but dominant for $\rho \gg \lambda$ (where ρ is the energy density):

$$\frac{|\kappa^2 \mathcal{S}_{\mu\nu} / \lambda|}{|\kappa^2 T_{\mu\nu}|} \sim \frac{|T_{\mu\nu}|}{\lambda} \sim \frac{\rho}{\lambda} \quad [43]$$

- $\mathcal{E}_{\mu\nu}$, the projection of the bulk Weyl tensor on the brane, encodes corrections from KK or five-dimensional graviton effects. From the brane-observer viewpoint, the energy–momentum corrections in $\mathcal{S}_{\mu\nu}$ are local, whereas the KK corrections in $\mathcal{E}_{\mu\nu}$ are nonlocal, since they incorporate five-dimensional gravity wave modes. These nonlocal corrections cannot be determined purely from data on the brane. In the perturbative analysis of RS2 which leads to the corrections in the gravitational potential, eqn [32], the KK modes that generate this correction are responsible for a nonzero $\mathcal{E}_{\mu\nu}$; this term is what carries the modification to the weak-field field equations.

The effective field equations are not a closed system. One needs to supplement them by five-dimensional equations governing $\mathcal{E}_{\mu\nu}$, which are obtained from the five-dimensional Einstein equations.

Cosmological Dynamics

A (1+4)-dimensional spacetime with spatial 4-isotropy (four-dimensional spherical/ plane/ hyperbolic symmetry) has a natural splitting into hypersurfaces of symmetry, which are (1+3)-dimensional surfaces with 3-isotropy and 3-homogeneity, that is, Friedmann–Robertson–Walker (FRW) surfaces. In particular, the AdS₅ bulk of the RS2 brane world, which admits a foliation into Minkowski surfaces, also admits an FRW foliation since it is 4-isotropic. The generalization of AdS₅ that preserves 4-isotropy and solves the five-dimensional Einstein equation is Schwarzschild AdS₅, and this bulk therefore admits an FRW foliation. It follows that an FRW cosmological brane world can be embedded in Schwarzschild AdS₅ spacetime.

The black hole in the bulk is felt on the brane via the $\mathcal{E}_{\mu\nu}$ term. The bulk black hole gives rise to “dark radiation” on the brane via its Coulomb effect. The FRW brane can be thought of as moving radially along the fifth dimension, with the junction conditions determining the velocity via the Friedmann equation. Thus, one can interpret the expansion of the universe as motion of the brane through the static bulk. In the special case of no black hole and no brane motion, the brane is empty and has Minkowski geometry, that is, the original RS2 brane world is recovered, in different coordinates.

An intriguing aspect of the cosmological metric is that five-dimensional gravitational wave signals can take “shortcuts” through the bulk in traveling

between points A and B on the brane. The travel time for such a graviton signal is less than the time taken for a photon signal (which is stuck to the brane) from A to B.

Cosmological dynamics on the brane are governed by the modified Friedmann equation:

$$H^2 = \frac{\kappa^2}{3}\rho\left(1 + \frac{\rho}{2\lambda}\right) + \frac{m}{a^4} + \frac{1}{3}\Lambda - \frac{K}{a^2} \quad [44]$$

where $H = \dot{a}/a$ is the Hubble expansion rate, $a(t)$ is the scale factor, K is the curvature index, and m is the mass of the bulk black hole.

The ρ^2/λ term is the high-energy term. When $\rho \gg \lambda$, in the early universe, then $H^2 \propto \rho^2$. This means that a given energy density produces a greater rate of expansion that it would in standard four-dimensional gravity. As a consequence, inflation in the early universe is modified in interesting ways, some of which may leave a signature in cosmological observations.

The m/a^4 term in eqn [44] is the “dark radiation,” so called because it redshifts with expansion like ordinary radiation. But, unlike ordinary radiation, it is not a form of detectable matter, but the imprint on the brane of the gravitational field in the bulk (the Coulomb effect of the bulk black hole). This additional effective relativistic degree of freedom is constrained by nucleosynthesis in the early universe. Any extra radiative energy not thermally coupled to radiation affects the rate of production of light elements, and observed abundances place tight constraints on such extra energy. The dark radiation can be no more than $\sim 3\%$ of the radiation energy density at nucleosynthesis:

$$\frac{3m}{\kappa^2 \rho_{\text{nuc}}} \lesssim 0.03 \quad [45]$$

The other modification to the Hubble rate is via the high-energy correction ρ/λ . In order to recover the observational successes of general relativity, the high-energy regime where significant deviations occur must take place before nucleosynthesis, that is, cosmological observations impose the lower limit

$$\lambda > (1 \text{ MeV})^4 \Rightarrow M_{(5)} > 10^4 \text{ GeV} \quad [46]$$

This is much weaker than the limit imposed by table-top experiments, which limit the curvature radius to $\ell \lesssim 0.2 \text{ mm}$, leading to

$$\lambda > (100 \text{ GeV})^4 \Rightarrow M_{(5)} > 10^8 \text{ GeV} \quad [47]$$

The high-energy regime during radiation domination is short-lived. Since ρ^2/λ decays as a^{-8} during the radiation era, it will rapidly drop below one, and the universe will enter the low-energy four-dimensional regime. However, traces of the high-energy era may be left in the perturbation spectra that leave an imprint in the cosmic microwave background radiation.

In conclusion, simple brane-world models of RS2 type provide a rich phenomenology for exploring some of the ideas that are emerging from M theory. The higher-dimensional degrees of freedom for the gravitational field, and the confinement of standard model fields to the visible brane, lead to a complex but fascinating interplay between gravity, particle physics, and geometry, which enlarges and enriches general relativity in the direction of a quantum gravity theory. High-precision astronomical data mean that cosmology is a potential laboratory for testing and constraining these brane worlds. The models predict extra-dimensional signatures in the cosmic microwave background and other observations, and these predictions can in principle be tested against data.

See also: String Theory: Phenomenology; Supergravity; Superstring Theories.

Further Reading

- Brax P and van de Bruck C (2003) Cosmology and brane worlds: a review. *Classical and Quantum Gravity* 20: R201 (arXiv: hep-th/0303095) (arXiv: hep-th/0303095).
- Cavaglia M (2003) Black hole and brane production in TeV gravity: a review. *International Journal of Modern Physics A* 18: 1843 (arXiv:hep-ph/0210296).
- Langlois D (2003) Cosmology in a brane-universe. *Astrophysics and Space Science* 283: 469 (arXiv:astro-ph/0301022).
- Maartens R (2004) Brane-world gravity. *Living Reviews in Relativity* 7: 7 (arXiv:gr-qc/0312059).
- Quevedo F (2002) Lectures on string/brane cosmology. *Classical and Quantum Gravity* 19: 5721 (arXiv:hep-th/0210292).
- Rubakov V (2001) Large and infinite extra dimensions. *Physics-Uspekhi* 44: 871 (arXiv:hep-ph/0104152).
- Wands D (2002) String-inspired cosmology. *Classical and Quantum Gravity* 19: 3403 (arXiv:hep-th/0203107).

Branes and Black Hole Statistical Mechanics

S R Das, University of Kentucky, Lexington, KY, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

In classical general relativity, a black hole is a solution of Einstein's equations with a region of spacetime which is causally disconnected from the asymptotic region at infinity. The boundary of such a region is called the "event horizon." The spacetime around the simplest black hole in three space dimensions is described by the Schwarzschild metric

$$ds^2 = -\left(1 - \frac{2GM}{rc^2}\right) dt^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 + r^2 d\Omega^2 \quad [1]$$

where G is Newton's gravitational constant, c is the velocity of light, and we have used spherical coordinates with $d\Omega$ the line element on an S^2 . A nonrotating, uncharged star which is too massive to form a neutron star will eventually collapse, and at late times the metric will be given by [1]. The horizon is a null surface $S^2 \times t$ and the radius of the S^2 is $r_{\text{horizon}} = 2GM/c^2$. The Schwarzschild solution has generalizations to black holes with charge and angular momentum and no-hair theorems guarantee that a black hole has no other characteristic property. All these solutions can be generalized to other theories like supergravity in various dimensions.

In 1974, Hawking showed that due to pair production of particles near the horizon, black holes radiate thermally. Hawking's calculation is valid for black holes whose masses are much larger than the Planck mass: for such black holes, the curvature at the horizon is weak and normal semiclassical quantization is valid. Remarkably, the properties of Hawking radiation are quite universal. A black hole can be characterized by an entropy called the Bekenstein–Hawking entropy. The leading result for the entropy S_{BH} for all black holes in any theory with the standard Einstein–Hilbert action is given by

$$S_{\text{BH}} = \frac{A_{\text{H}}}{4G} \quad [2]$$

where A_{H} denotes the area of the horizon. The temperature T_{H} is given by

$$T_{\text{H}} = \frac{\kappa}{2\pi} \quad [3]$$

where κ is the surface gravity at the horizon. The principle of detailed balance further ensures that the radiation rate of some species of particle i , $\Gamma_i(k)$, in some given momentum range $(k, k + dk)$ is related to the corresponding absorption cross section $\sigma_i(k)$ by

$$\Gamma(k) = \frac{\sigma_i(k)}{e^{\omega/T_{\text{H}}} \pm 1} \frac{d^d k}{(2\pi)^d} \quad [4]$$

where ω is the energy and d denotes the number of spatial dimensions. The \pm sign refers to fermions (bosons), respectively. A nontrivial k dependence of σ_i signifies a departure from black-body behavior. Consequently, $\sigma_i(k)$ is often called a grey-body factor. Equations [2] and [3] may be derived by combining Hawking's calculation of the radiation with standard thermodynamic relations. Alternatively, they follow from the leading semiclassical approximations of path-integral formulations of Euclidean gravity based on the standard Einstein–Hilbert action. For an account of black-hole thermodynamics, see Wald (1994).

Unlike usual thermodynamic systems, black holes appear to pose a deep puzzle. In usual systems, thermodynamics is a coarse-grained description of a system which is in a highly degenerate state. Typically, such systems are described in terms of a few macroscopic parameters such as the total energy, the total volume, the total charge. For each set of values of these macroscopic parameters, there are a large number of microscopic states which can be described in terms of the constituents such as atoms or molecules. This degeneracy manifests itself as an entropy S which is related to the number of microscopic states for a given set of values of the macroscopic parameters, Ω by Boltzmann's relation

$$S = \log(\Omega) \quad [5]$$

where units have been chosen such that the Boltzmann constant is unity. For a black hole, the macrostates are specified by its mass, charge, and angular momentum. No-hair theorems, however, seem to suggest that there are no other properties and hence no obvious candidate for microstates. In the absence of such a statistical basis, one would be inevitably led to the conclusion that there is loss of information in processes involving black holes.

In a consistent quantum theory of gravity, there would be such a statistical basis since quantum mechanics is unitary. String theory is a strong candidate for a unified theory which contains gravity. Indeed, string theory provides a microscopic description for a class of black holes.

Black Hole Solutions in String Theory

Perturbatively, the basic excitations of string theory are fundamental closed and open strings characterized by a string tension T_s and hence a length scale, the string length $l_s = 1/\sqrt{2\pi T_s}$. Consistency requires that the string should be able to propagate in ten spacetime dimensions and should be supersymmetric at the fundamental level. Formulated in this fashion, there are several consistent string theories: type IIA, type IIB, and heterotic string theory (which contain only closed strings perturbatively) and type I theory (which contains both open and closed strings).

At energies much smaller than $1/l_s$, only the massless modes of the string can be excited. For all these string theories, the massless spectrum of closed strings contains the graviton and the low-energy dynamics is given by the appropriate supersymmetric generalization of general relativity, supergravity. In addition, the closed-string spectrum contains a neutral scalar field, the dilaton ϕ , whose expectation value gives rise to a dimensionless parameter governing interactions, called the string coupling g_s :

$$g_s = e^{\langle\phi\rangle} \quad [6]$$

The ten-dimensional gravitational constant is given by

$$G_{10} = 8\pi^6 g_s^2 l_s^8 \quad [7]$$

Ten-dimensional supergravity has a wide variety of black hole solutions, the simplest of which is the straightforward generalization of the Schwarzschild solution.

Black p -Brane Solutions

More significantly, there are solutions which are charged with respect to the various gauge fields that appear in the supergravity spectrum. Generically, these charged solutions represent extended objects. For accounts of such solutions, see [Maldacena \(1996\)](#).

Consider, for example, the supergravity which follows from type IIB string theory. This theory has a pair of 2-form gauge fields B_{MN} and B'_{MN} and a 4-form gauge field A_{MNPQ} with a self-dual field strength. Just as an ordinary point electric charge produces a 1-form gauge field, a $(p+1)$ -form gauge field may be sourced by an electrically charged p -dimensional extended object. The corresponding field strength is a $(p+2)$ -form, whose Hodge dual in d spacetime dimensions is a $(d-p-2)$ form. This shows that there should be magnetically charged

$(d-p-4)$ -dimensional extended objects as well. These extended objects are called “branes.”

In the type IIB example, there should be two kinds of one-dimensional extended objects which carry electric charge under B_{MN}, B'_{MN} , called the F-string and the D-string, respectively. There are also two kinds of five-dimensional branes which carry magnetic charges under B_{MN}, B'_{MN} , called the NS 5-brane and D5 brane, respectively. Finally, there should be a 3-brane, since the corresponding 5-form field strength is self-dual as well as a D7 brane. A similar catalog can be prepared for other string theories, as well as for 11-dimensional supergravity, which is the low-energy limit of M-theory.

The classical solutions for a set of p -branes of the same kind generally have inner and outer horizons which have the topology $t \times S^{8-p} \times R^p$. The outer horizon is then associated with a Hawking temperature and a Bekenstein–Hawking entropy. Of particular interest are extremal limits. In this limit, the inner and outer horizons coincide and the mass density is simply proportional to the charge. Given some charge, the extremal solution has the lowest energy. Extremal limits are interesting because in supergravity these correspond to solutions in which some of the supersymmetries (in this case, half of the supersymmetries) are retained – such solutions are called Bogomolny–Prasad–Sommerfeld (BPS) saturated solutions. The charge in question appears as a central charge in the extended supersymmetry algebra. This fact may be used to show that such BPS solutions are absolutely stable. Indeed, for the particular solution considered here, the Hawking temperature $T_H \rightarrow 0$, so that there is no Hawking radiation, as required by stability. Furthermore, the entropy $S_{BH} \rightarrow 0$. The horizon shrinks to a point which appears as a naked null singularity.

All the ten dimensions of string theory need not be noncompact. In fact, to describe the real world, one must have a solution of string theory in which six of the dimensions are wrapped up and form a compact space. In principle, however, one can compactify any number of dimensions. In the above example of a p -brane, it is trivial to compactify the directions along which the brane is extended to a p -dimensional torus, T^p , which can be chosen to be a product of p circles each of radius R . At length scales much smaller than R , the theory then becomes a $(10-p)$ -dimensional theory. The p -brane appears as a black hole with a spherical horizon and, since the original p -form gauge field now behaves as an ordinary 1-form gauge field with a nonzero time component, this is an electrically charged black hole.

D1–D5– N System and Five-Dimensional Black Holes

For reasons which will become clear in the next section, it is useful to get extremal black holes with large horizon areas, so that Hawking's semiclassical formulas are valid. It turns out that such solutions involve branes of various types which intersect each other and are suitably wrapped on compact internal spaces. Such black holes then have necessarily different kinds of charges. It turns out that the simplest case is a five-dimensional black hole with three kinds of charges, which is obtained by brane systems wrapped on a compact five-dimensional space. An example is a type IIB solution which has D5 branes which are wrapped on either $T^4 \times S^1$ or $K3 \times S^1$, together with D1 branes wrapped on the S^1 as well as some momentum along the S^1 . From the noncompact five-dimensional point of view, this is a black hole with three kinds of gauge charges: the D5 charge Q_5 , the D1 charge Q_1 , and a Kaluza–Klein charge N coming from the momentum $P = N/R$ along the circle of radius R .

When the internal space is $T^4 \times S^1$ the five-dimensional Einstein frame metric is given by

$$ds^2 = -[f(r)]^{-2/3} \left(1 - \frac{r_0^2}{r^2} \right) dt^2 + [f(r)]^{1/3} \left[\frac{dr^2}{(1 - r_0^2/r^2)} + r^2 d\Omega_3^2 \right] \quad [8]$$

where

$$f(r) = \left(1 + \frac{r_0^2 \sinh^2 \alpha_1}{r^2} \right) \left(1 + \frac{r_0^2 \sinh^2 \alpha_5}{r^2} \right) \times \left(1 + \frac{r_0^2 \sinh^2 \sigma}{r^2} \right) \quad [9]$$

and the three charges are

$$Q_1 = \frac{V r_0^2 \sinh 2\alpha_1}{32\pi^4 g_s l_s^6}, \quad Q_5 = \frac{r_0^2 \sinh 2\alpha_5}{2g_s l_s^2} \quad [10]$$

$$N = \frac{VR^2}{32\pi^4 l_s^8 g_s^2} r_0^2 \sinh 2\sigma$$

where V is the volume of the T^4 and R is the radius of the circle S^1 .

The ADM mass of the black hole is

$$M_{\text{ADM}} = \frac{RVr_0^2}{32\pi^4 g_s^2 l_s^8} \times [\cosh 2\alpha_1 + \cosh 2\alpha_5 + \cosh 2\sigma] \quad [11]$$

The Bekenstein–Hawking entropy is given by

$$S_{\text{BH}} = \frac{RVr_0^3}{8\pi^3 l_s^8 g_s^2} \cosh \alpha_1 \cosh \alpha_5 \cosh \sigma \quad [12]$$

while the Hawking temperature is

$$T_{\text{H}} = \frac{1}{2\pi r_0 \cosh \alpha_1 \cosh \alpha_5 \cosh \sigma} \quad [13]$$

The extremal limit of this solution is given by

$$r_0 \rightarrow 0, \quad \alpha_1, \alpha_5, \sigma \rightarrow \infty \quad [14]$$

$$Q_1, Q_5, N = \text{fixed}$$

The extremal solution is a BPS saturated state and retains four of the original supersymmetries. In this limit, the inner and outer horizons coincide. However, the horizon is now a smooth S^3 with a finite area in the Einstein frame metric. Consequently, the extremal Bekenstein–Hawking entropy is also finite and may be seen to be

$$S_{\text{BH}}^{\text{3-charge extremal}} = 2\pi \sqrt{Q_1 Q_5 N} \quad [15]$$

The temperature, however, is zero in this limit, which is consistent with the stability of a BPS saturated state.

The above five-dimensional black hole is in fact a generalization of the Reissner–Nordstrom black hole. Similar solutions with large horizon areas in the extremal limit can be constructed in four dimensions. One such construction is in the IIB theory wrapped on T^6 in which there are four sets of D3 branes which wrap four different T^3 's contained in the T^6 . Black holes with lower supersymmetry may be obtained by replacing the T^6 by a Calabi–Yau space.

Duality and Branes

String theory has a rich set of symmetries called duality symmetries which relate different kinds of string theories that are suitably compactified. These symmetries relate different classical solutions. For example, application of these symmetries relate the five-dimensional black holes above with other five-dimensional black holes with different kinds of charges. Furthermore, at the level of supergravity, these various theories may be derived from a yet unknown 11-dimensional theory called the M-theory whose low-energy limit is 11-dimensional supergravity.

Branes in String Theory

For a given string theory, the perturbative spectrum consists of strings. However, at the nonperturbative

level, there are, in addition, extended objects of other dimensionalities. Duality symmetries imply that these extended objects are as “fundamental” as the strings themselves. Such extended objects are also called branes. For an exhaustive account of branes in string theory, see [Johnson \(2003\)](#).

Like their counterparts in supergravity, branes in string theory are typically charged with respect to some gauge fields. While supergravity solutions are possible with any value of the charge, in string theory the brane charges have to be quantized. Multiple units of the minimum quantum of charge can appear as collections of branes each with unit charge or, alternatively, branes which wrap around compact cycles in space a multiple number of times.

D-Branes

The extended objects in string theory are described in terms of their collective excitations. These are best understood for the class of branes called D-branes in the type II theory, discovered by Polchinski. These are D1, D3, D5, and D7 branes in type IIB and D0, D2, D4, and D6 branes in type IIA theory. Dp branes are characterized by the fact that they couple to, and act as sources for, $(p+1)$ -form gauge fields which belong to the Ramond–Ramond sector of the theory. Collective excitations of a p -dimensional extended object in field theory are expected to be described by waves on its $(p+1)$ -dimensional world volume. The collective coordinate action would be a quantum field theory which has vectors, corresponding to longitudinal oscillations of the brane, and scalars which correspond to transverse oscillations. For D-branes in string theory, the theory of collective excitations is a string field theory of open strings whose endpoints lie on the brane. (This is the origin of the nomenclature D-brane: an open string whose ends are constrained to lie on the brane has a world-sheet description in which the bosonic fields corresponding to transverse target space coordinates have Dirichlet boundary conditions.) The lowest-energy states of open superstrings are ordinary massless gauge fields and their supersymmetric partners so that the low-energy limit of the string field theory is a supersymmetric gauge theory.

The fact that the underlying theory is a string theory has an important consequence. For a system of N parallel D-branes of the same type, one would have open strings which join different branes as well as the same brane. The low-energy theory then becomes a supersymmetric nonabelian gauge theory with gauge group $U(N)$. In a suitable

gauge, the off-diagonal gauge fields and their supersymmetric partners (which include scalar fields in the adjoint representation) are the low-energy degrees of freedom of open strings which connect different branes.

The mass density or tension T_p of a single Dp brane is given by

$$T_p = \frac{1}{g_s (2\pi)^p l_s^{p+1}} \quad [16]$$

This couples to the $(p+1)$ -form gauge field with a charge

$$\mu_p = g_s T_p \quad [17]$$

and the Yang–Mills coupling constant for the collective theory on the brane world volume is given by

$$g_{\text{YM-D}p}^2 = (2\pi)^{p-2} g_s l_s^{p-3} \quad [18]$$

The ground state of a single Dp brane is a BPS state which preserves 16 of the 32 supersymmetries of the original theory. One consequence of this is that two or more parallel Dp branes of the same type form a threshold bound state preserving the same supersymmetries, with no net force between them. As a result, the tension of N parallel Dp branes is simply NT_p .

Branes of different dimensionalities can also form bound states. Of particular interest are configurations which can form threshold bound states which preserve some supersymmetries. For example, a set of N_1 parallel Dp branes can form a threshold bound state with a set of N_2 parallel $D(4+p)$ branes with all the p branes lying entirely along the $(4+p)$ -branes. This configuration is also a BPS saturated state preserving eight of the original supersymmetries and would have charges under both $(p+1)$ -form and $(p+5)$ -form gauge potentials. The BPS nature ensures that the total mass density is the sum of the individual mass densities.

NS Branes

The other extended objects in string theory are called NS branes since they couple to p -form gauge fields which arise from the Neveu–Schwarz/Neveu–Schwarz sector of the world-sheet theory. These are present in all the five string theories and appear in two types. The first is a macroscopic fundamental string which may be wound around a compact direction. The second is called a solitonic 5-brane. While the collective dynamics of a fundamental string is the standard world-sheet description of string theory, the description for the NS 5-brane is rather complicated and not known in full detail. The rest of this article deals exclusively with D-branes.

D-Branes and Black Branes

The idea that black holes correspond to highly degenerate states in string theory is quite old and dates back to 't Hooft (1990) and Susskind (1993). In the following two sections we discuss such black holes which are described by D-branes. For reviews see Maldacena (1996), Das and Mathur (2001), and David *et al.* (2002).

We have so far discussed the string-theoretic branes in two different ways. In the first description, branes are solutions of the low-energy equations of motion – this is the setting in which branes provide conventional descriptions of black holes. In the second description, branes are certain states in the quantum theory of superstrings. More specifically, D-branes are described in terms of states of the open-string field theory which lives on the branes. The first description is necessarily approximate. On the other hand, the second description is exact in principle, although in practice one might not know how to write down and analyze the string-field theory in an exact fashion.

The description in terms of open-string field theory should reduce to the description in terms of a classical solution when the charges and masses become large. If black-hole thermodynamics has a microscopic origin, D-branes should be highly degenerate states in this limit and the entropy should be given by the Boltzmann formula. Furthermore, Hawking radiation should be understood as an ordinary decay process.

For a system of Q_p parallel Dp branes, the mass is Q_p/g_s , while Newton's gravitational constant $G \sim g_s^2$. Gravitational effects are controlled by $GM \sim g_s Q_p$. A semiclassical limit in closed-string theory requires $g_s \rightarrow 0$, while a nontrivial gravitational effect in this limit requires $g_s Q_p$ finite, which implies one must have $Q_p \gg 1$. Furthermore, when $g_s Q_p \gg 1$ the typical curvatures are small compared to the string scale and the semiclassical string theory reduces to classical supergravity. This is the limit in which branes are well described as classical solutions.

Similar considerations apply for brane systems with multiple charges. For example, in the D1–D5– N system the classical solution becomes a good description when all the quantities $g_s Q_1$, $g_s Q_5$, and $g_s^2 N$ become large. (The relevant quantity which comes with the momentum has g_s^2 rather than g_s because the mass contribution from the momentum is simply N/R without any inverse power of g_s .) However, g_s is the square of the coupling constant of the open-string theory living on the brane – in fact, eqn [18] shows this relation in the low-energy limit.

It is well known that in a $U(Q_p)$ gauge theory the real coupling constant is $g_{\text{YM}} \sqrt{Q_p} \sim \sqrt{g_s Q_p}$. This means that the semiclassical limit corresponds to a strongly coupled string-field theory which reduces to strongly coupled gauge theory in the low-energy limit and the picture of D-branes as a collection of open strings is not very useful. In fact, known calculational methods in gauge theory or open-string theory are not valid in this regime.

Microscopic Entropy for Two-Charge Systems

The prospects are much better for extremal black holes, which appear as BPS states in string theory. This is because the spectra of BPS states do not depend on the coupling. The degeneracy of such states may therefore be calculated at weak coupling, where techniques are well known and the result can be extrapolated to strong coupling without change.

The simplest BPS state is the ground state of a set of parallel D-branes of the same type. This state is indeed 128-fold degenerate, which would imply a microscopic entropy. This entropy, however, is small and therefore invisible in the corresponding classical solution. Indeed, the classical solution shows that in the extremal limit the horizon area is zero, leading to a vanishing Bekenstein–Hawking entropy.

The next interesting class of states consists of threshold bound states with two kinds of charges. Consider, for example, the D1–D5 system on $T^4 \times S^1$ considered above with no momentum along the D1's. By known duality transformations, this is equivalent to a fundamental IIB string which is wound Q_5 times around the S^1 and with a net momentum $P = Q_1/2\pi Q_5 R$ (where R is the radius of the S^1), with four of the transverse directions compactified on a T^4 . For this system, it is easy to count the number of states for given values of Q_1 and Q_5 at weak string coupling by simply enumerating the perturbative oscillator states of the string. For large values of Q_1 and Q_5 , we can alternatively calculate this entropy by using a canonical ensemble of eight massless bosons corresponding to the eight transverse polarizations and their supersymmetric partners – eight massless fermions – moving on the string with some temperature T and a chemical potential α for the total momentum.

Consider a noninteracting gas of f massless bosons and f fermions living on a circle with circumference L . The average number of left- and right-moving particles with some energy e , denoted by ρ_L, ρ_R , respectively, are

$$\rho_i(e) = \frac{1}{e^{e/T_i} \pm 1}, \quad i = L, R \quad [19]$$

where the \pm sign refers to fermions and bosons, respectively, and we have introduced left- and right-moving temperatures T_L, T_R . The physical temperature is

$$\frac{1}{T} = \frac{1}{2} \left(\frac{1}{T_L} + \frac{1}{T_R} \right) \quad [20]$$

The extensive quantities, such as the energy E , momentum P , and entropy S , then become the sum of left- and right-moving pieces:

$$E = E_L + E_R, \quad P = P_L + P_R, \quad S = S_L + S_R \quad [21]$$

and the distribution function [19] leads to the following thermodynamic relations:

$$T_i = \sqrt{\frac{3E_i}{L\pi f}} = \frac{4S_i}{\pi f L}, \quad i = L, R \quad [22]$$

Since the total momentum $P = P_R + P_L = E_R - E_L$ is nonzero, the lowest-energy state is clearly the one in which all the particles move in the same direction, for example, right moving. This is a BPS state and corresponds to the extremal solution in supergravity. Then $E = E_R = P = P_R$. This approach to the black hole entropy was initiated by Das and Mathur (1996) and Callan and Maldacena (1996).

For our two-charge system, $f = 8, P = 2\pi Q_1/L$, and $L = 2\pi R Q_1 Q_5$. Using [22] we get

$$S_{\text{micro}}^{2\text{-charge-II}} = 2\pi \sqrt{2Q_1 Q_5} \quad [23]$$

This is the microscopic entropy for the fundamental string with momentum in the type II theory. By duality, this is also the microscopic entropy of the D1–D5 system. This is a large number which should agree with the macroscopic entropy calculated from the corresponding classical solution.

The discussion is almost identical for the fundamental heterotic string, except that now we have 24 right-moving bosons, eight left-moving bosons, and eight left-moving fermions, and the BPS state consists only of right movers. If n_w denotes the winding number and n_p the quantized momentum the extremal heterotic string entropy is

$$S_{\text{micro}}^{2\text{-charge heterotic}} = 4\pi \sqrt{n_p n_w} \quad [24]$$

The supergravity solution for the D1–D5 system may be obtained by substituting $\sigma = 0$ in eqns [8]–[13]. In the extremal limit, the classical Bekenstein–Hawking entropy vanishes as is clear from the expression [15], in which $N = 0$. This appears to be in contradiction with the fact that the state has a large microscopic entropy.

The key point, however, is that the two-charge solution has a singular horizon where the string frame curvature is large. Consequently, low-energy tree-level supergravity breaks down near the horizon and higher-derivative terms (e.g., higher powers of curvature) become important. This issue has been best studied for the fundamental heterotic string compactified on T^6 . This is dual to the D1–D5 system in type IIB theory compactified on $K3 \times T^2$. The classical supergravity solution is then a singular black hole in four spacetime dimensions. In one of the first papers on the string-theoretic understanding of black hole thermodynamics, Sen (1995) showed that, for large n_p, n_w , string-loop effects are small near the horizon so that the only relevant corrections are higher-derivative terms coming from integrating out the massive modes of the string at tree level. Furthermore, a robust scaling argument shows that regardless of the detailed nature of the derivative corrections, the macroscopic entropy defined through the horizon area must be of the form $a\sqrt{n_p n_w}$, where a is a pure number. Finally, one can define a “stretched horizon” as the surface where the curvature becomes of the order of the string scale and the area of the stretched horizon is indeed proportional to $\sqrt{n_p n_w}$. This result gives a strong indication that string theory provides a microscopic basis for black hole thermodynamics, although the coefficient a cannot be determined without more detailed knowledge of higher-derivative terms.

Microscopic Entropy of Extremal Three-Charge System

Brane bound states with three kinds of charge provide examples of black holes whose extremal limits have large horizons with curvatures much smaller than the string scale. In this case, a microscopic count of states in string theory should exactly account for the Bekenstein–Hawking formula, without corrections coming from higher derivatives. This is indeed true, as first found by Strominger and Vafa (1996). In the following, we will outline how this calculation can be done in the D1–D5– N system on $K3 \times S^1$ or $T^4 \times S^1$ following the treatment of Dijkgraaf *et al.* (1996).

D1 branes can be considered as “instanton strings” in the six-dimensional supersymmetric $U(Q_5)$ gauge theory of D5 branes (actually, these should be called solitonic strings rather than instantons, since the configurations are time independent). The total instanton number is the D1-brane charge Q_1 . The moduli space of these instantons is then a blown-up version of the

orbifold $(T^4)^{Q_1 Q_5}/S(Q_1 Q_5)$ or $(K3)^{Q_1 Q_5}/S(Q_1 Q_5)$ and is $4Q_1 Q_5$ dimensional. Since any instanton configuration is independent of time x^0 and the S^1 direction x^5 , the collective coordinate dynamics is a $(1+1)$ -dimensional field theory which lives in the (x^0, x^5) space. At low energies, this flows to a conformal field theory with a central charge $c=6Q_1 Q_5$ since there are $4Q_1 Q_5$ bosons each contributing 1 to the central charge and an equal number of fermions each contributing $1/2$. The BPS state with momentum N/R is a purely right- or left-moving state in this conformal field theory which has a conformal weight N . From general principles of conformal invariance, the degeneracy of such states for large N is given by Cardy's formula

$$d(N) \sim e^{2\pi\sqrt{cN/6}} \quad [25]$$

so that the microscopic entropy is

$$S_{3\text{-charge}}^{\text{micro}} = \log d(n) = 2\pi\sqrt{cN/6} \quad [26]$$

Substituting the value of $c=6Q_1 Q_5$, this is in exact agreement with the Bekenstein–Hawking entropy of the classical solution given in [15].

Nonextremal Black Holes and Hawking Radiation

The BPS property of ground states of D-brane systems enables us to compute the degeneracy of microstates exactly in the regime of parameters where the state can be reliably described as a black hole solution in the low-energy theory. However, extremal black holes have vanishing temperature and do not radiate. To understand the microscopic origins of Hawking radiation, one has to go away from extremality. Such states are not supersymmetric and an extrapolation of weak-coupling calculations to strong coupling is not *a priori* justified. Nevertheless, it turns out that for small departures from extremality, weak-coupling results still reproduce semiclassical answers for entropy, temperature, and luminosity.

Near-Extremal Entropy

Nonextremal properties are best understood for the D1–D5– N system on $T^4 \times S^1$. In the orbifold limit, the conformal field theory which describes the low-energy dynamics is equivalent to a gas of strings which are wound around the S^1 and which can oscillate along the T^4 . The total winding number is $k=Q_1 Q_5$ and may be achieved by sets of strings which are multiply wound in various ways. As argued below, entropically the most favored configuration is a single long string wound around $Q_1 Q_5$

times. Thus, the thermodynamics may be analyzed exactly along the lines of the fundamental string in the previous section. The thermodynamic relations are given by [22] with $f=4$ and $L=2\pi R Q_1 Q_5$. The extremal state consists entirely of right movers and $E=E_R=N/R$. Substituting these values in [22] yields the correct formula for the microscopic entropy

$$S_{\text{micro}}^{3\text{-charge}} = 2\pi\sqrt{Q_1 Q_5 N} \quad [27]$$

The same expression follows if $f=4Q_1 Q_5$ and $L=2\pi R$ corresponding to $Q_1 Q_5$ singly wound strings. However, for statistical methods to hold, the entropy must be much larger than the number of flavors. The ratio of the entropy to the number of flavors is $S/f \sim \sqrt{N/Q_1 Q_5}$ for multiple singly wound strings and is not guaranteed to be large when all of Q_1, Q_5, N are large. On the other hand, this ratio is $S/f \sim \sqrt{Q_1 Q_5 N}$ for the long string. This shows that the long string is always entropically favored.

A departure from the extremal state is achieved by adding a left-moving momentum $2\pi n/L$ as well as a right-moving momentum $2\pi n/L$ to the extremal state, thus adding energy to the system but maintaining the total momentum. For the long string, this yields

$$S_R = 2\pi\sqrt{Q_1 Q_5 N + n}, \quad S_L = 2\pi\sqrt{n} \quad [28]$$

For small departures from extremality, $n \ll N$, the expressions for the total entropy and temperature as a function of the excess energy $\Delta E = 2n/Q_1 Q_5$ agree exactly with the near-extremal Bekenstein–Hawking entropy and the Hawking temperature of the classical solution, as shown by Callan and Maldacena (1996) and by Horowitz and Strominger.

The necessity of the long string appears in another important physical consideration. For statistical mechanics to be valid, the specific heat of the system has to be larger than unity. This implies that for the case considered here the energy gap ΔE must be larger than $1/RQ_1 Q_5$, which is precisely what the long string yields.

Hawking Radiation

A nonextremal state described above is unstable, since a left mover can annihilate a right mover into a closed-string mode which may leave the brane system and propagate to the asymptotic region. The resulting closed-string state will be in a thermal state whose temperature is the physical temperature of the initial state. This process is the microscopic

description of Hawking radiation. The decay rate is related to the absorption cross section of the corresponding mode by the principle of detailed balance, encoded in eqn [4].

From the point of view of the classical solution, the absorption cross section can be calculated by solving the linearized wave equation in the background geometry and calculating the ratio of the incident and reflected waves. It follows from these calculations that at low energies, absorption (and hence emission) are dominated by massless minimally coupled scalars. In fact, for any spherically symmetric black hole in any number of dimensions, there is a general theorem which ensures that the low-energy limit of this absorption cross section is exactly equal to the horizon area.

In the microscopic model for the three-charge black hole, this absorption cross section may be calculated by the usual rules of quantum mechanics. In the long-string limit and in the approximation that the modes on the long string form a dilute gas, the result has been derived by Das and Mathur (1996):

$$\sigma(\omega) = \frac{2\pi G_{10} Q_1 Q_5}{V} \omega \frac{e^{\omega/T} - 1}{(e^{\omega/2T_R} - 1)(e^{\omega/2T_L} - 1)} \quad [29]$$

where V is the volume of the T^4 and T is the physical temperature given by [20]. For a near-extremal hole $T_R \gg T_L$, so that $T \sim 2T_L$. Then in the extreme low-energy limit $\omega \ll T_R$, so that the corresponding Bose factor may be approximated as $1/(e^{\omega/2T_R} - 1) \sim 2T_R/\omega$. The cross section [29] becomes

$$\begin{aligned} \sigma &= \frac{4\pi Q_1 Q_5 G_{10} T_R}{V} = \frac{4G_{10} S_R}{(2\pi R)V} \\ &= 4G_5 S_{\text{extremal}} = A_H \end{aligned} \quad [30]$$

where G_5 is the five-dimensional Newton’s gravitational constant. We have used the relation [22] with $L = 2\pi R Q_1 Q_5$ and $f = 4$. The fact that in the near-extremal limit S_R is simply the extremal entropy and the fact that the extremal entropy reproduces the Bekenstein–Hawking formula has been used as well. Thus, the microscopic cross section exactly reproduces the semiclassical result at low energies. Even more remarkably, the full cross section [29] agrees with the semiclassical answer for the gray-body factor for parameters which correspond to the dilute-gas regime, as shown by Maldacena and Strominger.

It is rather surprising that the results for microscopic absorption cross section calculated at weak coupling agree with the semiclassical answers, since the relevant process involves states which are not

supersymmetric and therefore a naive extrapolation to strong coupling is not *a priori* justified. There are strong indications, however, that low-energy nonrenormalization theorems are at work. This agreement has been established not only for black holes with finite-horizon areas, but also for other systems with no horizons – most significantly, a set of parallel 3-branes – and forms the basis for Maldacena’s conjecture about AdS/CFT Correspondence (see AdS/CFT Correspondence).

Effects of Higher-Derivative Terms

The classical low-energy limit of string theory is supergravity. The effects of the massive modes of the string as well as effect of string loops is to add terms to the supergravity action which involve higher number of spacetime derivatives, for example, terms containing higher powers of the curvature. In the presence of such terms, the Bekenstein–Hawking formula for black hole entropy [2] receives corrections which can be calculated in a systematic fashion. It turns out that for a class of extremal black holes, this corrected entropy as computed in the modified supergravity is also in exact agreement with a microscopic calculation.

One example of this agreement is provided by four-dimensional extremal black holes in type IIA string theory compactified on a Calabi–Yau manifold. These are obtained by wrapping D4 branes on three different 4-cycles on the Calabi–Yau and having in addition a number of D0 branes. Let p^A , $A = 1, \dots, 3$ denote the three D4 charges and q_0 denote the D0 charge. The microscopic entropy of the BPS state can be computed by embedding this in M-theory:

$$\begin{aligned} S_{\text{micro}}^{\text{CY-Black hole}} &= 2\pi \sqrt{\frac{1}{6} |q_0| (C_{ABC} p^A p^B p^C + c_{2A} p^A)} \quad [31] \end{aligned}$$

where C_{ABC} is the intersection number of the 4-cycles and c_2 denotes the second Chern class of the Calabi–Yau space. When all the charges p^A are large, the term involving c_2 is subdominant. In this case, the result agrees with the Bekenstein–Hawking entropy of the corresponding classical solution. When the charges are not all large (so that the second term is appreciable), the curvatures of the supergravity solution become large at the horizon and higher-derivative corrections to the action cannot be ignored. In this particular case, it turns out that these higher-derivative corrections are string-loop corrections and can be computed using general properties of $N = 2$ supersymmetry, so that one can compute corrections to near-horizon geometry. Furthermore, one has to now modify the

expression for macroscopic entropy using the formalism of Wald. Putting these together, it is found that the macroscopic entropy following from the modified supergravity is in exact agreement with [31]. This subject is reviewed in [Mohaupt \(2000\)](#).

These methods have also been applied to the problem of two-charge black holes in heterotic string theory on $T6$ or, equivalently, type IIA on $K3 \times T^2$ ([Dabholkar 2004](#)). Recall that in this case the horizon of the usual supergravity solution is singular. It has been found that leading-order higher-derivative corrections smoothen out the horizon into a $AdS_2 \times S^2$ spacetime and the modified expression for the macroscopic entropy is again in exact agreement with the microscopic answer [23].

Geometry of Microstates

A satisfactory solution of the information-loss paradox requires a much more detailed understanding of black holes in string theory. The discussion above shows that black holes have microstates which may be described well in the weak-coupling regime. It is interesting to ask whether there is a description of these microstates in the strong-coupling regime in terms of the effective geometry perceived by suitable probes. This question has been answered for the two-charge system in great detail (see [Mathur \(2004\)](#)). It turns out that the D1–D5 microstates can be described by perfectly smooth metrics with no horizons, and they asymptote to the standard two-charge metric discussed above. The location of the erstwhile stretched horizon marks the point where the different microstates start differing from each other significantly. Since each such geometry does not have a horizon, neither does it have any entropy – this is consistent with their identification with nondegenerate microstates. Indeed, the number of such microstates correctly accounts for the microscopic entropy. Whether a similar picture holds for the three-charge system remains to be seen in detail, although there are some indications that this may be true. In this approach, it is not yet fully understood how a horizon emerges and why the entropy scales as the horizon area.

Outlook

One key feature of the understanding of black hole statistical mechanics from the dynamics of branes is the fact that a problem in gravity is mapped to a problem in a theory without gravity, for example, open-string field theory. In fact, the closed strings in the bulk are already contained in the spectrum of the

open strings. This is a consequence of the basic duality between open strings and closed strings. Furthermore, the open-string theory lives in a lower-dimensional spacetime. This is a manifestation of the holographic principle. As argued by Maldacena, the presence of a horizon implies that the low-energy limit retains all the modes of the closed strings near the horizon, while it truncates the open-string theory to a gauge theory. Open–closed duality then reduces to gauge–string duality. This provides a strong evidence that black holes obey the normal laws of quantum mechanics and hence their time evolution is unitary.

One of the most outstanding problems in the subject is a proper understanding of neutral black holes. Most of the quantitative results described above depend on supersymmetry, which allows extrapolation of weak-coupling answers to the strong-coupling domain. Some of these results can be extended to situations which have small departures from supersymmetry, for example, near-extremal black holes. States corresponding to neutral black holes are, however, far from supersymmetry and known calculational techniques fail. There are good reasons to expect, however, that the general philosophy – in particular the holographic principle – is still valid. Finally, so far string theory has been able to attack problems of eternal black holes. A satisfactory understanding of the information-loss problem requires an understanding of the dynamics of black hole formation and subsequent evaporation. Unfortunately, very little is known about this at the moment.

See also: AdS/CFT Correspondence; Black Hole Mechanics; Supergravity; Superstring Theories.

Glossary

- ADM (Arnowitt–Deser–Misner) mass** – Mass of a gravitational background which is asymptotically flat.
- AdS_n (anti-de Sitter space)** – A space (or spacetime) with constant negative curvature in n dimensions.
- BPS state (Bogomolny–Prasad–Sommerfeld state)** – In a theory of extended supersymmetry, a state that is invariant under a nontrivial subalgebra of the full supersymmetry algebra. These states always carry conserved charges, and supersymmetry determines the mass exactly in terms of the charges.
- Calabi–Yau space** – Complex Kahler manifold with vanishing first Chern class.
- Compactify (n. compactification)** – To consider a field or string theory in a spacetime some of whose spatial dimensions are compact.
- Dirichlet boundary condition** – The boundary condition which fixes the value of a field on the boundary.

Duality Equivalence of systems which appear to be distinct. For string theories, such equivalences relate string theories on different spacetimes as well as theories with different coupling constants.

Einstein–Hilbert action – The standard action for gravity which leads to Einstein’s equation, $S = (1/16\pi G) \int d^d x \sqrt{g} R$, where R is the Ricci scalar, g denotes the determinant of the metric, and G is Newton’s gravitational constant.

Instanton – A classical solution of Euclidean field theory with finite action.

Kaluza–Klein gauge field – In a compactified theory, the gauge field which arises from the metric of the higher-dimensional theory.

K3 – The unique Calabi–Yau manifold in four dimensions having an $SU(2)$ holonomy.

Loop levels – In a Feynman diagram expansion of a field theory, terms which contribute in higher orders of the Planck constant \hbar .

Macroscopic entropy – Entropy associated with gravitational backgrounds via the Bekenstein–Hawking formula or its generalization.

Microscopic entropy – Entropy which follows from the degeneracy of states of a system via Boltzmann’s relation.

Minimally coupled scalar – A scalar field whose equation of motion is the standard Klein–Gordon equation where the derivatives are covariant derivatives.

Neveu–Schwarz/Neveu–Schwarz states – In type I and II string theories, bosonic closed-string states whose left- and right-moving parts are bosonic.

No-hair theorem – A theorem in general relativity which states that black holes with nonsingular horizons are uniquely characterized by their mass, angular momenta, and charges which can couple to long-range gauge fields.

Orbifold – A coset space M/G where G is a group of discrete symmetries of a manifold M . If G has a fixed point, the space is singular.

p-Form – A fully antisymmetric p -index tensor.

Ramond–Ramond states – In type I and II string theories, bosonic closed-string states whose left- and right-moving parts are fermionic.

Reissner–Nordstrom black hole – Black hole solution of general relativity with electric Maxwell charge.

S^n – n -Dimensional sphere.

Supergravity – Supersymmetric extension of general relativity.

Supersymmetry – A symmetry between bosons and fermions.

Threshold bound state – A bound state which is marginally bound, that is, the binding energy is zero.

Tree level – In a Feynman diagram expansion of a field theory, terms which contribute to lowest order of the Planck constant \hbar .

$U(N)$ – The group of $N \times N$ unitary matrices. If the determinant is unity, the subgroup is called $SU(N)$.

Further Reading

Callan CG and Maldacena M (1996) D-brane approach to black hole quantum mechanics. *Nuclear Physics B* 472: 591 (arXiv:hep-th/9602043).

Dabholkar A (2004) Exact counting of black hole microstates, arXiv:hep-th/0409148.

Das SR and Mathur SD (1996) Comparing decay rates for black holes and D-branes. *Nuclear Physics B* 478: 561 (arXiv:hep-th/9606185).

Das SR and Mathur SD (2001) The quantum physics of black holes: results from string theory. *Annual Review of Nuclear and Particle Science* 50: 153 (arXiv:gr-qc/0105063).

David JR, Mandal G, and Wadia SR (2002) Microscopic formulation of black holes in string theory. *Physics Reports* 369: 549 (arXiv:hep-th/0203048).

Dijkgraaf R, Moore GW, and Verlinde E (1996) Elliptic genera of symmetric products and second quantized strings. *Communications in Mathematical Physics* 185: 197 (arXiv:hep-th/9608096).

’t Hooft G (1990) The black hole interpretation of string theory. *Nuclear Physics B* 335: 138.

Johnson C (2003) *D-Branes*. Cambridge: Cambridge University Press.

Maldacena JM (1996) Black holes in string theory, arXiv:hep-th/9607235.

Maldacena J, Strominger A, and Witten E (1997) Black hole entropy in M-theory. *Journal of High Energy Physics* 9712: 002 (arXiv:hep-th/9711053).

Mathur SD (2004) Where are the states of a black hole?, arXiv:hep-th/0401115.

Mohaupt T (2000) Black hole entropy, special geometry and strings. *Fortschritte der Physik* 49: 3 (arXiv:hep-th/0007195).

Sen A (1995) Extremal black holes and elementary string states. *Modern Physics Letters A* 10: 2081.

Strominger A and Vafa C (1996) Microscopic origin of the Bekenstein–Hawking entropy. *Physics Letters B* 379: 99 (arXiv:hep-th/9601029).

Susskind L (1993) Some speculations about black hole entropy in string theory, arXiv:hep-th/9309145.

Wald R (1994) *Quantum Field Theory In Curved Space-Time and Black Hole Thermodynamics*. Chicago, IL: University of Chicago Press.

Breaking Water Waves

A Constantin, Trinity College, Dublin,
Republic of Ireland

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Watching the sea or a lake it is often possible to trace a wave as it propagates on the water's surface. One can roughly distinguish two types of breaking waves. All waves break while reaching the shore but certain waves break far from the shore. In the first case, the change in water depth or the presence of an obstacle (e.g., a rock) seems to cause wave breaking, while for certain waves within the second category, these factors appear not to be essential. It is a matter of observation that for many waves that break in the open water a drastic increase in their slope near breaking is noticeable. This leads us to the following mathematical definition: the wave profile gradually steepens as it propagates until it develops a point where the slope is vertical and the wave is said to have broken (Whitham 1980). Throughout this article, we are concerned with wave breaking that is not caused by a drastic change of the topography of the bottom; for a discussion of wave breaking at the beach we refer to Johnson (1997). The governing equations for water waves (see the next section) are too difficult to be dealt with in their full generality. Therefore, to gain some insight, one has to find simpler models that are more tractable mathematically. Investigating the properties of the model, certain predictions can be made. The conclusions reached will reflect reality only to some limited extent. The value of a model depends on the number and the degree of accuracy of physically useful deductions that can be made from it – the “truth” of the model is meaningless as all experiments contain inaccuracies and effects other than those accounted for (while deriving the model) cannot be totally excluded. We intend to discuss the way in which a recent model due to Camassa and Holm (1993) can lead to a better understanding of breaking water waves. Firstly we survey a few classical nonlinear partial differential equations that model the propagation of water waves over a flat bed (within the confines of the linear theory one cannot cope with the wave breaking phenomenon) and discuss their relevance to the study of breaking waves. We then analyze the breaking of waves within the context of the Camassa–Holm equation: existence of breaking waves, criteria that guarantee that a certain initial shape develops into a breaking wave, specific

features of wave breaking (blow-up rate and blow-up set for certain types of breaking waves). We conclude the presentation with a discussion of the way in which solutions to the Camassa–Holm equation can be continued after wave breaking.

The Governing Equations

The water waves that one typically sees propagating on the surface of the sea or on a lake are, as a matter of common experience, approximately two dimensional. That is, the motion is identical in any direction parallel to the crest line. To describe these waves, it suffices to consider a cross section of the flow that is perpendicular to the crest line. Choose Cartesian coordinates (x, y) with the y -axis pointing vertically upwards and the x -axis being the direction of wave propagation, while the origin lies at the mean water level. Let $(u(t, x, y), v(t, x, y))$ be the velocity field of the flow, let $y = -d$ be the flat bed (for some fixed $d > 0$), and let $y = \eta(t, x)$ be the water's free surface. Homogeneity (constant density) is a physically reasonable assumption for gravity waves (Johnson 1997), and it implies the equation of mass conservation

$$u_x + v_y = 0 \quad [1]$$

The inviscid setting is realistic since experimental evidence confirms that the length scales associated with an adjustment of the velocity distribution due to laminar viscosity or turbulent mixing are long compared to typical wavelengths. Under the assumption of inviscid flow the equation of motion is Euler's equation

$$\begin{aligned} u_t + uu_x + vv_y &= -P_x \\ v_t + uv_x + vv_y &= -P_y - g \end{aligned} \quad [2]$$

where $P(t, x, y)$ denotes the pressure and g is the gravitational constant of acceleration. The free surface decouples the motion of the water from that of the air so that (Johnson 1997) the dynamic boundary condition

$$P = P_0 \quad \text{on } y = \eta(t, x) \quad [3]$$

must hold if we neglect surface tension, where P_0 is the (constant) atmospheric pressure. Moreover, since the same particles always form the free surface, we have the kinematic boundary condition

$$v = \eta_t + u\eta_x \quad \text{on } y = \eta(t, x) \quad [4]$$

On the flat bed we have the kinematic boundary condition

$$v = 0 \quad \text{on } y = -d \quad [5]$$

expressing the fact that the flow is tangent to the horizontal bed (or, equivalently, that water cannot penetrate the rigid bed). The governing equations for water waves are [1]–[5]. Other than the fact that they are highly nonlinear, a main difficulty in analyzing the governing equations lies in the fact that we deal with a free boundary problem: the free surface $y = \eta(t, x)$ is not specified *a priori*. In our discussion, we suppose that initially (at time $t = 0$), a disturbance of the flat surface of still water was created and we analyze the subsequent motion of the water. The balance between the restoring gravity force and the inertia of the system governs the evolution of the mass of water and our primary objective is the behavior of the free surface.

An important category of flows are those of zero vorticity, characterized by the additional assumption

$$u_y = v_x \quad [6]$$

The vorticity of a flow, $\omega = u_y - v_x$, measures the local spin or rotation of a fluid element. In flows for which [6] holds the local whirl is completely absent and for this reason such flows are called irrotational. Relation [6] ensures the existence of a velocity potential, namely a function $\phi(t, x, y)$ defined up to a constant via

$$\phi_x = u, \quad \phi_y = v$$

Notice that [1] ensures that ϕ is a harmonic function, that is, $(\partial_x^2 + \partial_y^2)\phi = 0$. In this way, the powerful methods of complex analysis become available for the study of irrotational flows. Thus, while most water flows are with vorticity, the study of irrotational flows can be defended mathematically on grounds of beauty. Concerning the physical relevance of irrotational water flows, experimental evidence indicates that for waves entering a region of still water the assumption of irrotational flow is realistic (Johnson 1997). Moreover, as a consequence of Kelvin’s circulation theorem (Acheson 1990), a water flow that is irrotational initially has to be irrotational at all later times. It is thus reasonable to consider that water motions starting from rest will remain irrotational at later times.

Nonlinear Model Equations

Starting from the governing equations [1]–[6] one can derive a variety of model equations using the non-dimensionalization and scaling approach: a suitable set of nondimensional variables is introduced, which, after scaling, leads to the appearance of parameters. The sizes and relative sizes of these parameters then govern the type of phenomenon that is of interest. An asymptotic expansion in one or several parameters

yields an equation that is usually of significance in some region of space/time. The aim of this process is to obtain a simpler model that can be used to gain some understanding and to make some predictions for specific physical processes. This scaling method yields the Korteweg–de Vries (KdV) equation

$$\eta_t + \eta\eta_x + \eta_{xxx} = 0, \quad t > 0, x \in \mathbb{R} \quad [7]$$

as a model for the unidirectional propagation of shallow water waves over a flat bed (Johnson 1997). In [7] the function $\eta(t, x)$ represents the height of the water’s free surface above the flat bed. We would like to emphasize that the “shallow water” regime does not refer to water of insignificant depth – it indicates that the typical wavelength is much larger than the typical depth (e.g., tidal waves are considered to be shallow water waves although they affect the motion of the deep sea). The KdV model admits the solitary wave solutions

$$\eta_c(t, x) = 3c \operatorname{sech}^2\left(\frac{\sqrt{c}}{2}(x - ct)\right), \quad c \in \mathbb{R} \quad [8]$$

For any fixed $c > 0$, the profile η_c propagates without change of form at constant speed c on the surface on the water, that is, it represents a traveling wave. Since the profiles [8] of the traveling waves drop rapidly to the undisturbed water level $\eta = 0$ ahead and behind the crest of the wave, η_c are called solitary waves. Notice that [8] shows that taller solitary waves travel faster. They have other special properties: an initial profile consisting of two solitary waves, with the taller preceding the smaller one, evolves in such a way that the taller wave catches up the other, there is a period of complicated nonlinear interaction but eventually both solitary waves emerge completely unscathed! This special type of nonlinear interaction (the superposition principle is not valid since KdV is a nonlinear equation) in which solitary waves regain their form upon collision occurs only for special equations, in which case the solitary waves are called solitons. A further interesting property of the KdV model, relevant for the understanding of the interaction of solitons, is the fact that it is completely integrable (McKean 1998): there is a transformation which converts the equation into an infinite sequence of linear ordinary differential equations which can be trivially integrated. Moreover, the KdV-solitons η_c are stable: an initial profile that is close to the form of a soliton will evolve into a wave that at any later times has a form close to that of a soliton (Benjamin 1972). Despite all these intriguing features of the KdV-model, for all initial profiles $x \mapsto \eta(0, x)$ within the Sobolev space $H^1(\mathbb{R})$ of square-integrable functions with a square-integrable distributional derivative, eqn [7] has a unique solution

defined for all times $t \geq 0$ (cf. Kenig *et al.* (1996)) so that the KdV model cannot be used to shed light on the wave breaking phenomenon.

Whitham (1980) suggested the equation

$$\eta_t + \eta\eta_x + \int_{\mathbb{R}} k(x-y)\eta_y(t,y)dy = 0 \quad [9]$$

for the free surface profile $x \mapsto \eta(t,x)$, with the singular kernel

$$k(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \left(\frac{\tanh(\xi)}{\xi} \right)^{1/2} e^{i\xi x} d\xi$$

to model wave breaking. It can be shown (see Constantin and Escher (1998) and references therein) that [9] describes wave breaking: there are smooth initial profiles $x \mapsto \eta(0,x)$ such that the resulting unique solution of [9] exists on a maximal time interval $[0, T)$ with

$$\sup_{(t,x) \in [0,T) \times \mathbb{R}} \{\eta(t,x)\} < \infty$$

$$\inf_{x \in \mathbb{R}} \{\eta_x(t,x)\} \rightarrow -\infty \quad \text{as } t \uparrow T$$

(the solution remains bounded but its slope becomes infinite in finite time). However, in contrast to the KdV model, eqn [9] is not integrable and does not possess soliton solutions. As emphasized by Whitham (1980), it is intriguing to find models for water waves which exhibit both soliton interaction and wave breaking.

The Camassa–Holm equation

$$\eta_t - \eta_{txx} + 3\eta\eta_x = 2\eta_x\eta_{xx} + \eta\eta_{xxx} \quad [10]$$

was first obtained by Fokas and Fuchssteiner (1981/82) as a nonlinear partial differential equation with infinitely many conservation laws. Camassa and Holm (1993) derived [10] as a model for shallow water waves, established that the equation possesses soliton solutions and found that it is formally integrable (for a discussion of the integrability issues we refer to Constantin (2001), and Lenells (2002)). Moreover, the solitons of [10] are stable (Constantin and Strauss 2003). An astonishing plentitude of structures is tied into the Camassa–Holm equation: [10] is a re-expression of geodesic flow on the diffeomorphism group (Constantin 2000, Kouranbaeva 1999), a property that can be used to show that the least action principle holds in the sense that there is a unique flow transforming a wave profile into a nearby profile within the class of flows that minimize the kinetic energy (see the discussion in Constantin (2000) and Constantin and Kolev (2003)). Interestingly, the Camassa–Holm equation also models wave breaking. More precisely (see the discussion in Constantin (2000)), for any initial data $x \mapsto \eta_0(x) = \eta(0,x)$ in

$H^3(\mathbb{R})$ there is a unique solution of [10] defined on some maximal time interval $[0, T)$ and the solution stays uniformly bounded on $[0, T)$ with

$$\lim_{t \uparrow T} \left(\inf_{x \in \mathbb{R}} \{\eta_x(t,x)\} (T-t) \right) = -2 \text{ if } T < \infty$$

In addition to this, for a large class of initial data, there is precisely one point where the slope of the wave becomes infinite at breaking time (Constantin 2000): if $\eta_0 \not\equiv 0$ is odd and such that $\eta_0(x) - \eta_0''(x) \geq 0$ for all $x \leq 0$, then the corresponding wave $t \mapsto [x \mapsto \eta(t,x)]$ will break in finite time $T < \infty$ and

$$\lim_{t \uparrow T} \eta_x(t,0) = -\infty$$

whereas

$$|\eta_x(t,x)| \leq K + K \frac{\cosh(x)}{|\sinh(x)|}$$

$$t \in [0, T), \quad x \neq 0$$

for some constant $K > 0$. Thus, the Camassa–Holm model is an integrable infinite-dimensional Hamiltonian system with stable solitons and eqn [10] admits also breaking waves as local solutions (see Constantin and Escher (1998) and McKean (1998) and references therein for further results on wave breaking for the Camassa–Holm equation).

We conclude our discussion by pointing out that it is possible to continue solutions of the Camassa–Holm equation past the breaking time. For this purpose it is convenient to rewrite [10] as the nonlinear nonlocal conservation law

$$\eta_t + \eta\eta_x + \frac{1}{2} \partial_x \int_{\mathbb{R}} e^{-|x-y|} \left(\eta^2 + \frac{\eta_x^2}{2} \right) dy = 0 \quad [11]$$

reminiscent to some extent to the form of [7] and [9] and obtained by formally applying the operator $(1 - \partial_x^2)^{-1}$ to [10] in view of the fact that

$$(1 - \partial_x^2)^{-1} f = P * f \quad \text{for } f \in L^2(\mathbb{R})$$

the kernel of the convolution being

$$P(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}$$

By introducing a new set of independent and dependent variables it is possible to resolve all singularities due to wave breaking in the sense that [11] is transformed into a semilinear system, the unique solution of which can be obtained as a fixed point of a contractive operator (Bressan and Constantin 2005). In terms of [11], a semigroup of global conservative solutions (in the sense that the total energy

$$\frac{1}{2} \int_{\mathbb{R}} (\eta^2 + \eta_x^2) dx$$

equals a constant, for almost every time), depending continuously on the initial data $\eta(0, \cdot) \in H^1(\mathbb{R})$, is thus constructed.

See also: Compressible Flows: Mathematical Theory; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Integrable Systems: Overview; Interfaces and Multicomponent Fluids.

Further Reading

- Acheson DJ (1990) *Elementary Fluid Dynamics*. New York: Oxford University Press.
- Benjamin TB (1992) The stability of solitary waves. *Proceedings of the Royal Society of London Series A* 328: 153–183.
- Bressan A and Constantin A (2005) Global conservative solutions of the Camassa–Holm equation, *Preprints on Conservation Laws* 2005-016 (www.math.ntnu.no/conservation/2005/016).
- Camassa R and Holm DD (1993) A new integrable shallow water equation with peaked solitons. *Physical Review Letters* 71: 1661–1664.
- Constantin A (2000) Existence of permanent and breaking waves for a shallow water equation: a geometric approach. *Annales de l'Institut Fourier (Grenoble)* 50: 321–362.
- Constantin A (2001) On the scattering problem for the Camassa–Holm equation. *Proceedings of the Royal Society of London Series A* 457: 953–970.
- Constantin A and Escher J (1998) Wave breaking for nonlinear nonlocal shallow water equations. *Acta Mathematica* 181: 229–243.
- Constantin A and Kolev B (2003) Geodesic flow on the diffeomorphism group of the circle. *Commentarii Mathematici Helvetica* 78: 787–804.
- Constantin A and Strauss WA (2000) Stability of peakons. *Communications on Pure and Applied Mathematics* 53: 603–610.
- Fokas AS and Fuchssteiner B (1981/82) Symplectic structures, their Bäcklund transformations and hereditary symmetries. *Physica D* 4: 47–66.
- Gesztesy F and Holden H (2003) *Soliton Equations and their Algebraic-Geometric Solutions*. Cambridge: Cambridge University Press.
- Johnson RS (1997) *A Modern Introduction to the Mathematical Theory of Water Waves*. Cambridge: Cambridge University Press.
- Johnson RS (2002) Camassa–Holm, Korteweg–de Vries and related models for water waves. *Journal of Fluid Mechanics* 455(2002): 63–82.
- Kenig CE, Ponce G, and Vega LA (1996) A bilinear estimate with applications to the KdV equation. *Journal of the American Mathematical Society* 9: 573–603.
- Kouranbaeva S (1999) The Camassa–Holm equation as a geodesic flow on the diffeomorphism group. *Journal of Mathematical Physics* 40: 857–868.
- Lenells J (2002) The scattering approach for the Camassa–Holm equation. *Journal of Nonlinear Mathematical Physics* 9: 389–393.
- McKean HP (1979) Integrable systems and algebraic curves. In: *Global Analysis, Lecture Notes in Mathematics*, vol. 755, pp. 83–200. Berlin: Springer.
- McKean HP (1998) Breakdown of a shallow water equation. *Asian Journal of Mathematics* 2: 867–874.
- Whitham GB (1980) *Linear and Nonlinear Waves*. New York: Wiley.

BRST Quantization

M Henneaux, Université Libre de Bruxelles, Bruxelles, Belgium

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The BRST symmetry was originally introduced in the seminal papers by [Becchi et al. \(1976\)](#) and [Tyutin \(1975\)](#) for Yang–Mills gauge theories as a tool for controlling the renormalization of the models in a consistent (gauge-independent) way. This symmetry was discovered as a residual symmetry of the gauge-fixed action. It was realized later that, in fact, the BRST construction is quite general, in the sense that it covers arbitrary gauge theories and not just Yang–Mills gauge models. Furthermore, it is intrinsic, in that no gauge choice is actually necessary to define it.

The purpose of this review is to explain the general, intrinsic features of the BRST formalism applicable to “any” gauge theory. The proper setting for discussing these issues is that of homological algebra ([Stasheff \(1998\)](#), and references therein). This article first explains

the necessary algebraic material underlying the construction and then illustrates it in the cases of the Hamiltonian BRST formalism and the Lagrangian BRST formalism.

A Result from Homological Algebra

The main result of homological algebra needed in the BRST construction deals with a differential complex \mathcal{C} with two gradings. The first grading is an \mathbb{N} -degree and is called the “resolution degree,” or “r-degree.” The second grading is a \mathbb{Z} -degree and is called the total ghost number. It is denoted by gh . We assume that there are two odd derivations δ and s_0 that have the following properties:

$$\begin{aligned} r(\delta) &= -1, & \text{gh}(\delta) &= 1 \\ r(s_0) &= 0, & \text{gh}(s_0) &= 1 \end{aligned} \quad [1]$$

and

$$\delta^2 = 0, \quad s_0 \delta + \delta s_0 = 0, \quad s_0^2 = -[\delta, s_1] \quad [2]$$

for some derivation s_1 of r-degree 1 and ghost number 1. The bracket $[\cdot, \cdot]$ is the graded commutator – in this specific case, the anticommutator. We also assume that the homology of δ vanishes at nonzero value of the r-degree, both in the original complex \mathcal{C} ,

$$H_k(\delta, \mathcal{C}) = 0, \quad k > 0 \quad [3]$$

(which is equivalent to $\delta a = 0, r(a) > 0 \Rightarrow a = \delta b$) and in the space of derivations,

$$[\alpha, \delta] = 0, \quad r(\alpha) \neq 0 \Rightarrow \alpha = [\beta, \delta] \quad [4]$$

where α and β are both derivations in \mathcal{C} . The r-degree of a homogeneous linear operator α is defined through $r(\alpha(x)) = r(\alpha) + r(x)$ for any element $x \in \mathcal{C}$ and is negative when α decreases the r-degree.

In $H_0(\delta, \mathcal{C})$, the (odd) derivation s_0 defines a differential. The cohomology of s_0 modulo δ , denoted $H^k(s_0, H_0(\delta, \mathcal{C}))$, is the cohomology of s_0 in $H_0(\delta, \mathcal{C})$. It is explicitly defined through the cocycle condition

$$s_0 a = \delta m \quad [5]$$

with coboundaries of the form

$$s_0 b + \delta n \quad [6]$$

The central result underlying the BRST construction is:

Theorem 1 *Given the above setting, there exists an odd derivation s in \mathcal{C} with the following properties:*

$$s = \delta + s_0 + s_1 + \dots \quad [7]$$

$$r(s_k) = k, \quad \text{gh}(s_k) = 1 \quad [8]$$

$$s^2 = 0 \quad [9]$$

Furthermore, one has

$$H^k(s, \mathcal{C}) = H^k(s_0, H_0(\delta, \mathcal{C})) \quad [10]$$

The proof is straightforward (see, e.g., Henneaux and Teitelboim (1992)). In particular, the proof of [10] is a standard spectral sequence argument with a sequence that collapses after the second step. It is interesting to note that, contrary to s_0 , which is only a differential modulo δ , s is a true differential. The construction of s provides a model for $H^k(s_0, H_0(\delta, \mathcal{C}))$. The differential s is not unique, but this does not affect the subsequent discussion.

In physical applications, the total ghost number is a derived quantity. The primary gradings are the resolution degree and the “filtration degree” called the pure ghost number and denoted pgh . It is an \mathbb{N} -degree and one has

$$\text{gh} = \text{pgh} - r \quad [11]$$

The r-degree is known as the antighost or antifield number, depending on the context (see below). When $r(x) = 0$, one has $\text{gh}(x) = \text{pgh}(x)$. Since the pure ghost number is non-negative, this implies that

$$H^k(s, \mathcal{C}) = 0, \quad k < 0 \quad [12]$$

A Geometric Application

Geometric Setting

Theorem 1 is relevant to the following situation. Consider a surface Σ in a manifold M , defined by equations

$$f_a = 0 \quad [13]$$

which may or may not be independent. (We assume for definiteness that the variables in M are bosonic, that is, that M is an ordinary manifold – as opposed to a supermanifold. The graded case can be covered without difficulty by including appropriate sign factors at the relevant places.) Assume that Σ is partitioned by orbits generated by vector fields X_α defined everywhere in M , tangent to Σ and closing on Σ in the Lie bracket,

$$[X_\alpha, X_\beta] = C^\gamma_{\alpha\beta} X_\gamma + \text{“more”} \quad [14]$$

where “more” denotes terms that vanish on Σ . We assume, for simplicity, that the vector fields X_α are linearly independent of Σ , although this is not necessary. The formalism can be developed in the nonindependent case, but it then requires more variables. We are interested in the quotient space Σ/\mathcal{O} of the surface Σ by the orbits. To guide the geometrical intuition, we shall assume that this quotient space is a smooth manifold (the fiber of the orbits, etc.), and we shall suggestively adopt notations adapted to this best possible case. The approach, being purely algebraic, is in fact more general. (Accordingly, the notations should be understood with a liberal mind.)

The aim here is to describe the algebra of “observables,” that is, the algebra $C^\infty(\Sigma/\mathcal{O})$ of functions on the quotient space Σ/\mathcal{O} . The terminology “observables” anticipates the physical situation discussed below, where the orbits are the “gauge orbits.” In order to describe algebraically the algebra of observables, one observes that this algebra is obtained

through a two-step procedure. First, one restricts the functions from M to Σ . Second, one imposes the invariance condition along the orbits. To each of these steps corresponds a separate differential.

Longitudinal Complex

The longitudinal complex is associated with the second step. One can consider on Σ an “exterior derivative operator D along the gauge orbits.” This operator is defined on functions on Σ as

$$Df = X_\alpha(f)C^\alpha \quad [15]$$

where the 1-forms C^α dual to the X 's are called ghosts. In the physical context, the form-degree is the pgh described earlier, and so $\text{pgh}(C^\alpha) = 1$. The action of D on the ghosts is given by

$$DC^\alpha = -\frac{1}{2}C^\gamma{}_{\alpha\beta}C^\alpha C^\beta \quad [16]$$

The longitudinal complex \mathcal{L}_Σ is the complex of exterior forms along the gauge orbits. In our representation used here, it is given by the space of polynomials in the ghosts C^α with coefficients that are functions on Σ . The exterior derivative D is defined on this space by extending the formulas [15] and [16] so that it is an odd derivation. One clearly has (on Σ)

$$D^2 = 0 \quad [17]$$

The functions on the quotient space Σ/\mathcal{O} are just the elements of the zeroth cohomological group $H^0(D, \mathcal{L}_\Sigma)$,

$$H^0(D, \mathcal{L}_\Sigma) = C^\infty(\Sigma/\mathcal{O}) \quad [18]$$

In general, $H^k(D, \mathcal{L}_\Sigma) \neq 0$.

Koszul–Tate Differential δ

The Koszul–Tate differential δ implements the first step in the reduction procedure. More precisely, it provides an algebraic resolution of the algebra $C^\infty(\Sigma)$ of the smooth functions on the surface Σ .

That algebra can be identified with the quotient algebra

$$C^\infty(\Sigma) = C^\infty(M)/\mathcal{N} \quad [19]$$

where \mathcal{N} is the ideal of functions that vanish on Σ . The Koszul–Tate complex \mathcal{K} is defined by adding one new generator for each equation $f_a = 0$ defining Σ , denoted t_a^* and assigned r-degree 1. In the algebra $C^\infty(M) \otimes \wedge(t_a^*)$ (where $\wedge(t_a^*)$ is the exterior algebra on t_a^*), one defines δ through

$$\delta f = 0 \quad \forall f \in C^\infty(M), \quad \delta t_a^* = f_a \quad [20]$$

and extends it as an odd derivation. It is clear that $r(\delta) = -1$ and that $\delta^2 = 0$. Because the

functions on M are annihilated by δ , they are clearly cycles at r-degree zero. Because the left-hand side f_a of the equations $f_a = 0$ are exact (equal to δt_a^*), the ideal \mathcal{N} coincides with the set of boundaries in degree zero.

Thus,

$$H_0(\delta, \mathcal{K}) = C^\infty(\Sigma) \quad [21]$$

We see accordingly that δ successfully enforces the restriction to the surface Σ through its homology in degree zero.

However, if the equations $f_a = 0$ are not independent, this is not the end of the story. Indeed, any identity $Z_A^a f_a = 0$ on the functions f_a leads to a nontrivial cycle $Z_A^a t_a^*$ in r-degree 1, $\delta(Z_A^a t_a^*) = 0$. This is undesirable. To cure this drawback, one introduces further generators t_A^* in r-degree 2, one for each identity $Z_A^a f_a = 0$, and defines

$$\delta t_A^* = Z_A^a t_a^*, \quad r(t_A^*) = 2 \quad [22]$$

in order to “kill” the unwanted cycles $Z_A^a t_a^*$. The Koszul complex \mathcal{K} is thus enlarged to contain these new (even) variables and redefined as

$$\mathcal{K} = C^\infty(M) \otimes \wedge(t_a^*) \otimes S(t_A^*) \quad [23]$$

where $S(t_A^*)$ is the symmetric algebra in t_A^* . The operator δ is extended to \mathcal{K} as an odd derivation. One has $\delta^2 = 0$ and the property [21] is unaffected by the inclusion of the new generators. Furthermore, by construction,

$$H_1(\delta, \mathcal{K}) = 0 \quad [24]$$

If there is no “identity on the identities,” we shall assume that the process stops. Otherwise, one needs to introduce further generators in r-degree 3 and possibly higher. When all the appropriate variables are included, there is no homology at higher r-degree. Thus,

$$H_k(\delta, \mathcal{K}) = 0, \quad k > 0 \quad [25]$$

Combining δ with D

We now turn to the problem of combining the Koszul–Tate complex with the longitudinal complex, so as to implement the full reduction. To that end, we define \mathcal{C} by adding the ghosts to \mathcal{K} ,

$$\mathcal{C} = \mathcal{K} \otimes \wedge(C_\alpha) = 0 \quad [26]$$

We then extend the action of the Koszul–Tate differential in the simplest way which preserves all gradings, namely

$$\delta C_\alpha = 0 \quad [27]$$

It is clear that the homology of δ in \mathcal{C} is given by

$$H_0(\delta, \mathcal{C}) = \mathcal{L}_\Sigma, \quad H_k(\delta, \mathcal{C}) = 0 \quad (k > 0) \quad [28]$$

One can also extend the longitudinal derivative D to the whole complex \mathcal{C} because the vector fields X_α are defined throughout M and so, the definitions [15] and [16] make sense in \mathcal{C} . One defines the action of D on the generators t^* by requiring that

$$D\delta + \delta D = 0 \quad [29]$$

This is easily verified to be possible. However, the (odd) derivation so obtained fails to be a differential in \mathcal{C} when the vector fields X_α do not close off the surface Σ . In that case, the gauge transformations are not integrable off Σ ; one says that they form an “open algebra.” One has then $D^2 = 0$ only on Σ , or, more precisely,

$$D^2 = -\delta s_1 - s_1 \delta \quad [30]$$

for some (odd) derivation s_1 (that vanishes in the “closed algebra” case). But this situation is precisely the one discussed earlier, with the Koszul–Tate differential being indeed δ , as anticipated by the notation, and the longitudinal differential D playing the role of s_0 (the degrees also match). Applying the theorem discussed there, we can conclude:

Theorem 2 *There exists a differential s in \mathcal{C} ,*

$$s = \delta + D + s_1 + \dots, \quad s^2 = 0 \quad [31]$$

such that

$$H^0(s, \mathcal{C}) = C^\infty(\Sigma/\mathcal{O}) \quad [32]$$

This is an immediate consequence of [Theorem 1](#) and [eqns \[18\] and \[28\]](#). The differential s is known in the physical applications described below as the BRST differential.

Hamiltonian BRST Construction

As a first application of the above setting, we consider the Hamiltonian description of gauge systems. As already known, gauge systems are characterized in the Hamiltonian description by constraints and, for this reason, are called “constrained Hamiltonian systems.” Furthermore, the gauge transformations generate gauge orbits on the constraint surface and the physical observables are the functions on the quotient space of the constraint surface by the gauge orbits.

A further important feature arises in the Hamiltonian formalism: the gauge transformations are

canonical transformations that are generated by the first-class constraints. Assuming that all the second-class constraints have been eliminated and that the bracket being used is the Dirac bracket, one sees that there is a vector field X_α for each constraint function $f_\alpha, \alpha \equiv a$. (The functions f_α are thus assumed to be independent since the vector fields X_α are assumed to be so. If not, further variables are needed, but the analysis proceeds along the same ideas.)

This implies, in turn, that there is a pairing between the ghosts C^a associated with the longitudinal exterior derivative and the generators t_a^* of the Koszul–Tate complex. This pairing enables one to extend the bracket structure defined on the phase space to the pairs (C^a, t_a^*) by declaring that these are canonically conjugate. The variables t_a^* are the momenta conjugate to the ghosts, $[t_a^*, C^b] = \delta_a^b$. Accordingly, the complex \mathcal{C} relevant to the Hamiltonian situation,

$$\mathcal{C} = C^\infty(P) \otimes \wedge(C^a) \wedge (t_a^*) \quad [33]$$

has a phase-space structure (here, $P \equiv M$ is the manifold obtained after eliminating the second-class constraints, equipped with the Dirac bracket). The space \mathcal{C} is known as the “extended phase space.” The r-degree is called “antighost number” in the Hamiltonian context.

By the general theorem described in the previous section, one knows that the cohomology at $\text{gh} = 0$ of the BRST differential is isomorphic to the algebra of the observables. Thus, there are two alternative ways to describe this physical algebra, either through reduction, by eliminating the redundant (gauge) variables, or cohomologically in an extended space containing additional variables, the ghosts, and their momenta.

There is an additional interesting feature of the BRST construction in the Hamiltonian case: the BRST transformation is a canonical transformation in the extended phase space, in the sense that

$$sF = [\Omega, F] \quad [34]$$

for some “BRST generator” Ω of ghost number 1 ($F, \Omega \in \mathcal{C}$). The nilpotency s^2 of the BRST differential is equivalent to

$$[\Omega, \Omega] = 0 \quad [35]$$

That s is canonically generated implies that the cohomological BRST groups come with a natural bracket structure: the Poisson bracket of the extended phase space passes on to the BRST cohomological groups. In particular, $H^0(s, \mathcal{C})$, equipped with this bracket structure, is isomorphic (as Poisson algebra) to the algebra of physical observables.

Lagrangian BRST Construction

The analysis of the Lagrangian BRST construction, due to [Batalin and Vilkovisky \(1981\)](#) (“antifield formalism”), proceeds in the same way because the covariant description of the space of observables involves also the same geometric ingredients. The surface Σ is now the “stationary surface,” that is, the space of solutions to the equations of motion. The space M in which it is embedded is the space of all field histories. The gauge symmetry acts on this space. Furthermore, the gauge vector fields are tangent to Σ since a solution is mapped on a solution by a gauge transformation. The integral submanifolds are the gauge orbits. The observables are the functions on the quotient space.

Since the equations of motion follow from an action principle, there are as many equations as there are fields φ^i . The corresponding generators t_a^* in the Koszul–Tate complex (at degree 1) are called “antifields conjugate to the fields” and are denoted φ_i^* . The r-degree is known as “antifield” (or also “antighost”) number. The gauge symmetry of the action implies Noether identities on the equations of motion. These are, therefore, not independent. According to the above general discussion, there are further generators in the Koszul–Tate complex, at degree 2. More precisely, there are as many new generators in degree 2 as there are Noether identities or independent gauge symmetries. These are called antifields conjugate to the ghosts and denoted C_α^* .

In the longitudinal complex, one has the ghosts C^α , with as many ghosts as there are gauge symmetries. Thus, the BRST complex is the space

$$\mathcal{C} = C^\infty(M) \otimes \wedge(C^\alpha) \otimes \wedge(\varphi_i^*) \otimes S(C_\alpha^*) \quad [36]$$

where M is the space of all field histories. There is now a natural pairing between the original field variables φ^i and the antifields φ_i^* , as well as between the ghosts C^α and the antifields C_α^* . One thus defines a bracket in which the fields φ^i and the ghosts C^α on the one hand, and the antifields φ_i^* and C_α^* on the other, are declared to be conjugate. This bracket is denoted by parentheses,

$$(\varphi^i, \varphi_j^*) = \delta_j^i, \quad (C^\alpha, C_\beta^*) = \delta_\beta^\alpha \quad [37]$$

However, since the bracket pairs variables with degrees that add up to -1 , it is in fact an “odd bracket,” called the “antibracket.”

The BRST differential is again canonically generated, but this time in the antibracket,

$$sF = (S, F), \quad F \in \mathcal{C} \quad [38]$$

where the generator S is an even function of the fields, the ghosts and the antifields, with $\text{gh} = 0$ (the

ghost number is carried by the odd antibracket). The nilpotency $s^2 = 0$ of the BRST differential is equivalent to the crucial “master equation,”

$$(S, S) = 0 \quad [39]$$

Because the BRST differential is canonically generated, there is a natural Poisson bracket in cohomology. This bracket is not the Poisson bracket of observables (at $\text{gh} = 0$) because it changes the ghost number by one unit. One can, however, relate it to the Poisson bracket of observables ([Barnich and Henneaux 1996](#)); furthermore, it plays an important role in the study of the consistent deformations of the action.

Spacetime Locality

In the context of local field theory, one is often interested in a particular class of functions of the field histories, namely the so-called space of local functionals. A local functional is, by definition, the integral of a local n -form (where n is the spacetime dimension). A local n -form reads, in local coordinates,

$$\omega = f(x) d^n x \quad [40]$$

where $f(x)$ depends on the fields at x as well as on a finite number of their derivatives. When the ghosts and the antifields are included, the local functions depend on them in the same way.

The previous general cohomological result was derived in the space of all function(al)s, without locality restriction. When changing the space of cochains, one may change the cohomology. For instance, a local functional which is BRST-trivial in the space of all functionals may become nontrivial in the space of local functionals. This indeed happens here because the homology of the Koszul–Tate differentials usually no longer vanishes at strictly positive r-degree in the space of local functionals, where it is related to local conservation laws. As a result, the analysis of the BRST cohomology in the space of local functionals is an interesting and nontrivial problem. In particular, the cohomological groups $H^k(s)$ in the space of local functionals may not vanish at negative ghost numbers.

BRST Quantization

The quantization of a dynamical system can proceed along different lines. For gauge models, the path-integral approach is most efficiently pursued in the context of the antifield formalism. We shall briefly outline here the general principles underlying the

operator approach, which is based on the Hamiltonian formalism.

In the operator approach, all the variables, including the ghosts and the conjugate momenta, are realized as operators in a space endowed with a nonpositive-definite inner product (because of the ghosts and the gauge modes). Real dynamical variables become formally Hermitian operators. Ignoring anomalies, the BRST generator Ω becomes an operator that fulfills the conditions

$$\Omega^* = \Omega, \quad \Omega^2 = 0 \quad [41]$$

(which allows for nontrivial solutions $\Omega \neq 0$ because the inner product is not positive definite). The second relation is a consequence of the classical Poisson bracket relation $[\Omega, \Omega] = 0$ and the fact that the graded Poisson bracket of two odd objects becomes the anticommutator.

To remove the ghost and gauge redundancy, which has no physical content, one must impose a condition that selects physical states. The appropriate condition is motivated by the general cohomological result connecting the BRST cohomology with the algebra of physical observables. One imposes the condition

$$\Omega|\psi\rangle = 0 \quad [42]$$

Because of [41], states of the form $\Omega|\chi\rangle$ are solutions of [42], but they have a vanishing inner product with any other physical states, including themselves. They are called null states. The physical states are given by the BRST state cohomology. The physical operators are given by the BRST operator cohomology at $gh=0$ and induce a well-defined action in the state cohomology. In particular, the Hamiltonian, being gauge invariant in the original theory, is represented by a BRST cohomological class, so that the time evolution maps physical states on physical states.

The whole scheme is (formally) consistent because exact BRST operators have vanishing matrix elements between states annihilated by the BRST operator Ω , while null states $|\phi\rangle$ are such that $\langle\psi|A|\phi\rangle=0$ whenever A is a BRST-closed operator, $[A, \Omega]=0$, and $|\psi\rangle$ a physical state. Problems may arise, however, if the classical relations $[\Omega, \Omega]=0$ and $[H, \Omega]=0$ are not satisfied in presence of extra terms of order \hbar , that is,

$$\Omega^2 \neq 0 \quad \text{or} \quad H\Omega + \Omega H \neq 0 \quad [43]$$

In such cases, one says that they are anomalies. These are usually fatal to the consistency of the theory.

Some Applications

The number of applications of the BRST formalism is so large that it would be out of place to try being

exhaustive here. Some of its main successes are outlined here, with suggestions for “Further reading.”

Renormalization of Gauge Theories

First, there is the original context of perturbative renormalization and anomalies for gauge theories of the Yang–Mills type. The relevant cohomology here is the BRST cohomology in the space of local functionals involving the fields, the ghosts, and the antifields. The antifields are also known in this context as Zinn-Justin sources for the BRST variations of the fields and ghosts, since Zinn-Justin was the first to introduce them (with that meaning). Many authors have contributed to the full computation of the local BRST cohomology. A review is given in Barnich *et al.* (2000), where extensions to other theories are also indicated.

String Theory

Modern string theory would be inconceivable without the BRST formalism. This started with the pioneering paper by Kato and Ogawa (1983), where the critical dimension of the bosonic string was derived from the condition that Ω^2 should vanish (quantum mechanically), and where it was shown that the string physical states could be identified with the state BRST cohomology. The reader is referred to excellent monographs on modern string theory (see “Further reading”).

Deformations of Gauge Models

The study of consistent deformations of a given gauge theory (i.e., the problem of introducing consistent couplings) is also efficiently dealt with in the BRST context. References to applications may be found in Henneaux (1998).

See also: Anomalies; Batalin–Vilkovisky Quantization; BF Theories; Constrained Systems; Functional Integration in Quantum Physics; Graded Poisson Algebras; Indefinite Metric; Perturbative Renormalization Theory and BRST; Quantum Chromodynamics; Quantum Field Theory: A Brief Introduction; Renormalization: General Theory; String Field Theory; Supermanifolds; Topological Sigma Models.

Further Reading

- Barnich G, Brandt F, and Henneaux M (2000) Local BRST cohomology in gauge theories. *Physics Reports* 338: 439.
- Barnich G and Henneaux M (1996) Isomorphisms between the Batalin–Vilkovisky antibracket and the Poisson bracket. *Journal of Mathematical Physics* 37: 5273.

- Batalin IA and Vilkovisky GA (1977) Relativistic S -matrix of dynamical systems with boson and fermion constraints. *Physics Letters* B69: 309.
- Batalin IA and Vilkovisky GA (1981) Gauge algebra and quantization. *Physics Letters* B102: 27.
- Becchi C, Rouet A, and Stora R (1976) Renormalization of gauge theories. *Annals of Physics*, NY 98: 287.
- Fradkin ES and Vilkovisky GA (1975) Quantization of relativistic systems with constraints. *Physics Letters* B55: 224.
- Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*, vols. 1 and 2. Cambridge: Cambridge University Press.
- Henneaux M (1998) Consistent interactions between gauge fields: the cohomological approach. *Contemporary Mathematics* 219: 93.
- Henneaux M and Teitelboim C (1992) *Quantization of Gauge Systems*. Princeton: Princeton University Press.
- Kato M and Ogawa K (1983) Covariant quantization of strings based on BRS invariance. *Nuclear Physics* B212: 443.
- Kugo T and Ojima I (1979) Local covariant operator formalism of nonabelian gauge theories and quark confinement problem. *Progress of Theoretical Physics (Suppl.)* 66: 1.
- Polchinski J (1998) *String Theory*, vols. 1 and 2. Cambridge: Cambridge University Press.
- Stasheff JD (1998) The (secret?) homological algebra of the Batalin–Vilkovisky approach. *Contemporary Mathematics* 219: 195.
- Tyutin IV (1975) Gauge invariance in field theory and statistical physics in the operator formalism. Preprint Lebedev-75-39.

C

C*-Algebras and their Classification

G A Elliott, University of Toronto, Toronto, Canada

© 2006 Elsevier Ltd. All rights reserved.

The study of algebras of Hilbert space operators, closed under the adjoint operation and in the weak operator topology, was begun by John von Neumann shortly after the discovery of quantum mechanics, and partly with the aim of understanding the monolithic ideas proposed by Heisenberg and Schrödinger.

Seventy-five years later, the theory of these algebras has become a monolith in its own right (*see* von Neumann Algebras: Introduction, Modular Theory and Classification Theory; von Neumann Algebras: Subfactor Theory), with more internal structure and with more external reference to physics and, as it turns out, to other areas of mathematics than could possibly have been imagined at the outset. (The most striking example of an application to mathematics is perhaps the discovery of the Jones knot polynomial (*see* The Jones Polynomial); note that this has also had repercussions for physics.)

Twenty-five years after the beginning of the theory of von Neumann algebras, as these algebras are now called, Gelfand and Naimark noticed that a second class of algebras of operators on a Hilbert space, closed under the adjoint operation, was worthy of study, namely those closed in the norm topology. Gelfand and Naimark made two important discoveries concerning this class of operator algebras, now called C*-algebras.

First, Gelfand and Naimark showed that, in the commutative case, at least when the C*-algebra is considered only up to isomorphism – with its identity as a concrete algebra of operators suppressed – the information contained in a C*-algebra is purely topological. More precisely, Gelfand and Naimark showed that the category of unital commutative C*-algebras, with unit-preserving algebra homomorphisms (these necessarily preserve the adjoint operation), is equivalent in a contravariant way (i.e., with reversal of arrows) to the category of compact Hausdorff spaces, with continuous maps. The compact space associated with a

unital commutative C*-algebra under the Gelfand–Naimark correspondence may be viewed as the space of maximal proper ideals, with a natural topology (the hull-kernel, or Jacobson, topology), and is called the spectrum. This space may also be viewed as the set of (unital, linear, multiplicative) maps from the algebra into the complex numbers, in which case the topology is that of pointwise convergence.

Second, using this result, Gelfand and Naimark proved that arbitrary C*-algebras could be axiomatized in a simple way abstractly, as *-algebras – that is, as algebras over the complex numbers with a conjugate linear anti-automorphism of order 2 – with certain special properties. It is now known that the only property that needs to be assumed is the existence of a (necessarily unique) Banach space norm related to the *-algebra structure by means of the so-called C*-algebra identity:

$$\|x^*x\| = \|x^*\| \|x\| \quad [1]$$

This is clearly related to – and in fact implies – the normed algebra inequality

$$\|xy\| \leq \|x\| \|y\| \quad [2]$$

One reason that the Gelfand–Naimark axiomatization of C*-algebras is important is that it underlines how natural it is to consider a C*-algebra abstractly, i.e., independently of any particular representation. Indeed, while one of the fundamental phenomena of von Neumann algebra theory (discovered by Murray and von Neumann) is that, essentially – in rather a strong sense – there is only one way to represent a given von Neumann algebra on a Hilbert space (and there is even a canonical way, called the standard representation!), it is an equally fundamental phenomenon of C*-algebra theory that, except in extremely special cases, this is no longer true.

For instance, although the C*-algebra of compact operators on a given Hilbert space has, up to unitary equivalence, only a single irreducible representation – this is what underlies the fact, proved by von Neumann, referred to as the uniqueness of the

Heisenberg commutation relations for a quantum-mechanical system with finitely many degrees of freedom – as soon as one considers a physical system with infinitely many degrees of freedom, one finds that the naturally associated C^* -algebra has infinitely many – indeed, uncountably many – unitary equivalence classes of irreducible representations, and it is impossible to parametrize these in any reasonable way.

This striking dichotomy presents itself also in other contexts, more elementary perhaps than the physics of infinitely many degrees of freedom. Consider the dynamical system consisting of a circle and a fixed rotation acting on it. If the rotation is of finite order – i.e., if the angle is a rational multiple of 2π – then the naturally associated C^* -algebra is relatively easy to study. In the case of angle zero, it is the unital commutative C^* -algebra with Gelfand–Naimark spectrum the torus. In the general case of a rational angle, the space of unitary equivalence classes of irreducible representations is still naturally parametrized by the torus. (And this is the same as the space of primitive ideals – the kernels of the irreducible representations – with the Jacobson topology.)

In the irrational case – the case of a rotation by an irrational multiple of 2π (still elementary from a geometrical point of view; note that the calendar is based on such a system!) – the irreducible representations are no longer parametrized up to unitary equivalence by the torus – and the space of primitive ideals consists of a single point – the C^* -algebra is simple. (But it is decidedly not simple to study!)

This fundamental dichotomy in the classification of C^* -algebras – conjectured by Gaarding and Wightman in the quantum-mechanical setting and by Mackey in the geometrical one – was established by Glimm. Glimm proved (in the setting of separability; most of his results were generalized later to the nonseparable case) that a large number of *a priori* different ways that a C^* -algebra could behave well were in fact one and the same behavior: either all present for a given C^* -algebra, or all catastrophically absent!

Some of the properties considered by Glimm, and shown to be equivalent (for a separable C^* -algebra) were as follows. First of all, every representation of the C^* -algebra on a Hilbert space should be of type I, i.e., should generate a von Neumann algebra of type I. (A von Neumann algebra was said by Murray and von Neumann to be of type I if it contained a minimal projection of central support one, i.e., a projection not contained in a proper direct summand and minimal with this property.) Second, in every irreducible representation (not necessarily injective) on a Hilbert space, the image of the

C^* -algebra should contain the compact operators. Third, any two irreducible representations with the same kernel should be unitarily equivalent. Fourth, it should be possible to parametrize the unitary equivalence classes of irreducible representations by a real number in a natural way (respecting the natural Borel structure introduced by Mackey).

The first of the equivalent properties listed above, that all representations of a C^* -algebra should be of type I, suggested a name for the property – that the C^* -algebra itself should be of type I. This property of a C^* -algebra, identified by Glimm – or, rather, its opposite, which as mentioned above is much more common (just as irrational numbers are more common than rationals, or systems with infinitely many degrees of freedom are, at least in theory, much more common than those with finitely many degrees of freedom) – is a fundamental unifying principle of nature.

Besides commutative C^* -algebras – as mentioned above, just another way of looking at topological spaces (compact Hausdorff spaces, that is) – and besides the C^* -algebra associated to a rotation or to a physical system with infinitely many degrees of freedom, what are some of the naturally occurring examples of C^* -algebras – of type I or not!

First, let us take a closer look at what arises from a system with infinitely many degrees of freedom – in the fermion case. As shown by Jordan and Wigner, one obtains what, as a C^* -algebra, is very easy to describe, namely, just the infinite tensor product in the category of unital C^* -algebras of copies of the algebra of 2×2 matrices over the complex numbers. As it happens, in work earlier than that referred to above, Glimm had considered such infinite tensor product C^* -algebras, also allowing the components to be matrix algebras of order different from two. This raised a problem of classification – for those C^* -algebras, all of which were simple and not of type I. (The only simple unital C^* -algebra of type I is a single matrix algebra, or a finite tensor product of matrix algebras!)

In a pioneering classification paper (the first paper on the classification of C^* -algebras being perhaps that of Gelfand and Naimark, in which the commutative case was described), Glimm obtained the classification of infinite tensor products of matrix algebras, showing that it was a direct extension of the classification of finite tensor products, i.e., just of the matrix algebras themselves. As described later by Dixmier, Glimm's classification was as follows. Given a sequence n_1, n_2, \dots of natural numbers (equal to one or more), form the infinite product in a natural way – just by keeping track of the total number of times each prime number appears in the

finite products $n_1 \dots n_k$ (a multiplicity which may be either finite or infinite). Call such a formal infinite product a generalized integer – or, perhaps, a supernatural number! Two (countably) infinite tensor products of matrix algebras are isomorphic (just as in the finite tensor product case) if and only if the corresponding supernatural numbers are equal.

In formulating Glimm's classification of infinite tensor products of matrix algebras in this way, Dixmier pointed out that each supernatural number determines a subgroup of the rational numbers (those with denominator dividing the supernatural number) and that every subgroup of the rational numbers containing the integers arises in this way. He then gave an alternative derivation of Glimm's theorem by recovering this subgroup of the rational numbers as a natural invariant of the algebra, namely, as the subgroup generated by the values on projections of the unique normalized trace. (By a trace is meant here a unitarily invariant positive linear functional.) This could even be interpreted as an alternative statement of Glimm's theorem.

Soon afterwards, Bratteli considered an extension of Glimm's class of C*-algebras, namely, the inductive limits of arbitrary sequences of finite-dimensional C*-algebras, and gave a classification of these algebras in terms of the embedding multiplicity data in the sequences. This was exactly analogous to the original classification of Glimm, but now vastly more complex, with the multiplicity data of the sequence encoded in what is now called a Bratteli diagram. (Note that a finite-dimensional C*-algebra is just a direct sum of matrix algebras over the complex numbers.) Bratteli diagrams have proved to be very important, and in particular have been shown by Putnam and others to be useful for the study of minimal homeomorphisms of the Cantor set.

Bratteli's extension of Glimm's tensor product classification was followed by a corresponding extension by the present author of Dixmier's approach to Glimm's result. It was no longer possible to express the appropriate data in terms of traces (even in the case of a unique normalized trace). Instead, the present author recalled the concept of equivalence of projections introduced by Murray and von Neumann forty years earlier, together with the fact, proved by Murray and von Neumann, that equivalence is compatible with addition of orthogonal projections. (Two projections in a *-algebra are equivalent if they are equal to x^*x and xx^* for some element x .) The resulting elementary invariant – the set of equivalence classes of projections with the operation of addition whenever defined (whenever the equivalence classes

to be added have orthogonal representatives) – one might refer to this as a local abelian semigroup – which was used by Murray and von Neumann to divide von Neumann algebras into what they called types I, II, and III – was shown by the author to determine Bratteli's algebras up to isomorphism.

Bratteli called his algebras approximately finite-dimensional C*-algebras, or AF algebras. The author referred to his invariant simply as the range of the (abstract) dimension, and pointed out that this structure determined an enveloping ordered abelian group, which he called the dimension group. It was soon noticed that the dimension group was related to the K-group introduced by Grothendieck in algebraic geometry (see K-Theory), and by Atiyah and Hirzebruch (see K-Theory) in topology.

Grothendieck's K-group was defined for an arbitrary ring with unit, and Atiyah and Hirzebruch in effect considered the special case of the ring of continuous functions on a compact Hausdorff space – in other words, a commutative C*-algebra – in the process showing that the deep phenomenon of Bott periodicity could be expressed in terms of this invariant. The invariant itself (see below) is essentially the same as that of Murray and von Neumann. In the special case that the ring is an AF algebra, the K-group coincides with the dimension group. (The K-group has a natural ordered, or pre-ordered, structure, although this was often suppressed.)

Let us consider the definition of the K-group of a not necessarily unital C*-algebra; it is in this setting that the statement of Bott periodicity attains its simplest form.

First, in the unital case, one constructs the abelian local semigroup (addition just partially defined) of Murray–von Neumann equivalence classes of projections, as described above in the case of an AF algebra. Let us call this the dimension range. As stated above, for AF algebras this is all that needs to be done – the enveloping group of the dimension range is already the K-group. In the general case, one must repeat the construction for the algebra of 2×2 matrices over the given algebra, with the given algebra considered as embedded as the upper left-hand corner of the matrix algebra. The dimension range of the given algebra then maps naturally into (but not necessarily onto) the dimension range of the matrix algebra. One should then repeat this construction, doubling the order of the matrix algebra at every stage (or, alternatively, increasing it just by one). The enveloping group of the (algebraic) inductive limit of this sequence of local semigroups is then the K-group of the given algebra. (Alternatively, one may just consider immediately the *-algebra of all infinite matrices over the given

C*-algebra with only finitely many nonzero entries, and form the dimension range of this *-algebra – and the enveloping group of this abelian local semi-group, now in fact a semigroup.)

In the case of a nonunital C*-algebra, one adjoins a unit (as may be done, for instance, by representing the C*-algebra faithfully on a Hilbert space, and showing that the C*-algebra obtained by adjoining the identity operator is independent of the representation – actually, one need only check that the *-algebra structure is unique, as the C*-algebra norm on a C*-algebra is always determined by the *-algebra structure). The K-group of the resulting unital C*-algebra then maps naturally into the K-group of the natural one-dimensional quotient, and the kernel of this map is, for reasons that will become clearer later, defined to be the K-group of the nonunital algebra.

Atiyah and Hirzebruch in fact referred to the K-group of the C*-algebra as K_0 – the reason being that there is another very natural group to consider, namely, the K-group of the suspension of the C*-algebra. (The suspension, SA , of a C*-algebra A is defined as the C*-algebra of all continuous functions from the real line \mathbb{R} into A which converge to zero at $\pm\infty$, with the pointwise *-algebra operations and the supremum norm. It may also be defined as the (unique) C*-algebra tensor product $A \otimes C_0(\mathbb{R})$, where $C_0(\mathbb{R})$ denotes the suspension of the C*-algebra \mathbb{C} of complex numbers.) Denoting the K_0 -group of the suspension of a given C*-algebra by K_1 , one might expect this process to continue, but in fact it is periodic ($K_0, K_1, K_0, K_1, \dots$). Bott periodicity states that there is a natural isomorphism of K_2 with K_0 . (C*-algebras can also be defined with the field of real numbers as scalars, and in this case the period of Bott periodicity is eight.)

Another way of stating Bott periodicity, or, more precisely, of embedding it into the K-theory of C*-algebras, is as follows. Given a short exact sequence of C*-algebras,

$$0 \rightarrow J \rightarrow A \rightarrow A/J \rightarrow 0 \quad [3]$$

i.e., given a C*-algebra A and a closed two-sided ideal J (the quotient *-algebra is then a C*-algebra with the quotient norm) – A is sometimes referred to as an extension of J by A/J – consider the natural short (not necessarily exact) sequences

$$K_0(J) \rightarrow K_0(A) \rightarrow K_0(A/J) \quad [4]$$

and

$$K_1(J) \rightarrow K_1(A) \rightarrow K_1(A/J) \quad [5]$$

(K_0 and K_1 are functors!). There exist natural connecting maps $K_1(A/J) \rightarrow K_0(J)$ and $K_0(A/J) \rightarrow K_1(J)$ – the

first referred to as the index map, and the second (sometimes referred to as the odd-order index map) obtained from this immediately from Bott periodicity (as stated above) – such that the periodic six-term sequence

$$\begin{array}{ccccc} K_0(J) & \rightarrow & K_0(A) & \rightarrow & K_0(A/J) \\ & & \uparrow & & \downarrow \\ K_1(A/J) & \leftarrow & K_1(A) & \leftarrow & K_1(J) \end{array}$$

is exact. (The periodicity stated above can also be recovered from this.)

Given that the functor K_0 classifies AF algebras, one might expect the functor K_1 to be useful for classification purposes also. In fact, this is the case. (Indeed, as shown by Brown, the K_1 -functor is already important for the theory of AF algebras – in spite of, or even because of (!), the fact that the K_1 -group of an AF algebra is zero.) Using the six-term exact sequence of Bott periodicity described above, corresponding to an extension of C*-algebras, together with results of the present author, Brown showed that any extension of one AF algebra by another is again an AF algebra.

A rather large class of simple unital C*-algebras has by now been classified by means of the invariants K_0 and K_1 – together with the class of the unit in K_0 , and the order (or pre-order) structure on K_0 – and also taking into account the compact convex set of tracial states on the C*-algebra (a positive linear functional on a C*-algebra is called a trace if it has the same value on x^*x and xx^* for every element x , and a tracial state if it is a state, that is, has norm 1, or has value 1 on the unit in the case the algebra has a unit). In addition to the set of tracial states, together with its natural topology and convex structure, one should also keep track of the natural pairing between traces and K_0 (any trace on a unital C*-algebra has the same value on two equivalent projections – equal to x^*x and xx^* for some element x – and hence gives rise to an additive real-valued functional on K_0).

In terms of these invariants (which might, broadly speaking, be called K-theoretical), it has been possible to classify the simple unital C*-algebras (not of type I) arising as inductive limits (i.e., as the completions of increasing unions) of sequences of finite direct sums of matrix algebras over separable commutative C*-algebras, these assumed to have spectra of dimension at most three, on the one hand (work of the present author together with Guihua Gong and Liangqing Li, a culmination of earlier work of these authors together with a number of others), and, on the other hand, it has been possible (work of Kirchberg and Phillips, also based on earlier work by a number of authors) to classify the

C*-algebra tensor products (in a natural sense) of these C*-algebras with what is called the Cuntz C*-algebra O_∞ (see below). In the first of these two cases, the compact convex set of tracial states – always a Choquet simplex – is an arbitrary (metrizable) such space.

In the second case, this space is empty (as it is for O_∞ in particular). In both cases, K_0 and K_1 are arbitrary countable abelian groups, with the proviso that K_0 is not the sum of a torsion group and a cyclic group. In the first case, the order structure on K_0 , the class of the unit element, and the pairing of K_0 with the space of traces have certain special properties; as it turns out, these can be expressed in a simple way. (The class of the unit need only be positive and nonzero.) In the second case, the order structure on K_0 is degenerate – every element is positive – and the class of the unit can be arbitrary (including zero!).

Let us just note that the Cuntz C*-algebra O_∞ is the unital C*-algebra generated by an infinite sequence s_1, s_2, \dots of isometries with orthogonal ranges (in other words, elements s_i such that $s_i^* s_i$ is the unit and $s_j^* s_i = 0$ if $j \neq i$). One need not require the C*-algebra to have the universal property with respect to these generators and relations as it is in fact unique (up to an isomorphism preserving these generators). In particular, this C*-algebra is simple. (If one considers a finite sequence of isometries with orthogonal ranges, and assumes in addition that the sum of these is the unit, one also obtains a simple C*-algebra, the Cuntz C*-algebra O_n , $n=2, 3, \dots$). The K_0 -group and K_1 -group of O_∞ are, respectively, \mathbb{Z} and 0. (The K_0 -group and K_1 -groups of O_n for $n=2, 3, \dots$ are, respectively, $\mathbb{Z}/(n-1)\mathbb{Z}$ and 0.)

Both classes of C*-algebras considered in the classification result stated above, although described in rather a concrete way (in terms of inductive limits and tensor products), can also be characterized axiomatically, in a way that makes it clear that they are, in fact, much more general than they seem. (These axiomatizations are due to Lin and to Kirchberg and Phillips. Typically, the abstract axioms are easier to establish in a given case than the inductive limit form described above.)

In view of this, and the fact that one of the axioms is a notion of amenability (the analogous property for C*-algebras of a notion that has also been considered for von Neumann algebras) and since amenable von Neumann algebras (on a separable Hilbert space) have been classified completely (in remarkable work of Connes, together with many others, starting with Murray and von Neumann – and, one must also mention, ending with Haagerup,

who settled a particularly stubborn case), it is natural to ask whether the K -theoretical invariants described above might be sufficient to classify all amenable separable C*-algebras, say, those which are simple and unital.

The work of Villadsen has shown that additional invariants must in fact be considered, if one is to deal with arbitrary amenable simple C*-algebras, and this has been confirmed in subsequent work of Rørdam and of Toms. (Villadsen's examples were obtained by removing the condition of low dimension on the spectra of the commutative C*-algebras appearing in the inductive limit decomposition considered above.) The very nature of these authors' work, however, has been to introduce additional invariants, all of which it seems natural to consider as, broadly speaking, K -theoretical. (And all of which, as it happens, are already familiar.)

The question of the classifiability, in terms of simple invariants (K -theoretical in nature, at least in the broad sense, and including the spectrum which is indispensable in the nonsimple case), of all (separable) amenable C*-algebras would therefore still appear to be on the agenda.

Already, in any case, just like the analogous question for von Neumann algebras (now settled), this question would appear to have had a noticeable influence on the development of the subject – not least in underlining the importance of K -theoretical methods, which have proved to be pertinent both in connection with the index theory of differential operators on geometrical structures – from foliations to fractals – and in connection with questions in physics, related to quantum statistical mechanics (see e.g., Quantum Hall Effect), to quantum field theory (e.g., the standard model), and even to string theory and M -theory.

See also: Axiomatic Quantum Field Theory; Bosons and Fermions in External Fields; The Jones Polynomial; K -Theory; Positive Maps on C*-Algebras; Quantum Hall Effect; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; von Neumann Algebras: Subfactor Theory.

Further Reading

- Davidson KR (1996) *C*-Algebras by Example*. Fields Institute Monographs, 6. Providence, RI: American Mathematical Society.
- Dixmier J (1969) *Les C*-Algèbres et leurs Représentations*, 2nd edn. Paris: Gauthier-Villars.
- Elliott GA (1995) The classification problem for amenable C*-algebras. In: Chatterji SD (ed.) *Proceedings of the International Congress of Mathematicians*, vols. 1, 2, pp. 922–932. (Zürich, 1994). Basel: Birkhäuser.

Evans DE and Kawahigashi Y (1998) *Quantum Symmetries on Operator Algebras*. Oxford: Oxford University Press.
 Fillmore PA (1996) *A User's Guide to Operator Algebras*. New York: Wiley.
 Kadison RV and Ringrose J (1983–92) *Fundamentals of the Theory of Operator Algebras* (4 volumes). New York: Academic Press.
 Lin H (2001) *An Introduction to the Classification of Amenable C^* -Algebras*. Singapore: World Scientific.

Pedersen GK (1979) *C^* -Algebras and their Automorphism Groups*, London Math. Soc. Monographs. London: Academic Press.
 Rørdam M (2002) *Classification of Nuclear, Simple C^* -Algebras*, Encyclopaedia of Mathematical Sciences, vol. 126, pp. 1–145. Berlin: Springer.
 Sakai S (1971) *C^* -Algebras and W^* -Algebras*. Berlin: Springer.

Calibrated Geometry and Special Lagrangian Submanifolds

D D Joyce, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Calibrated Geometry

“Calibrated geometry,” introduced by Harvey and Lawson (1982), is the study of special classes of “minimal submanifolds” N of a Riemannian manifold (M, g) , defined using a closed form φ on M called a calibration. For example, if (M, J, g) is a Kähler manifold with Kähler form ω , then complex k -submanifolds of M are calibrated with respect to $\varphi = \omega^k/k!$. Another important class of calibrated submanifolds are special Lagrangian submanifolds in Calabi–Yau manifolds, which is the focus of the section “Special Lagrangian geometry.”

Calibrations and Calibrated Submanifolds

We begin by defining “calibrations” and “calibrated submanifolds.”

Definition 1 Let (M, g) be a Riemannian manifold. An “oriented tangent k -plane” V on M is a vector subspace V of some tangent space $T_x M$ to M with $\dim V = k$, equipped with an orientation. If V is an oriented tangent k -plane on M then $g|_V$ is a Euclidean metric on V ; so, combining $g|_V$ with the orientation on V gives a natural volume form vol_V on V , which is a k -form on V .

Now let φ be a closed k -form on M . φ is said to be a calibration on M , if for every oriented k -plane V on M , $\varphi|_V \leq \text{vol}_V$. Here, $\varphi|_V = \alpha \cdot \text{vol}_V$ for some $\alpha \in \mathbb{R}$, and $\varphi|_V \leq \text{vol}_V$ if $\alpha \leq 1$. Let N be an oriented submanifold of M with dimension k . Then each tangent space $T_x N$ for $x \in N$ is an oriented tangent k -plane. We say that N is a calibrated submanifold if $\varphi|_{T_x N} = \text{vol}_{T_x N}$ for all $x \in N$.

It is easy to show that calibrated submanifolds are automatically “minimal submanifolds.” We prove this in the compact case, but noncompact calibrated submanifolds are locally volume-minimizing as well.

Proposition 2 Let (M, g) be a Riemannian manifold, φ a calibration on M , and N a compact φ -submanifold in M . Then N is volume-minimizing in its homology class.

Proof Let $\dim N = k$, and let $[N] \in H_k(M, \mathbb{R})$ and $[\varphi] \in H^k(M, \mathbb{R})$ be the homology and cohomology classes of N and φ . Then

$$[\varphi] \cdot [N] = \int_{x \in N} \varphi|_{T_x N} = \int_{x \in N} \text{vol}_{T_x N} = \text{Vol}(N)$$

since $\varphi|_{T_x N} = \text{vol}_{T_x N}$ for each $x \in N$, as N is a calibrated submanifold. If N' is any other compact k -submanifold of M with $[N'] = [N]$ in $H_k(M, \mathbb{R})$, then

$$\begin{aligned} [\varphi] \cdot [N] &= [\varphi] \cdot [N'] = \int_{x \in N'} \varphi|_{T_x N'} \leq \int_{x \in N'} \text{vol}_{T_x N'} \\ &= \text{Vol}(N') \end{aligned}$$

since $\varphi|_{T_x N'} \leq \text{vol}_{T_x N'}$ because φ is a calibration. The last two equations give $\text{Vol}(N) \leq \text{Vol}(N')$. Thus, N is volume-minimizing in its homology class. \square

Now let (M, g) be a Riemannian manifold with a calibration φ , and let $\iota: N \rightarrow M$ be an immersed submanifold. Whether N is a φ -submanifold depends upon the tangent spaces of N . That is, it depends on ι and its first derivative. So, for N to be calibrated with respect to φ is a first-order partial differential equation on ι . But if N is calibrated then N is minimal, and for N to be minimal is a second-order partial differential equation on ι .

One moral is that the calibrated equations, being first order, are often easier to solve than the minimal submanifold equations, which are second order. So calibrated geometry is a fertile source of examples of minimal submanifolds.

Calibrated Submanifolds and Special Holonomy

A calibration φ on (M, g) is only interesting if there exist plenty of φ -submanifolds N in M , locally or globally. Since $\varphi|_{T_x N} = \text{vol}_{T_x N}$ for each $x \in N$, φ -submanifolds will be abundant only if the family \mathcal{F}_φ of calibrated tangent k -planes V with $\varphi|_V = \text{vol}_V$

is “reasonably large” – say, if \mathcal{F}_φ has small codimension in the family of all tangent k -planes V on M . A maximally boring example is the k -form $\varphi=0$, which is a calibration but has no calibrated tangent k -planes, so no φ -submanifolds.

Thus, most calibrations φ will have few or no φ -submanifolds, and only special calibrations φ with \mathcal{F}_φ large will have interesting calibrated geometries. Now the field of Riemannian holonomy groups is a natural companion for calibrated geometry, because it gives a simple way to generate interesting calibrations φ which automatically have \mathcal{F}_φ large.

Let $G \subset O(n)$ be a possible holonomy group of a Riemannian metric. In particular, we can take G to be one of the holonomy groups $U(m)$, $SU(m)$, $Sp(m)$, G_2 , or $Spin(7)$ from Berger’s classification. Then G acts on the k -forms $\Lambda^k(\mathbb{R}^n)^*$ on \mathbb{R}^n , so we can look for G -invariant k -forms on \mathbb{R}^n . Suppose φ_0 is a nonzero, G -invariant k -form on \mathbb{R}^n .

By rescaling φ_0 we can arrange that for each oriented k -plane $U \subset \mathbb{R}^n$, we have $\varphi_0|_U \leq \text{vol}_U$, and that $\varphi_0|_U = \text{vol}_U$ for at least one such U . Let H be the stabilizer subgroup of this U in G . Then $\varphi_0|_{\gamma \cdot U} = \text{vol}_{\gamma \cdot U}$ by G -invariance, so $\gamma \cdot U$ is a calibrated k -plane for all $\gamma \in G$. Thus, the family \mathcal{F}_0 of φ_0 -calibrated k -planes in \mathbb{R}^n contains G/H , so it is “reasonably large,” and it is likely that the calibrated submanifolds will have an interesting geometry.

Now let M be a manifold of dimension n , and g a metric on M with Levi-Civita connection ∇ and holonomy group G . Then there is a k -form φ on M with $\nabla\varphi=0$, corresponding to φ_0 . Hence $d\varphi=0$, and φ is closed. Also, the condition $\varphi_0|_U \leq \text{vol}_U$ for all oriented k -planes U in \mathbb{R}^n implies that $\varphi|_V \leq \text{vol}_V$ for all oriented tangent k -planes V in M . Thus, φ is a calibration on M . The family \mathcal{F}_φ of calibrated tangent k -planes on M fibers over M with fiber \mathcal{F}_0 ; so, it is “reasonably large.”

This gives a general method for finding interesting calibrations on manifolds with reduced holonomy. Here are the most significant examples.

- Let $G=U(m) \subset O(2m)$. Then G preserves a 2-form ω_0 on \mathbb{R}^{2m} . If g is a metric on M with holonomy $U(m)$, then g is Kähler with complex structure J , and the 2-form ω on M associated to ω_0 is the Kähler form of g .

One can show that ω is a calibration on (M, g) , and the calibrated submanifolds are exactly the “holomorphic curves” in (M, J) . More generally, $\omega^k/k!$ is a calibration on M for $1 \leq k \leq m$, and the corresponding calibrated submanifolds are the complex k -dimensional submanifolds of (M, J) .

- Let $G=SU(m) \subset O(2m)$. Then G preserves a complex volume form $\Omega_0 = dz_1 \wedge \cdots \wedge dz_m$ on

\mathbb{C}^m . Thus, a Calabi–Yau m -fold (M, g) with $\text{Hol}(g) = SU(m)$ has a holomorphic volume form Ω . The real part $\text{Re}\Omega$ is a calibration on M , and the corresponding calibrated submanifolds are called special Lagrangian submanifolds.

- The group $G_2 \subset O(7)$ preserves a 3-form φ_0 and a 4-form $*\varphi_0$ on \mathbb{R}^7 . Thus, a Riemannian 7-manifold (M, g) with holonomy G_2 comes with a 3-form φ and 4-form $*\varphi$, which are both calibrations. The corresponding calibrated submanifolds are called associative 3-folds and coassociative 4-folds.
- The group $Spin(7) \subset O(8)$ preserves a 4-form Ω_0 on \mathbb{R}^8 . Thus a Riemannian 8-manifold (M, g) with holonomy $Spin(7)$ has a 4-form Ω , which is a calibration. The Ω -submanifolds are called Cayley 4-folds.

It is an important general principle that to each calibration φ on an n -manifold (M, g) with special holonomy constructed in this way, there corresponds a constant calibration φ_0 on \mathbb{R}^n . Locally, φ -submanifolds in M resemble the φ_0 -submanifolds in \mathbb{R}^n , and have many of the same properties. Thus, to understand the calibrated submanifolds in a manifold with special holonomy, it is often a good idea to start by studying the corresponding calibrated submanifolds of \mathbb{R}^n .

In particular, singularities of φ -submanifolds in M will be locally modeled on singularities of φ_0 -submanifolds in \mathbb{R}^n . (In the sense of geometric measure theory, the tangent cone at a singular point of a φ -submanifold in M is a conical φ_0 -submanifold in \mathbb{R}^n .) So by studying singular φ_0 -submanifolds in \mathbb{R}^n , we may understand the singular behavior of φ -submanifolds in M .

Special Lagrangian Geometry

We now focus on one class of calibrated submanifolds, special Lagrangian submanifolds in Calabi–Yau manifolds. Calabi–Yau 3-folds are used to make the spacetime vacuum in string theory, and special Lagrangian 3-folds are the classical versions of A-branes, or supersymmetric 3-cycles, in Calabi–Yau 3-folds. Special Lagrangian geometry aroused great interest amongst string theorists because of its rôle in the SYZ conjecture, providing a geometric basis for “mirror symmetry” of Calabi–Yau 3-folds.

Calabi–Yau Manifolds

Here is our definition of Calabi–Yau manifold. Readers are warned that there are several different definitions of Calabi–Yau manifolds in use in the literature. Ours is unusual in regarding Ω as part of the given structure.

Definition 3 Let $m \geq 2$. A Calabi–Yau m -fold is a quadruple (M, J, g, Ω) such that (M, J) is a compact m -dimensional complex manifold, g a Kähler metric on (M, J) with Kähler form ω , and Ω a holomorphic $(m, 0)$ -form on M called the holomorphic volume form, which satisfies

$$\omega^m/m! = (-1)^{m(m-1)/2} (i/2)^m \Omega \wedge \bar{\Omega} \quad [1]$$

The constant factor in [1] is chosen to make $\text{Re } \Omega$ a calibration. It follows from [1] that g is Ricci-flat, Ω is constant under the Levi-Civita connection, and the holonomy group of g has $\text{Hol}(g) \subseteq \text{SU}(m)$.

Let (M, J) be a compact, complex manifold, and g a Kähler metric on M , with Ricci curvature R_{ab} . Define the Ricci form ρ of g by $\rho_{ac} = J_a^b R_{bc}$. Then ρ is a closed real $(1, 1)$ -form on M , with de Rham cohomology class $[\rho] = 2\pi c_1(M) \in H^2(M, \mathbb{R})$, where $c_1(M)$ is the first Chern class of M in $H^2(M, \mathbb{Z})$. The Calabi conjecture specifies which closed $(1, 1)$ -forms can be the Ricci forms of a Kähler metric on M .

The Calabi conjecture *Let (M, J) be a compact, complex manifold, and g' a Kähler metric on M , with Kähler form ω' . Suppose that ρ is a real, closed $(1, 1)$ -form on M with $[\rho] = 2\pi c_1(M)$. Then there exists a unique Kähler metric g on M with Kähler form ω , such that $[\omega] = [\omega'] \in H^2(M, \mathbb{R})$, and the Ricci form of g is ρ .*

Note that $[\omega] = [\omega']$ says that g and g' are in the same Kähler class. The conjecture was posed by Calabi in 1954, and was eventually proved by Yau in 1976. Its importance to us is that when the canonical bundle K_M is trivial, so that $c_1(M) = 0$, we can take $\rho \equiv 0$, and then g is Ricci-flat. Since K_M is trivial, it has a nonzero holomorphic section, a holomorphic $(m, 0)$ -form Ω . As g is Ricci-flat, it follows that $\nabla \Omega = 0$, where ∇ is the Levi-Civita connection of g . Rescaling Ω by a complex constant makes [1] hold, and then (M, J, g, Ω) is a Calabi–Yau m -fold. This proves:

Theorem 4 *Let (M, J) be a compact complex m -manifold with K_M trivial. Then every Kähler class on M contains a unique Ricci-flat Kähler metric g . There exists a holomorphic $(m, 0)$ -form Ω , unique up to change of phase $\Omega \mapsto e^{i\theta} \Omega$, such that (M, J, g, Ω) is a Calabi–Yau m -fold.*

Using algebraic geometry, one can produce many examples of complex m -folds (M, J) satisfying these conditions, such as the Fermat $(m + 2)$ -tic

$$\{[z_0, \dots, z_{m+1}] \in \mathbb{C}\mathbb{P}^{m+1} : z_0^{m+2} + \dots + z_{m+1}^{m+2} = 0\} \quad [2]$$

Therefore, Calabi–Yau m -folds are very abundant.

Special Lagrangian Submanifolds

Definition 5 Let (M, J, g, Ω) be a Calabi–Yau m -fold. Then $\text{Re } \Omega$ is a calibration on the Riemannian manifold (M, g) . An oriented real m -dimensional submanifold N in M is called a special Lagrangian submanifold (SL m -fold) if it is calibrated with respect to $\text{Re } \Omega$.

Here is an alternative definition of SL m -folds. It is often more useful than Definition 5.

Proposition 6 *Let (M, J, g, Ω) be a Calabi–Yau m -fold, with Kähler form ω , and N a real m -dimensional submanifold in M . Then N admits an orientation making it into an SL m -fold in M if and only if $\omega|_N \equiv 0$ and $\text{Im } \Omega|_N \equiv 0$.*

Regard N as an immersed submanifold, with immersion $\iota : N \rightarrow M$. Then $[\omega|_N]$ and $[\text{Im } \Omega|_N]$ are unchanged under continuous variations of the immersion ι . Thus, $[\omega|_N] = [\text{Im } \Omega|_N] = 0$ is a necessary condition not just for N to be special Lagrangian, but also for any isotopic submanifold N' in M to be special Lagrangian. This proves:

Corollary 7 *Let (M, J, g, Ω) be a Calabi–Yau m -fold, and N a compact real m -submanifold in M . Then a necessary condition for N to be isotopic to a special Lagrangian submanifold N' in M is that $[\omega|_N] = 0$ in $H^2(N, \mathbb{R})$ and $[\text{Im } \Omega|_N] = 0$ in $H^m(N, \mathbb{R})$.*

Deformations of Compact SL m -Folds

The deformation theory of compact special Lagrangian manifolds was studied by McLean (1998), who proved the following result:

Theorem 8 *Let (M, J, g, Ω) be a Calabi–Yau m -fold, and N a compact special Lagrangian m -fold in M . Then the moduli space \mathcal{M}_N of special Lagrangian deformations of N is a smooth manifold of dimension $b^1(N)$, the first Betti number of N .*

Sketch proof. Suppose for simplicity that N is an embedded submanifold. There is a natural orthogonal decomposition $TM|_N = TN \oplus \nu$, where $\nu \rightarrow N$ is the normal bundle of N in M . As N is Lagrangian, the complex structure $J : TM \rightarrow TM$ gives an isomorphism $J : \nu \rightarrow TN$. But the metric g gives an isomorphism $TN \cong T^*N$. Composing these two gives an isomorphism $\nu \cong T^*N$.

Let T be a small tubular neighborhood of N in M . Then we can identify T with a neighborhood of the zero section in ν . Using the isomorphism $\nu \cong T^*N$, we have an identification between T and a neighborhood of the zero section in T^*N . This can be chosen to identify the Kähler form ω on T with the natural symplectic

structure on T^*N . Let $\pi: T \rightarrow N$ be the obvious projection.

Under this identification, submanifolds N' in $T \subset M$ which are C^1 close to N are identified with the graphs of small smooth sections α of T^*N . That is, submanifolds N' of M close to N are identified with 1-forms α on N . We need to know: which 1-forms α are identified with SL m -folds N' ?

Now, N' is special Lagrangian if $\omega|_{N'} \equiv \text{Im } \Omega|_{N'} \equiv 0$. But $\pi|_{N'}: N' \rightarrow N$ is a diffeomorphism, so we can push $\omega|_{N'}$ and $\text{Im } \Omega|_{N'}$ down to N , and regard them as functions of α . Calculation shows that

$$\pi_*(\omega|_{N'}) = d\alpha \quad \text{and} \quad \pi_*(\text{Im } \Omega|_{N'}) = F(\alpha, \nabla\alpha)$$

where F is a nonlinear function of its arguments. Thus, the moduli space \mathcal{M}_N is locally isomorphic to the set of small 1-forms α on N such that $d\alpha \equiv 0$ and $F(\alpha, \nabla\alpha) \equiv 0$.

Now it turns out that F satisfies $F(\alpha, \nabla\alpha) \approx d(*\alpha)$ when α is small. Therefore, \mathcal{M}_N is locally approximately isomorphic to the vector space of 1-forms α with $d\alpha = d(*\alpha) = 0$. But by Hodge theory, this is isomorphic to the de Rham cohomology group $H^1(N, \mathbb{R})$, and is a manifold with dimension $b^1(N)$.

To carry out this last step rigorously requires some technical machinery: one must work with certain Banach spaces of sections of T^*N , $\Lambda^2 T^*N$ and $\Lambda^m T^*N$, use elliptic regularity results to prove that the map $\alpha \mapsto (d\alpha, F(\alpha, \nabla\alpha))$ has closed image in these Banach spaces, and then use the implicit function theorem for Banach spaces to show that the kernel of the map is what is expected.

Obstructions to Existence of Compact SL m -Folds

Let $\{(M, J_t, g_t, \Omega_t) : t \in (-\epsilon, \epsilon)\}$ be a smooth one-parameter family of Calabi–Yau m -folds. Suppose N_0 is an SL m -fold in (M, J_0, g_0, Ω_0) . When can we extend N_0 to a smooth family of SL m -folds N_t in (M, J_t, g_t, Ω_t) for $t \in (-\epsilon, \epsilon)$?

By Corollary 7, a necessary condition is that $[\omega_t|_{N_0}] = [\text{Im } \Omega_t|_{N_0}] = 0$ for all t . Our next result shows that locally, this is also a sufficient condition.

Theorem 9 *Let $\{(M, J_t, g_t, \Omega_t) : t \in (-\epsilon, \epsilon)\}$ be a smooth one-parameter family of Calabi–Yau m -folds, with Kähler forms ω_t . Let N_0 be a compact SL m -fold in (M, J_0, g_0, Ω_0) , and suppose that $[\omega_t|_{N_0}] = 0$ in $H^2(N_0, \mathbb{R})$ and $[\text{Im } \Omega_t|_{N_0}] = 0$ in $H^m(N_0, \mathbb{R})$ for all $t \in (-\epsilon, \epsilon)$. Then N_0 extends to a smooth one-parameter family $\{N_t : t \in (-\delta, \delta)\}$, where $0 < \delta \leq \epsilon$ and N_t is a compact SL m -fold in (M, J_t, g_t, Ω_t) .*

This can be proved using similar techniques to **Theorem 8**. Note that the condition $[\text{Im } \Omega_t|_{N_0}] = 0$

for all t can be satisfied by choosing the phases of the Ω_t appropriately, and if the image of $H_2(N, \mathbb{Z})$ in $H_2(M, \mathbb{R})$ is zero, then the condition $[\omega|_N] = 0$ holds automatically.

Thus, the obstructions $[\omega_t|_{N_0}] = [\text{Im } \Omega_t|_{N_0}] = 0$ in **Theorem 9** are actually fairly mild restrictions, and SL m -folds should be considered as pretty stable under small deformations of the Calabi–Yau structure.

Remark The deformation and obstruction theory of compact SL m -folds are extremely well behaved compared to many other moduli space problems in differential geometry. In other geometric problems (such as the deformations of complex structures on a complex manifold, or pseudoholomorphic curves in an almost-complex manifold, or instantons on a Riemannian 4-manifold), the deformation theory often has the following general structure.

There are vector bundles E, F over a compact manifold M , and an elliptic operator $P: C^\infty(E) \rightarrow C^\infty(F)$, usually first order. The kernel $\text{Ker } P$ is the set of infinitesimal deformations, and the cokernel $\text{Coker } P$ the set of obstructions. The actual moduli space \mathcal{M} is locally the zeros of a nonlinear map $\Psi: \text{Ker } P \rightarrow \text{Coker } P$.

In a generic case, $\text{Coker } P = 0$, and then the moduli space \mathcal{M} is locally isomorphic to $\text{Ker } P$, and so is locally a manifold with dimension $\text{ind}(P)$. However, in nongeneric situations $\text{Coker } P$ may be nonzero, and then the moduli space \mathcal{M} may be nonsingular, or have an unexpected dimension.

However, SL m -folds do not follow this pattern. Instead, the obstructions are topologically determined, and the moduli space is always smooth, with dimension given by a topological formula. This should be regarded as a minor mathematical miracle.

Mirror Symmetry and the SYZ Conjecture

Mirror symmetry is a mysterious relationship between pairs of Calabi–Yau 3-folds M, \hat{M} , arising from a branch of physics known as string theory, and leading to some very strange and exciting conjectures about Calabi–Yau 3-folds, many of which have been proved in special cases.

In the beginning (the 1980s), mirror symmetry seemed mathematically completely mysterious. But there are now two complementary conjectural theories, due to Kontsevich and Strominger–Yau–Zaslow, which explain mirror symmetry in a fairly mathematical way. Probably both are true, at some level. The second proposal, due to Strominger, Yau, and Zaslow (1996), is known as the SYZ conjecture. Here is an attempt to state it.

The SYZ conjecture Suppose M and \hat{M} are mirror Calabi–Yau 3-folds. Then (under some additional conditions), there should exist a compact topological 3-manifold B and surjective, continuous maps $f: M \rightarrow B$ and $\hat{f}: \hat{M} \rightarrow B$, such that

- (i) There exists a dense open set $B_0 \subset B$, such that for each $b \in B_0$, the fibers $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are nonsingular special Lagrangian 3-tori T^3 in M and \hat{M} . Furthermore, $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are in some sense dual to one another.
- (ii) For each $b \in \Delta = B \setminus B_0$, the fibers $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are expected to be singular special Lagrangian 3-folds in M and \hat{M} .

The fibrations f and \hat{f} are called special Lagrangian fibrations, and the set of singular fibers Δ is called the discriminant. In part (i), the nonsingular fibers of f and \hat{f} are supposed to be dual tori. What does this mean?

On the topological level, we can define duality between two tori T, \hat{T} to be a choice of isomorphism $H^1(T, \mathbb{Z}) \cong H_1(\hat{T}, \mathbb{Z})$. We can also define duality between tori equipped with flat Riemannian metrics. Write $T = V/\Lambda$, where V is a Euclidean vector space and Λ a lattice in V . Then the dual torus \hat{T} is defined to be V^*/Λ^* , where V^* is the dual vector space and Λ^* the dual lattice. However, there is no notion of duality between nonflat metrics on dual tori.

Strominger, Yau, and Zaslow argue only that their conjecture holds when M, \hat{M} are close to the “large complex structure limit.” In this case, the diameters of the fibers $f^{-1}(b), \hat{f}^{-1}(b)$ are expected to be small compared to the diameter of the base space B , and away from singularities of f, \hat{f} , the metrics on the nonsingular fibers are expected to be approximately flat. So, part (i) of the SYZ conjecture says that for $b \in B \setminus B_0$, $f^{-1}(b)$ is approximately a flat Riemannian 3-torus, and $\hat{f}^{-1}(b)$ is approximately the dual flat Riemannian torus.

Mathematical research on the SYZ conjecture has followed two broad approaches. The first could be described as symplectic topological. For this, we treat M, \hat{M} just as symplectic manifolds and f, \hat{f} just as Lagrangian fibrations. We also suppose B is a smooth 3-manifold and f, \hat{f} are smooth maps. Under these simplifying assumptions, Mark Gross, Wei-Dong Ruan, and others have built up a beautiful, detailed picture of how dual SYZ fibrations work at the global topological level.

The second approach could be described as local geometric. Here, we try to take the special Lagrangian condition seriously from the outset, and focus on the local behavior of special Lagrangian

submanifolds, and especially their singularities, rather than on global topological questions. In addition, we are interested in what fibrations of generic Calabi–Yau 3-folds might look like.

There is now a well-developed theory of SL m -folds with isolated singularities modeled on cones (Joyce 2003a). This is applied to SL fibrations and the SYZ conjecture in Joyce (2003a, b), leading to the tentative conclusions that for generic Calabi–Yau 3-folds M , special Lagrangian fibrations $f: M \rightarrow B$ will be only piecewise smooth, and have discriminants Δ of real codimension 1 in B , in contrast to smooth fibrations which have Δ of codimension 2. We also argue that for generic mirrors M, \hat{M} and f, \hat{f} , the discriminants $\Delta, \hat{\Delta}$ cannot be homeomorphic and so do not coincide. This contradicts part (ii) above.

A better way to formulate the SYZ conjecture may be in terms of families of mirror Calabi–Yau 3-folds M_t, \hat{M}_t and fibrations $f_t: M_t \rightarrow B, \hat{f}_t: \hat{M}_t \rightarrow B$ for $t \in (0, \epsilon)$ which approach the “large complex structure limit” as $t \rightarrow 0$. Then we could require the discriminants $\Delta_t, \hat{\Delta}_t$ of f_t, \hat{f}_t to converge to some common, codimension 2 limit Δ_0 as $t \rightarrow 0$.

It is an important, and difficult, open problem to construct examples of special Lagrangian fibrations of compact, holonomy $SU(3)$ Calabi–Yau 3-folds. None are currently known.

See also: Minimal submanifolds; Mirror Symmetry; A Geometric Survey; Moduli Spaces: An Introduction; Riemannian Holonomy Groups and Exceptional Holonomy.

Further Reading

- Gross M, Huybrechts D, and Joyce D (2003) *Calabi–Yau Manifolds and Related Geometries*, Universitext Series, Berlin: Springer.
- Harvey R and Lawson HB (1982) Calibrated geometries. *Acta Mathematica* 148: 47–157.
- Joyce DD (2000) *Compact Manifolds with Special Holonomy*. Oxford: Oxford University Press.
- Joyce DD (2003a) Special Lagrangian submanifolds with isolated conical singularities. V. Survey and applications. *Journal of Differential Geometry* 63: 279–347, math.DG/0303272.
- Joyce DD (2003b) Singularities of special Lagrangian fibrations and the SYZ conjecture. *Communications in Analysis and Geometry* 11: 859–907, math.DG/0011179.
- Joyce DD (2003c) $U(1)$ -invariant special Lagrangian 3-folds in \mathbb{C}^3 and special Lagrangian fibrations. *Turkish Mathematical Journal* 27: 99–114, math.DG/0206016.
- McLean RC (1998) Deformations of calibrated submanifolds. *Communications in Analysis and Geometry* 6: 705–747.
- Strominger A, Yau S-T, and Zaslow E (1996) Mirror symmetry is T-duality. *Nuclear Physics B* 479: 243–259, hep-th/9606040.

Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type

S N M Ruijsenaars, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Systems of Calogero–Moser–Sutherland (CMS) type form a class of finite-dimensional dynamical systems that are integrable both at the classical and at the quantum level. The CMS systems describe N point particles moving on a line or on a ring, interacting via pair potentials that are specific functions of four types, namely rational (I), hyperbolic (II), trigonometric (III), and elliptic (IV). They occur not only in a nonrelativistic (Galilei-invariant), but also in a relativistic (Poincaré-invariant) setting. Thus, one can distinguish a hierarchy of 16 physically distinct versions (classical/quantum, nonrelativistic/relativistic, type I–IV), the most general one being the quantum relativistic type IV system.

The nonrelativistic systems date back to pioneering work by Calogero, Sutherland, and Moser in the early 1970s. The pair potential structure of the interaction can be encoded in the root system A_{N-1} , and there also exist integrable versions for all of the remaining root systems. The classical systems are given by N Poisson commuting Hamiltonians with a polynomial dependence on the particle momenta p_1, \dots, p_N . Accordingly, the quantum versions are described by N commuting Hamiltonians that are partial differential operators.

The relativistic systems were introduced in the mid-1980s, at the classical level by Ruijsenaars and Schneider, and at the quantum level by Ruijsenaars. They converge to the nonrelativistic systems in the limit $c \rightarrow \infty$, where c is the speed of light. Again, the systems can be related to the root system A_{N-1} , and they admit integrable versions for other root systems. All of the commuting classical Hamiltonians depend exponentially on generalized momenta p_1, \dots, p_N . Hence, the associated commuting quantum Hamiltonians are analytic difference operators.

The above integrable systems can be further generalized by allowing supersymmetry or internal degrees of freedom (“spins”), coupled in quite special ways to retain integrability. In this article, however, the focus is on the 16 versions of the A_{N-1} -symmetric CMS systems without internal degrees of freedom. The primary aim is to acquaint the reader with their definition and integrability,

and with their most prominent features and inter-relationships. Second, we intend to give a rough sketch of the state of the art concerning explicit solutions for the various versions. This involves a concretization of the action-angle maps and eigenfunction transforms that simultaneously diagonalize the commuting dynamics, paying special attention to their remarkable duality properties.

It is beyond the scope of this article to review the hundreds of papers specifically dealing with CMS type systems, let alone the much larger literature where they play some role. Indeed, the systems have been encountered in a great many different contexts and they are related to a host of other integrable systems in various ways. Accordingly, they can be studied from the perspective of various subfields of mathematics and theoretical physics. First some of these perspectives and relations to seemingly quite different topics will be mentioned before embarking on the far more focused survey.

Staying first within the confines of the CMS type systems, some nonobvious limits yielding other familiar finite-dimensional integrable systems will be mentioned. To begin with, all of the A_{N-1} type systems give rise to systems with a Toda type (exponential “nearest neighbor”) interaction via a suitable limiting transition (basically a strong-coupling limit). This leads to integrable N -particle systems with a classical/quantum, nonrelativistic/relativistic, nonperiodic/periodic version; starting from the quantum relativistic periodic Toda system, the remaining seven versions can be obtained by suitable limits.

Next, we recall that the quantum system of N nonrelativistic bosons on the line or ring interacting via a pair potential of δ -function type is soluble via a Bethe ansatz, with the “line version” exhibiting quantum soliton behavior (factorized scattering). It has been shown that there exist scaling limits of eigenfunctions for suitable CMS systems that give rise to the latter Bethe type eigenfunctions for $N=2$, while convergence for $N>2$ is plausible, but has not been demonstrated thus far.

Via suitable analytic continuations preserving reality/formal self-adjointness, one can arrive at CMS systems with more than one species of particle (particles and “antiparticles”). Likewise, analytic continuations and appropriate limits of CMS systems associated with root systems other than A_{N-1} lead to a further proliferation of N -dimensional integrable systems. Typically, such limits refer either

to the commuting Hamiltonians (the Toda limit being a case in point) or to the joint eigenfunctions (as exemplified by the δ -function system limit); it seems difficult to control both sets of quantities at once.

Starting from the spin type CMS systems, another kind of limit can be taken. Specifically, by “freezing” the particles at equilibrium positions, it is possible to arrive at integrable spin chains of Haldane–Shastry and Inozemtsev type.

At this point, it is expedient to insert a brief remark on finite-dimensional integrable systems. As the term suggests, one may expect that, with due effort, such systems can be “integrated,” or, equivalently, “solved.” But it should be noted that the latter terms (let alone the qualifier “due effort”) have no unambiguous mathematical meaning. Certainly, “solving” involves obtaining explicit information on the action-angle map and joint eigenfunction transform at the classical and quantum level, resp., but *a priori* it is not at all clear how far one can proceed.

Focusing again on the CMS systems and their relatives, it should be stressed that, in many cases, one is still far removed from a complete solution, especially for the elliptic CMS systems. In this regard the previous remark serves not only as a caveat, but also to make clear why the various vantage points provided by different subfields in mathematics and physics are crucial: typically, they yield complementary insights and distinct representations for solutions, serving different purposes.

To be sure, in first approximation the mathematics involved at the classical and quantum level is symplectic geometry and Hilbert space theory, resp. In point of fact, however, far more ingredients have turned out to be quite natural and useful. On the classical level, these include the theory of groups, Lie algebras and symmetric spaces, linear algebra and spectral theory, Riemann surface theory, and more generally algebraic geometry.

On the quantum level, the viewpoint of harmonic analysis on symmetric spaces is particularly natural and fruitful for the nonrelativistic CMS systems and their arbitrary root-system versions, whereas quantum groups/algebras/symmetric spaces can be tied in with the relativistic systems and their versions for other root systems. (The $c \rightarrow \infty$ limit amounts to the $q \rightarrow 1$ limit in the quantum group picture.) As a matter of fact, the whole area of special functions and their q -analogs is intimately related to the quantum CMS type systems (cf. also the last section of this article). Finally, the occurrence of commuting analytic difference operators in the relativistic ($q \neq 1$) systems leads to largely uncharted territory

in the intersection of the theory of Hilbert space eigenfunction expansions and the theory of linear analytic difference equations.

The study of the thermodynamics ($N \rightarrow \infty$ limit with temperature ≥ 0 and density ≥ 0 fixed) associated with the trigonometric and elliptic CMS systems and their spin cousins yields its own circle of problems. It was initiated by Sutherland three decades ago, and even though a host of results on partition functions, correlation functions, fractional statistics, strong–weak coupling duality, relations to Yangians, etc., have meanwhile been obtained, many questions are still open. This area also has links with random-matrix theory, but the input from this field is thus far limited to certain discrete couplings.

The above N -dimensional integrable systems are related to a great many infinite-dimensional integrable systems, both at the classical and at the quantum level. On the one hand, there are structural analogs that have been used to advantage in the study of CMS systems, including Lax pair and R -matrix formulations, zero-curvature representations, bi-Hamiltonian formalism, Bäcklund transformations, time discretizations, and tools such as Baker–Akhiezer functions, Bethe ansatz, separation of variables, and Baxter-type Q -operators.

On the other hand, there are striking physical similarities between various soliton field theories (a prominent one being the sine-Gordon field theory) and infinite soliton lattices (in particular several Toda type lattices), and the CMS systems for special parameter values. Particularly conspicuous are the ties between the classical CMS systems and the KP and two-dimensional Toda hierarchies. The latter relations actually extend beyond the solitons, including rational and theta function solutions.

CMS systems are relevant in various other contexts not yet mentioned. A prominent one among these is a class of supersymmetric gauge field theories. In this quantum context, the classical CMS systems have surfaced in the description of moduli spaces encoding the vacuum structure (Seiberg–Witten theory). Equally surprising, certain classical CMS systems (with internal degrees of freedom) have found a second application in a quantum context, namely in the description of quantum chaos (level repulsion).

We conclude this introduction by listing additional disparate subjects where connections with CMS type systems have been found. These include the theory of Sklyanin, affine Hecke, Kac–Moody, Virasoro and W -algebras, equations of Knizhnik–Zamolodchikov, Yang–Baxter, Witten–Dijkgraaf–Verlinde–Verlinde, and Painlevé type, Gaudin,

Hitchin, Wess–Zumino, matrix and quasi-exactly solvable models, Dunkl–Cherednik and Polychronakos operators, the quantum Hall effect and quantum transport, two-dimensional Yang–Mills theory, functional equations, integrable mappings, Huygens’ principle, and the bispectral problem.

Classical Nonrelativistic CMS Systems

A system of N nonrelativistic equal-mass m particles on the line interacting via pair potentials can be described by a Hamiltonian

$$H = \frac{1}{2m} \sum_{j=1}^N p_j^2 + \sum_{1 \leq j < k \leq N} V(x_j - x_k), \quad m > 0 \quad [1]$$

The CMS systems are defined by four distinct choices of pair potential. The simplest choice reads

$$V(x) = g^2/mx^2, \quad g > 0 \quad (\text{I}) \quad [2]$$

Hence, the coupling constant g has dimension [action] (the product of [position] and [momentum]). This potential is clearly repulsive. Thus, each initial state in the phase space

$$\Omega = \{(x, p) \in \mathbb{R}^{2N} \mid x \in G\} \quad [3]$$

where G is the configuration space

$$G = \{x \in \mathbb{R}^N \mid x_N < \dots < x_1\} \quad [4]$$

is a scattering state.

The next level is given by the hyperbolic choice

$$V(x) = g^2\nu^2/m \sinh^2(\nu x), \quad \nu > 0 \quad (\text{II}) \quad [5]$$

Hence, ν has dimension [position]⁻¹, and the previous system arises by taking ν to 0. It is clear that [5] yields again a repulsive particle system, so that each state in Ω given by [3] is a scattering state.

The highest level in the hierarchy is the elliptic level, where

$$V(x) = g^2 \wp(x; \omega, \omega')/m, \quad \omega, -i\omega' > 0 \quad (\text{IV}) \quad [6]$$

and $\wp(x; \omega, \omega')$ denotes the Weierstrass \wp -function with periods 2ω and $2\omega'$. It is beyond the scope of this article to elaborate on the elliptic regime, even though it is of considerable interest. It reappears in later sections as the most general regime in which integrability holds true. Indeed, a prominent feature of the elliptic case [6] is that it can be specialized both to the hyperbolic case [5] and to the trigonometric case, given by

$$V(x) = g^2\nu^2/m \sin^2(\nu x) \quad (\text{III}) \quad [7]$$

To obtain the hyperbolic specialization, one should take $\omega' = i\pi/2\nu$ and send ω to ∞ ; then [6]

reduces to [5] (up to an additive constant). Likewise, [7] results from [6] by choosing $\omega = \pi/2\nu$ and taking $-i\omega'$ to ∞ .

The physical picture associated with the trigonometric and elliptic systems is quite different from that of the rational and hyperbolic ones. Of course, the potentials [7] and [6] are again repulsive, but now the internal motion is confined and oscillatory. More specifically, due to energy conservation the phase spaces

$$\begin{aligned} \Omega_{\text{III}} &= G_{\text{III}} \times \mathbb{R}^N, \\ G_{\text{III}} &= \{x_N < \dots < x_1, x_1 - x_N < \pi/\nu\} \end{aligned} \quad [8]$$

$$\begin{aligned} \Omega_{\text{IV}} &= G_{\text{IV}} \times \mathbb{R}^N, \\ G_{\text{IV}} &= \{x_N < \dots < x_1, x_1 - x_N < 2\omega\} \end{aligned} \quad [9]$$

are left invariant by the flow generated by the trigonometric and elliptic N -particle Hamiltonian, resp.

Alternatively, one may interpret the trigonometric Hamiltonian as describing particles constrained to move on a circle and interacting via the inverse square potential [2]. In this picture, the quantities $2\nu x_1, \dots, 2\nu x_N$ are viewed as angular positions on the circle, and one needs a suitable quotient of the phase space [8] by a discrete group action to describe a state of the system.

Turning to integrability aspects, we begin by noting that the total momentum Hamiltonian

$$P = \sum_{j=1}^N p_j \quad [10]$$

obviously Poisson commutes with the above defining Hamiltonians of the systems. For $N = 2$, therefore, integrability is plain. It is possible to write down explicitly the higher commuting Hamiltonians for $N > 2$ as well but, in the nonrelativistic setting, it is more illuminating to characterize them as the power traces or (equivalently) the symmetric functions of a so-called Lax matrix.

The Lax matrix is an $N \times N$ matrix-valued function on the phase space of the system. It plays a pivotal role not only for understanding integrability, but also for setting up an action-angle transformation. The latter issue is discussed again later. Here the more conspicuous features of the Lax matrix will be explained, focusing on the type II system for expository ease. Then one can choose

$$\begin{aligned} L_{jj} &= p_j, \quad L_{jk} = ig\nu/\sinh \nu(x_j - x_k), \\ & j, k = 1, \dots, N, \quad j \neq k \end{aligned} \quad [11]$$

Thus, L is Hermitean and we have

$$\text{tr } L = P, \quad \text{tr } L^2 = 2mH \quad [12]$$

(The rational Lax matrix results from [11] by taking $\nu \rightarrow 0$, and the trigonometric one by taking $\nu \rightarrow i\nu$. The elliptic Lax matrix has a similar structure, but it involves an extra “spectral” parameter.)

Although not obvious, it is true that all of the power traces

$$H_k = \frac{1}{k} \text{tr } L^k, \quad k = 1, \dots, N \quad [13]$$

are in involution (i.e., Poisson commute). One way to understand this involves the so-called Lax pair equation associated with the Hamiltonian flow generated by $H = H_2/m$. This involves a second $N \times N$ matrix function given by

$$\begin{aligned} M_{ji} &= \sum_{l \neq j} \frac{-ig\nu^2}{m \sinh^2 \nu(x_j - x_l)} \\ M_{jk} &= \frac{ig\nu^2 \cosh \nu(x_j - x_k)}{m \sinh^2 \nu(x_j - x_k)} \\ & \quad j \neq k \end{aligned} \quad [14]$$

When the positions and momenta in L and M evolve according to the H -flow, one has

$$\dot{L}_t = [M_t, L_t] \quad [15]$$

where $[\cdot, \cdot]$ is the matrix commutator. (Indeed, [15] amounts to the Hamilton equations, as is readily checked.) Since M is anti-Hermitian, it is not difficult to derive from this Lax pair equation that the flow is isospectral: L_t is related to L_0 by a unitary transformation $L_t = U_t L_0 U_t^*$ obtained from M_t , so that the spectrum of L_t is time independent.

This argument already shows the existence of N conserved quantities under the H -flow, namely the N eigenvalues of L . It is, however, simpler to work with either the power traces H_k given by [13] or with the symmetric functions S_k of L , given by

$$\det(\mathbf{1}_N + \lambda L) = \sum_{k=0}^N \lambda^k S_k \quad [16]$$

These Hamiltonians depend only on the eigenvalues of L , so they are also conserved under the flow. Note that

$$S_1 = P, \quad S_2 = P^2 - mH \quad [17]$$

To see why these Hamiltonians are in involution, one can invoke the long-time asymptotics of the H -flow. It reads

$$\begin{aligned} p(t) &\sim \hat{p}, & \hat{p}_N &< \dots < \hat{p}_1, \\ x_j(t) &\sim x_j^+ + t\hat{p}_j/m, \\ & \quad j = 1, \dots, N, \quad t \rightarrow \infty \end{aligned} \quad [18]$$

Accordingly, one gets

$$L_t \sim \text{diag}(\hat{p}_1, \dots, \hat{p}_N) = L_\infty, \quad t \rightarrow \infty \quad [19]$$

Since the time evolution is a canonical transformation and the Poisson brackets $\{H_k, H_l\}$ are time independent (by the Jacobi identity), it now readily follows from [19] that they vanish. (Indeed, H_k and H_l reduce to power traces of L_∞ , and the asymptotic momenta $\hat{p}_1, \dots, \hat{p}_N$ Poisson commute.)

Quantum Nonrelativistic CMS Systems

The canonical quantization prescription

$$p_j \rightarrow -i\hbar \partial / \partial x_j, \quad j = 1, \dots, N \quad [20]$$

(\hbar being the Planck constant) gives rise to an unambiguous quantum Hamiltonian

$$H = -\frac{\hbar^2}{2m} \sum_{j=1}^N \partial_j^2 + \sum_{1 \leq j < k \leq N} V(x_j - x_k) \quad [21]$$

for any classical Hamiltonian [1]. Thus, the defining Hamiltonians of the above systems give rise to well-defined partial differential operators (PDOs), which act on suitable dense subspaces of the Hilbert space $L^2(G_\kappa, dx)$, $\kappa = \text{I}, \dots, \text{IV}$, with G_{I} and G_{II} given by G in [4], and $G_{\text{III}}, G_{\text{IV}}$ by [8] and [9], respectively.

We recall that there is no general result ensuring that a classically integrable system admits an integrable quantum version. More precisely, when one substitutes [20] in N Poisson commuting Hamiltonians, it need not be true that they commute as quantum operators, even when no ordering ambiguities are present. For the power trace Hamiltonians such ambiguities do occur. (For example, [11] gives rise to a term in H_3 proportional to $p_1/\sinh^2 \nu(x_1 - x_2)$.) On the other hand, no noncommuting factors occur in the quantization of S_1, \dots, S_N . To verify this, one need only note that S_k equals the sum of all $k \times k$ principal minors of L , cf. [16]; choosing a diagonal element p_j in a summand, one therefore has no dependence on x_j in the remaining factors, hence no ordering ambiguity.

As a result, the prescription [20] yields N unambiguous operators $S_k(x, -i\hbar \nabla)$, which are moreover formally self-adjoint on $L^2(G_\kappa, dx)$ for each of the four cases $\kappa = \text{I}, \dots, \text{IV}$. Although by no means obvious, it is true that these operators do commute. Thus, integrability is preserved under quantization of the above systems. Now the power traces of a matrix can be expressed as polynomials in the symmetric functions (via the Newton

identities), so this yields an ordering ensuring that the quantized power traces commute as well.

Just as the action-angle transformation for a classically integrable system “diagonalizes” all of the Poisson commuting Hamiltonians at once (in the sense that the transformed Hamiltonians depend only on the action variables), one expects that there exists a unitary operator that transforms all of the commuting Hamiltonians to diagonal form. In the classical setting, the existence of this diagonalizing map follows (under suitable technical restrictions) from the Liouville–Arnold theorem, whereas in the quantum context the existence of such a joint eigenfunction transformation is a far more delicate issue. This problem is briefly discussed later again, noting here that the solutions obtained to date vary considerably in completeness and “explicitness” for the four regimes.

Classical Relativistic CMS Systems

The nonrelativistic spacetime symmetry group is the Galilei group. Its Lie algebra is represented by the time translation generator H given by [1], space translation generator P given by [10], and the Galilei boost generator

$$B = -m \sum_{j=1}^N x_j \quad [22]$$

More precisely, the Poisson brackets are given by

$$\{H, P\} = 0, \quad \{H, B\} = P, \quad \{P, B\} = Nm \quad [23]$$

so that the last bracket does not vanish (as is the case for the Galilei Lie algebra). This deviation is inconsequential, however, since the constant Nm (central extension) yields trivial Hamilton equations.

The relativistic spacetime symmetry group (Poincaré group) yields a Lie algebra that differs from [23] only in Nm being replaced by H/c^2 , where c is the speed of light. Clearly, the functions

$$\begin{aligned} H &= mc^2 \sum_{j=1}^N \cosh\left(\frac{p_j}{mc}\right) \\ P &= mc \sum_{j=1}^N \sinh\left(\frac{p_j}{mc}\right) \end{aligned} \quad [24]$$

together with B given by [22] give rise to these altered Poisson brackets. Physically, these three generators describe a system of N relativistic free mass- m particles in terms of their rapidities p_j/mc .

A natural ansatz to take interaction into account now reads

$$\begin{aligned} H &= mc^2 \sum_{j=1}^N \cosh\left(\frac{p_j}{mc}\right) V_j(x) \\ P &= mc \sum_{j=1}^N \sinh\left(\frac{p_j}{mc}\right) V_j(x) \\ V_j(x) &= \prod_{k \neq j} f(x_j - x_k) \end{aligned} \quad [25]$$

Indeed, it is plain that this still entails

$$\{H, B\} = P, \quad \{P, B\} = H/c^2 \quad [26]$$

But to obtain a relativistic particle system, the time and space translations must also commute. The corresponding requirement $\{H, P\} = 0$ yields a severe constraint on the “pair potential” function $f(x)$ in [25] whenever $N > 2$. (For $N = 2$, one gets $\{H, P\} = 0$ irrespective of the choice of f .)

As it turns out, the vanishing requirement is satisfied when

$$f^2(x) = a + b\varphi(x) \quad [27]$$

where a, b are constants and $\varphi(x)$ is the Weierstrass function already encountered. Taking, for example, $a, b > 0$, one can take the positive square root of the right-hand side of [27]. This choice of $f(x)$ yields the defining Hamiltonian of the relativistic elliptic system (type IV). In the three degenerate cases, it is convenient to choose

$$f(x) = \begin{cases} (1 + g^2/m^2 c^2 x^2)^{1/2} & \text{(I)} \\ (1 + \sin^2(vg/mc)/\sinh^2(vx))^{1/2} & \text{(II)} \\ (1 + \sinh^2(vg/mc)/\sin^2(vx))^{1/2} & \text{(III)} \end{cases} \quad [28]$$

It is an elementary exercise to check that this implies

$$\lim_{c \rightarrow \infty} (H - Nm c^2) = H_{\text{nr}}, \quad \lim_{c \rightarrow \infty} P = P_{\text{nr}} \quad [29]$$

where H_{nr} and P_{nr} are the above nonrelativistic time and space translation generators. Hence, the defining Hamiltonians of the relativistic systems reduce to their nonrelativistic counterparts in the limit $c \rightarrow \infty$.

The special character of the function [27] makes itself felt not only in ensuring Poincaré invariance, but also in entailing integrability. To begin with, note that the functions

$$S_{\pm N} = \exp\left(\pm \beta \sum_{j=1}^N p_j\right), \quad \beta = 1/mc \quad [30]$$

commute with H and P , so that integrability for $N=3$ is plain. More generally, the Hamiltonians

$$S_{\pm l} = \sum_{\substack{I \subset \{1, \dots, N\} \\ |I|=l}} \exp\left(\pm \beta \sum_{j \in I} p_j\right) \prod_{\substack{j \in I \\ k \notin I}} f(x_j - x_k), \quad [31]$$

$l = 1, \dots, N$

can be shown to mutually commute. Clearly, one has

$$S_{-l} = S_{-N} S_{N-l}, \quad l = 1, \dots, N-1 \quad [32]$$

and

$$H = (S_1 + S_{-1})/2m\beta^2, \quad P = (S_1 - S_{-1})/2\beta \quad [33]$$

As anticipated by the notation, the functions S_1, \dots, S_N may be viewed as the symmetric functions of a Lax matrix. More precisely, in the elliptic case this is true up to multiplicative constants that depend on a spectral parameter occurring in the Lax matrix. As before, only the Lax matrix for the type II system is specified here. In this case, one can dispense with the spectral parameter and choose

$$L_{jk} = e_j C_{jk} e_k, \quad j, k = 1, \dots, N \quad [34]$$

where

$$e_j = \exp(\nu x_j + \beta p_j/2) \prod_{l \neq j} f(x_j - x_l)^{1/2} \quad [35]$$

$$C_{jk} = \exp(-\nu(x_j + x_k)) \frac{\sinh(i\beta\nu g)}{\sinh \nu(x_j - x_k + i\beta g)} \quad [36]$$

In [35], $f(x)$ is the type II function given by [28]. The matrix C arises from Cauchy's matrix $1/(w_j - z_k)$ via a suitable substitution, and Cauchy's identity

$$\det \left(\frac{1}{w_j - z_k} \right)_{j,k=1}^N = \prod_{j=1}^N \frac{1}{w_j - z_j} \prod_{1 \leq j < k \leq N} \frac{(w_j - w_k)(z_j - z_k)}{(w_j - z_k)(z_j - w_k)} \quad [37]$$

ensures that [34] yields the Hamiltonians S_l of [31].

To conclude this section, we point out that the relation

$$L = \mathbf{1}_N + \beta L_{\text{nr}} + O(\beta^2), \quad \beta \rightarrow 0 \quad [38]$$

where L_{nr} denotes the nonrelativistic Lax matrix [11], can be used to deduce the involutivity of the nonrelativistic Hamiltonians from that of their relativistic counterparts.

Quantum Relativistic CMS Systems

When the canonical quantization prescription [20] is applied to the classical Hamiltonians [31] with

$f(x) = 1$, one obtains commuting quantum operators whose action is exemplified by

$$\exp\left(-\frac{\hbar}{mc} i \frac{d}{dx}\right) F(x) = F\left(x - i \frac{\hbar}{mc}\right) \quad [39]$$

That is, the operators act on functions that have an analytic continuation in x_1, \dots, x_N from the real line \mathbb{R} to a strip around \mathbb{R} in the complex plane \mathbb{C} , whose width is at least $2\hbar/mc$.

Operators of this type are called analytic difference operators (henceforth $\Lambda\Delta\text{Os}$). The choice $f(x) = 1$ amounts to the free case $g=0$ in [28]. For $g \neq 0$, however, the canonical quantization exemplified by [39] yields noncommuting $\Lambda\Delta\text{Os}$. Thus, the factor ordering following from [31] would entail that integrability breaks down at the quantum level.

As mentioned before, there is no general result guaranteeing that a different ordering that preserves integrability exists. Even so, this is true in the present case. Specifically, the function $f(x)$ can be factorized as $f_+(x)f_-(x)$, and then the $\Lambda\Delta\text{Os}$

$$S_{\pm l} = \sum_{\substack{I \subset \{1, \dots, N\} \\ |I|=l}} \prod_{\substack{j \in I \\ k \notin I}} f_{\mp}(x_j - x_k) \times \exp\left(\mp i\hbar\beta \sum_{j \in I} \partial_j\right) \prod_{\substack{j \in I \\ k \notin I}} f_{\pm}(x_j - x_k) \quad [40]$$

do commute. In the elliptic case [27], this factorization involves the Weierstrass σ -function, and commutativity can be encoded in a sequence of functional equations satisfied by the σ -function. For the type I–III systems the pertinent factorization of [28] is given by

$$f_{\pm}(x) = \begin{cases} (1 \pm i\beta g/x)^{1/2} & \text{(I)} \\ (\sinh \nu(x \pm i\beta g)/\sinh \nu x)^{1/2} & \text{(II)} \\ (\sin \nu(x \pm i\beta g)/\sin \nu x)^{1/2} & \text{(III)} \end{cases} \quad [41]$$

(Here one has $g > 0$, and the choice of square root is such that $f_{\pm}(x) \rightarrow 1$ for $g \downarrow 0$.)

The nonrelativistic limit $c \rightarrow \infty$ of the quantum Hamiltonians [33] can be determined by expanding S_1 and S_{-1} in a power series in $\beta = 1/mc$. In this way, one obtains once more [29], except for a small, but crucial change in H_{nr} : instead of the coupling constant dependence g^2 in the potential energy, one gets $g(g - \hbar)$. The extra term arises from the action of the term linear in β in the expansion of the exponential on the term linear in β in the expansion of the functions $f_{\pm}(x)$.

From the perspective of the nonrelativistic quantum CMS systems, the change $g^2 \rightarrow g(g - \hbar)$ appears ad hoc. As it transpires, however, the different

dependence on g ensures that the eigenfunctions of H_{nr} depend on g in a far simpler way. This will become clear shortly.

Action-Angle Transforms and Duality

Under certain technical assumptions, any integrable system given by N independent Poisson commuting Hamiltonians $S_1(x, p), \dots, S_N(x, p)$ on a $2N$ -dimensional phase space admits local canonical transformations to action-angle variables. Like the spectral theorem on the quantum level, this structural result is of limited practical value. Indeed, just as the spectral theorem yields no concrete information concerning eigenfunctions, bound-state energies, scattering, etc., associated with a given self-adjoint Hamiltonian, the Liouville–Arnold theorem only yields general insight in the type of motion that can occur and the geometric character of the local maps (in terms of invariant tori).

To fully comprehend (“solve”) a given integrable system, one should render the associated action-angle map as concrete as possible. For the CMS type systems, a complete solution to this problem has only been achieved for the systems of type I–III. The motion in the trigonometric systems is oscillatory, so that a closeup via the action-angle transform involves extensive geometric constructions. By contrast, the type I and II systems are scattering systems, and here the action-angle map can be tied in with the classical wave maps (Møller transformations).

We now sketch some salient features of the action-angle maps for systems of type I and II. In all cases the map (denoted Φ) is a canonical transformation from the phase space Ω (eqn [3]) with 2-form $dx \wedge dp$ to the phase space

$$\hat{\Omega} = \{(\hat{x}, \hat{p}) \in \mathbb{R}^{2N} \mid \hat{p} \in G\} \quad [42]$$

with 2-form $d\hat{x} \wedge d\hat{p}$. Thus, the actions $\hat{p}_1, \dots, \hat{p}_N$ vary over G given by [4] and the “angles” $\hat{x}_1, \dots, \hat{x}_N$ over \mathbb{R} . Consequently, $\hat{\Omega}$ amounts to Ω with x and p interchanged.

As should be the case, the transformed commuting Hamiltonians

$$\hat{S}_k = S_k \cdot \Phi^{-1}, \quad k = 1, \dots, N \quad [43]$$

depend only on the action vector \hat{p} . To be specific, they arise from $S_k(x, p)$ by taking $g=0$ (no interaction, hence no x dependence) and substituting $p \rightarrow \hat{p}$. Indeed, the actions \hat{p}_k are the $t \rightarrow \infty$ limits of the momenta $p_k(t)$, where the t dependence refers to the defining Hamiltonian of the system.

As it happens, the Lax matrix L is of decisive importance to concretize the action-angle map Φ ,

and in particular to reveal its hidden duality properties. The starting point is a commutation relation of $L(x, p)$ with a diagonal matrix $A(x)$ given by

$$\begin{aligned} A(x) &= \text{diag}(d(x_1), \dots, d(x_N)) \\ d(y) &= \begin{cases} y & \text{(I)} \\ \exp(2\nu y) & \text{(II)} \end{cases} \end{aligned} \quad [44]$$

Obviously, the symmetric functions $\check{D}_k(x)$ of $A(x)$ yield an integrable system on Ω , so the Hamiltonians

$$D_k(\hat{x}, \hat{p}) = (\check{D}_k \circ \Phi^{-1})(\hat{x}, \hat{p}), \quad k = 1, \dots, N \quad [45]$$

yield an integrable system on the action-angle phase space $\hat{\Omega}$. The crux of the matter is now that these systems are familiar: they are also systems of type I and II!

To be specific, let us denote the dual systems just described by a caret, and the nonrelativistic/relativistic systems by a suffix nr/rel, resp. Then the duality properties alluded to are given by

$$\begin{aligned} \hat{I}_{nr} &\simeq I_{nr}, & \hat{I}_{rel} &\simeq I_{nr} \\ \hat{II}_{nr} &\simeq I_{rel}, & \hat{II}_{rel} &\simeq I_{rel} \end{aligned} \quad [46]$$

and Φ^{-1} serves as the action-angle map for the dual systems.

In order to sketch why this state of affairs holds true for the II_{rel} system, recall that its Lax matrix is given by [34]. From this, one readily checks the commutation relation

$$\text{coth}(i\beta\nu g)[A, L] = 2e \otimes e - (AL + LA) \quad [47]$$

Since L is Hermitean, there exists a unitary U diagonalizing L . It can now be shown that the spectrum of L is positive and nondegenerate, and that U^*e has nonzero components. The gauge ambiguity in U (given by a permutation matrix and diagonal phase matrix) can, therefore, be fixed by requiring

$$\begin{aligned} U^*LU &= \text{diag}(\exp(\beta\hat{p}_1), \dots, \exp(\beta\hat{p}_N)), \\ \hat{p}_N &< \dots < \hat{p}_1 \end{aligned} \quad [48]$$

$$(U^*e)_j > 0, \quad j = 1, \dots, N \quad [49]$$

A suitable reparametrization of U^*e then yields the “angle” vector \hat{x} .

As a consequence, U^*AU becomes a function of \hat{x} and \hat{p} . In detail, one finds

$$(U^*AU)(\hat{x}, \hat{p}) = L(\beta/2, 2\nu; \hat{p}, \hat{x})^T \quad [50]$$

where $L(\nu, \beta; x, p)$ is given by [34] and T denotes the transpose. Therefore, the “dual Lax matrix” $\hat{A} = U^*AU$ is essentially equal to L , explaining the self-duality $\hat{II}_{rel} \simeq II_{rel}$ announced above.

With the action-angle transform under explicit control, much more can be said about the solutions to Hamilton’s equations for each of the commuting Hamiltonians, both as regards finite times and as regards long-time asymptotics (scattering). It is beyond the scope of this article to enlarge on this, but it is worth mentioning that the scattering reveals the solitonic character of the particles. Indeed, the set of asymptotic momenta $\hat{p}_1, \dots, \hat{p}_N$ is conserved under the scattering and the asymptotic position shifts are factorized in terms of pair shifts. A quite remarkable feature of the type I systems is that the shifts actually vanish (“billiard ball” scattering).

Eigenfunction Transforms and Duality

Both at the relativistic and at the nonrelativistic level the commuting quantum Hamiltonians S_1, \dots, S_N are formally self-adjoint on the Hilbert space $L^2(G_\kappa, dx), \kappa = \text{I}, \dots, \text{IV}$. Thus, it may be expected that it is possible to construct a unitary eigenfunction transform

$$\Phi_\kappa : L^2(G_\kappa, dx) \rightarrow L^2(\hat{G}_\kappa, d\mu_\kappa(p)), \quad \kappa = \text{I}, \dots, \text{IV} \quad [51]$$

diagonalizing S_k as multiplication by a real-valued function $M_k(p)$. Here \hat{G}_κ encodes the joint spectrum and $d\mu_\kappa(p)$ is a suitable measure on \hat{G}_κ .

Obviously, this expectation is borne out in the free case $g=0$. Then, Φ_κ is basically Fourier transformation, its kernel consisting of a sum of joint eigenfunctions

$$\exp(-ix \cdot \sigma(p)/\hbar), \quad \sigma \in S_N \quad [52]$$

with σ ranging over the permutation group S_N . For $\kappa = \text{I}, \text{II}$, one can take $G_\kappa = \hat{G}_\kappa = G$ (eqn [4]) and $d\mu_\kappa(p) = dp$. Here one gets

$$M_k(p) = \sum_{1 \leq i_1 < \dots < i_k \leq N} \begin{cases} p_{i_1} \cdots p_{i_k} \\ \exp(\beta p_{i_1}) \cdots \exp(\beta p_{i_k}) \end{cases} \quad [53]$$

in the nonrelativistic and relativistic case, resp. For $\kappa = \text{III}, \text{IV}$, one needs to take into account periodic boundary conditions on the walls of G_κ , yielding a discrete joint spectrum after the center-of-mass motion is omitted. (With the above choices of G_{III} and G_{IV} , cf. [8] and [9], the center-of-mass motion is a free motion along the line, so the total momentum still varies continuously.) Of course, the diagonalized S_k are once more given by [53], since the kernel of Φ_κ consists of free boson states.

Taking next $g > 0$, the above expectation has not been confirmed for all of the eight regimes involved. This is not only because in some cases not even the

existence of joint eigenfunctions has been shown, but also because in the relativistic case the unitarity of Φ_{II} and Φ_{IV} already breaks down for $N=2$ when g increases beyond a critical value, cf. [57] below. It is quite likely that this happens for $N > 2$ as well, but this is not readily apparent from the current fragmentary knowledge on joint eigenfunctions for $N > 2$.

The only two cases where the $g > 0$ joint eigenfunction transform is of an elementary nature are the III_{nr} and III_{rel} cases. Indeed, the joint eigenfunctions describing the internal motion are of the form

$$\psi_n(x) = W(x)^{1/2} P_n(x), \quad n \in \mathbb{N}^{N-1} \quad [54]$$

Here,

$$W(x) = \prod_{1 \leq j < k \leq N} w(x_j - x_k) \quad [55]$$

is a positive weight function on G_{III} and the $P_n(x)$ are multivariable orthogonal polynomials. Thus, $P_n(x)$ is a finite linear combination of the above free boson states, with p in [52] a linear function of n . For the III_{nr} case, these eigenfunctions were already found by Sutherland. (Here, the functions $P_n(x)$ amount to polynomials, often called the Jack polynomials, which arose in a statistics context.) The III_{rel} polynomials may be viewed as the special A_{N-1} case of Macdonald’s orthogonal q -polynomials for arbitrary root systems, with

$$q = \exp(-2\hbar\beta\nu) \quad [56]$$

(Note that q converges to 1 both in the nonrelativistic limit $c \rightarrow \infty$ and in the classical limit $\hbar \rightarrow 0$.)

For the II_{nr} case, the joint eigenfunctions were found and studied a couple of decades ago by Heckman and Opdam, yielding a multivariable hypergeometric transform. Indeed, for $N=2$, the eigenfunctions can be expressed in terms of the hypergeometric function ${}_2F_1$, as has been known since the early days of quantum mechanics. Likewise, the arbitrary- N I_{nr} joint eigenfunction transform (studied in detail by de Jeu) can be viewed as a multivariable Hankel transform, the $N=2$ kernel being essentially a Hankel function.

Much less is known concerning IV_{nr} eigenfunctions, and *a fortiori* for the associated transform Φ_{IV} . For $N=2$ the time-independent Schrödinger equation amounts to the Lamé equation. Hence, solutions are Lamé functions that can be studied in particular via Fuchs theory (regular singularities). A far more explicit form of the eigenfunctions dates back to work by Hermite in the nineteenth century. More precisely, provided the g dependence of the

defining Hamiltonian is changed from g^2 to $g(g - \hbar)$ (a change already encountered above), Hermite's results apply to couplings $g = l\hbar$, $l = 2, 3, 4, \dots$. His eigenfunctions have a structure that is nowadays referred to as the Bethe ansatz. For the same g values and arbitrary N , H_{nr} eigenfunctions of Bethe ansatz type were found and studied by Felder and Varchenko, but even for these g values much remains to be done to achieve a complete understanding of the Φ_{IV} transform.

A quite different approach, due to Komori and Takemura, does yield rather detailed information on Φ_{IV} for arbitrary $g > 0$. The key feature of their strategy is to view the IV_{nr} case as a perturbation of the III_{nr} case. This entails, however, that the validity of their results is restricted to large imaginary period of the φ -function.

For the IV_{rel} system, there are only rather complete results on Φ_{IV} for $N = 2$. More specifically, the eigenfunction transform is known to be unitary for

$$g \in [0, \hbar + \pi/\beta\nu] \quad [57]$$

and a dense set in a corresponding parameter space. (For g outside this interval, unitarity is violated.) The kernel of Φ_{IV} involves eigenfunctions of Bethe ansatz structure. For $g = l\hbar$, $l = 2, 3, \dots$ and arbitrary N , Bethe ansatz type H_{rel} eigenfunctions were found by Billey, generalizing the Felder–Varchenko results mentioned above.

It remains to discuss the I_{rel} and II_{rel} systems. To this end, we first recall the classical dualities [46]. It is natural to expect that these dualities are still present at the quantum level. For the I_{nr} case, this is readily confirmed: the transform is indeed invariant under interchange of x and p . In fact, the $N = 2$ center-of-mass Hankel transform even depends only on $(x_1 - x_2)(p_1 - p_2)$, so that self-duality is manifest in this case.

More generally, for $N = 2$ the expected dualities [46] are indeed present. The $\text{II}_{\text{nr } 2F_1}$ transform satisfies the I_{rel} analytic difference equation in $p_1 - p_2$ due to the contiguous relations obeyed by ${}_2F_1$. The II_{rel} transform is only unitary when g is restricted by [57], and it is indeed self-dual in the same sense as the action-angle map (Ruijsenaars).

Turning finally to the case $N > 2$, the multi-variable hypergeometric transform Φ_{II} does have the expected duality property. More specifically, its inverse diagonalizes the commuting I_{rel} AΔOs (Chalykh). For II_{rel} with $N > 2$ and $g = l\hbar$, $l = 2, 3, \dots$, Chalykh also finds elementary joint eigenfunctions with the expected self-duality. To date, no Hilbert space results for the $N > 2$ II_{rel} case have been obtained.

To conclude, we mention that the soliton scattering behavior at the classical level is preserved under quantization in all cases where this can be checked. That is, no new momenta are created in the scattering process and the S -matrix is factorized as a product of pair S -matrices. Moreover, for the type I cases, the S -matrix is a momentum-independent (but g -dependent) phase, as a quantum analog of the classical billiard ball scattering.

See also: Bethe Ansatz; Classical r -Matrices, Lie Bialgebras, and Poisson Lie Groups; Functional Equations and Integrable Systems; Integrable Discrete Systems; Integrable Systems and Algebraic Geometry; Integrable Systems in Random Matrix Theory; Integrable Systems: Overview; Isochronous Systems; Ordinary Special Functions; q -Special Functions; Quantum Calogero–Moser Systems; Seiberg–Witten Theory; Separation of Variables for Differential Equations; Sine-Gordon Equation; Toda Lattices.

Further Reading

- Babelon O, Bernard D, and Talon M (2003) *Introduction to Classical Integrable Systems*. Cambridge: Cambridge University Press.
- Calogero F (1971) Solution of the one-dimensional N -body problem with quadratic and/or inversely quadratic pair potentials. *Journal of Mathematical Physics* 12: 419–436.
- Calogero F (2001) *Classical Many-Body Problems Amenable to Exact Treatments*. Berlin: Springer.
- van Diejen JF and Vinet L (eds.) (2000) *Calogero–Moser–Sutherland Models*. Berlin: Springer.
- Fock V, Gorsky A, Nekrasov N, and Rubtsov V (2000) Duality in integrable systems and gauge theories. *Journal of High Energy Physics* 7(28): 1–39.
- Marshakov A (1999) *Seiberg–Witten Theory and Integrable Systems*. Singapore: World Scientific.
- Moser J (1975) Three integrable Hamiltonian systems connected with isospectral deformations. *Advances in Mathematics* 16: 197–220.
- Olshanetsky MA and Perelomov AM (1981) Classical integrable finite-dimensional systems related to Lie algebras. *Physics Reports* 71: 313–400.
- Olshanetsky MA and Perelomov AM (1983) Quantum integrable systems related to Lie algebras. *Physics Reports* 94: 313–404.
- Ruijsenaars SNM (1987) Complete integrability of relativistic Calogero–Moser systems and elliptic function identities. *Communications in Mathematical Physics* 110: 191–213.
- Ruijsenaars SNM (1999) Systems of Calogero–Moser type. In: Semenov G and Vinet L (eds.) *Proceedings of the 1994 Banff Summer School Particles and Fields*, pp. 251–352. Berlin: Springer.
- Ruijsenaars SNM and Schneider H (1986) A new class of integrable systems and its relation to solitons. *Annals of Physics (NY)* 170: 370–405.
- Sutherland B (1972) Exact results for a quantum many-body problem in one dimension II. *Physical Review A* 5: 1372–1376.

Canonical General Relativity

C Rovelli, Université de la Méditerranée et Centre de Physique Théorique, Marseilles, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Lagrangian formulations of general relativity (GR) were found by Hilbert and by Einstein himself, almost immediately after the discovery of the theory. The construction of Hamiltonian formulations of GR, on the other hand, has taken much longer, and has required decades of theoretical research.

The first such formulations were developed by Dirac and by Bergmann and his collaborators, in the 1950s. Their cumbersome formalism was simplified by the introduction of new variables: first by Arnowit, Deser, and Misner in the 1960s and then by Ashtekar in the 1980s. A large number of variants and improvements of these formalisms have been developed by many other authors. Most likely the process is not over, and there is still much to learn about the canonical formulation of GR.

A number of reasons motivate the study of canonical GR. In general, the canonical formalism can be an important step towards quantum theory; it allows the identification of the physical degrees of freedom, and the gauge-invariant states and observables of theory; and it is an important tool for analyzing formal aspects of the theory such as its Cauchy problem. All these issues are highly non-trivial, and present open problems, in GR.

In turn, the structural peculiarity and the conceptual novelty of GR have motivated re-analyses and extensions of the canonical formalism itself.

The following sections discuss the source of the peculiar difficulty of canonical GR, and summarize the formulations of the theory that are most commonly used.

The Origin of the Difficulties

The reason for the complexity of the Hamiltonian formulation of GR is not so much in the intricacy of its nonlinear field equations; rather, it must be found in the conceptual novelty introduced by GR at the very foundation of the structure of mechanics.

The dynamical systems considered before GR can be formulated in terms of states evolving in time. One assumes that a time variable t can be measured by a physical clock, and that certain observable quantities A of the system can be measured at every instant of time. If we know the state s of the system at some

initial time, the theory predicts the value $A(t)$ of these quantities for any given later instant of time t . The space of the possible initial states s is the phase space Γ_0 . Observables are real functions on Γ_0 . Infinitesimal time evolution can be represented as a vector field in Γ_0 . This vector field is determined by the Hamiltonian, which is also a function on Γ_0 . The integral lines $s(t)$ of this vector field determine the time evolution $A(t) = A(s(t))$ of the observables.

This conceptual structure is very general. It can be easily adapted to special-relativistic systems. However, it is not general enough for general-relativistic systems. GR is *not* formulated as the evolution of states and observables in a preferred time variable which can be measured by a physical clock. Rather, it is formulated as the *relative* (common) evolution of many observable quantities. Accordingly, in GR there is no quantity playing the same role as the conventional Hamiltonian. In fact, the canonical Hamiltonian density that one obtains from a Legendre transformation from a Lagrangian vanishes identically in GR.

The origin of this peculiar behavior of the theory is the following. The field equations are written as evolution equations in a time coordinate t . However, they are invariant under arbitrary changes of t . That is, if we replace t with an arbitrary function $t' = t'(t)$ in a solution of the field equations, we obtain another solution. This underdetermination does not lead to a lack of predictivity in GR, because we do not interpret the variable t as the measurable reading of a physical clock, as we do in non-general-relativistic theories. Rather, we interpret t as a nonobservable mathematical parameter, void of physical significance. Accordingly, the notions of “state at a given time” and “value of an observable at a given time” are very unnatural in GR.

A Hamiltonian formulation of GR requires a version of the canonical formalism sufficiently general to deal with this broader notion of evolution. Generalizations of the Hamiltonian formalism have been developed by many authors, such as Dirac (see below), Souriau, Arnold, Witten, and many others. The first step in this direction was taken by Lagrange himself: Lagrange gave a time-independent interpretation of the phase space as the space Γ of the solutions of the equations of motion (modulo gauges). As we shall see, however, consensus is still lacking on a fully satisfactory formalism.

Dirac Theory of Constrained Systems

Dirac has developed a Hamiltonian theory for mechanical systems with constraints, precisely in

view of its application to GR. Dirac's theory is beautiful, finds vast applications, and it is still commonly taken as the basis to discuss Hamiltonian GR, although GR does not fit very naturally into Dirac's scheme. In the following, only the part of Dirac's theory relevant for GR is summarized.

Consider a Lagrangian system with Lagrangian variables q^i , with $i = 1, \dots, n$. Call v^i the corresponding velocities. Let the system be defined by the Lagrangian $L(q^i, v^i)$. The momenta are defined as functions of q^i and v^i by $p_i(q^i, v^i) = \partial L(q^i, v^i) / \partial v^i$. The canonical Hamiltonian $H(q^i, p_i) = v^i(q^i, p_i) p_i - L(q^i, v^i(q^i, p_i))$ (summation over repeated indices is understood) is obtained by inverting the function $p_i(q^i, v^i)$ and expressing the velocities as functions of the momenta $v^i(q^i, p_i)$. The phase space Γ_0 is the space of the variables (q^i, p_i) . Infinitesimal time evolution is given by the vector field $V = v^i(q^i, p_i) \partial / \partial q^i + f_i(q^i, p_i) \partial / \partial p_i$, where velocities and forces are given by the Hamilton equations $v^i = \partial H / \partial p_i$ and $f_i = -\partial H / \partial q^i$.

More formally, the 2-form $\omega = dp_i \wedge dq^i$ endows Γ_0 with a symplectic structure. In the presence of such a structure, every function A determines a vector field V_A , defined by $i_{V_A} \omega = -dA$. By integrating this field, we have a flow in Γ_0 , called the flow generated by A . Time evolution is the flow generated by the Hamiltonian. Given two functions A and B , their Poisson brackets are defined by the function $\{A, B\} = -V_A(B) = V_B(A)$. Therefore, the time evolution of an observable A satisfies $dA/dt = \{A, H\}$. A dynamical system is completely characterized by the set $(\Gamma_0, \omega, \mathcal{A}, H)$, where $\mathcal{A} = (A_1, \dots, A_N)$ is the ensemble of the observables.

A constrained system, in the sense of Dirac, is a system for which the image of the function $v^i \rightarrow p_i(q^i, v^i)$ is smaller than R^n . We can characterize the image \mathcal{I} of the map $(q^i, v^i) \rightarrow (q^i, p_i)$ with a set of equations on Γ_0

$$C_\alpha(q^i, p_i) = 0 \quad [1]$$

where $\alpha = 1, \dots, m'$. These are called the primary constraints.

The "constraint surface" C is the largest subspace of \mathcal{I} which is preserved by time evolution. It can be characterized by adding additional constraints, still of the form (1), with $\alpha = m' + 1, \dots, m$. These additional constraints, called secondary constraints, can be computed as the Poisson brackets of the primary constraints with the Hamiltonian (plus the Poisson brackets of these secondary constraints with the Hamiltonian, and so on, until the Poisson brackets of all the constraints with the Hamiltonian vanish on in C). We say that an equation holds weakly if it holds on C .

A constrained system is "first class" if the Poisson brackets of the constraints among themselves vanishes weakly. Maxwell theory and GR are first-class constrained systems. In a first-class constrained system, the constraints generate flows that preserve C and foliate it into "orbits." The space of these orbits is called the physical phase space (see [Figure 1](#)).

This flow is interpreted as a "gauge" transformation, namely as a change of mathematical description of the same physical state. As first observed by Dirac, such interpretation is necessary if we demand a deterministic physical evolution, for the following reason. A first-class constrained system is a system in which the time evolution $q^i(t)$ of the Lagrangian variables is not completely determined by the equations of motion. (The relation between constraints and underdetermination of the evolution is simple to understand. In a Lagrangian system, the number of equations of motion is equal to the number of Lagrangian variables. If one of these equations is a constraint (between the initial velocities and initial coordinates), then one evolution equation is missing.) To recover a deterministic physical evolution, we must interpret two "mathematical" states that can evolve from the same initial data, as describing the same "physical" state. As shown by Dirac, the transformations generated by the constraints are precisely the ones that implement such an identification.

It follows that the physical states must be identified with the equivalence classes of the points of C under the gauge transformations generated by the constraints, namely with the orbits of their flow. It is easy to show that (locally) there is a unique symplectic 2-form ω_{ph} on Γ_{ph} such that its pullback to C is equal to the pullback of ω to C ($i_* \omega = \pi_* \omega_{\text{ph}}$, see [Figure 1](#)). Physical observables A_{ph} are functions on C that are gauge invariant, namely constant on

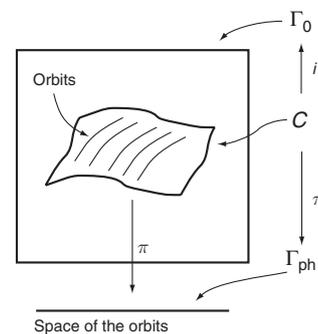


Figure 1 The structure of a first-class constrained system. Γ_0 : phase space, C : constraint surface, Γ_{ph} : physical phase space; i : imbedding of C in Γ ; π projection to orbit space (sending each point into its orbit).

the orbits. That is, they are functions on Γ_{ph} . The Hamiltonian is a physical observable. The dynamical system $(\Gamma_{\text{ph}}, \omega_{\text{ph}}, \mathcal{A}_{\text{ph}}, H)$, where \mathcal{A}_{ph} is the ensemble of the physical observables, is a complete description of the physical system, called the gauge-invariant formulation, with no more constraints or gauges.

For instance, the phase space of Maxwell theory is coordinatized by the Maxwell potential $A_\mu(\mathbf{x})$, $\mu=0, 1, 2, 3$, and its conjugate momentum $E^\mu(\mathbf{x})$. Since the time derivative of A_0 does not appear in the Maxwell action, the primary constraint is

$$E^0(\mathbf{x}) = 0 \quad [2]$$

The secondary constraint turns out to be the Gauss law,

$$\partial_a E^a(\mathbf{x}) = 0 \quad [3]$$

where $a=1, 2, 3$. The first generates arbitrary transformations of A_0 , while the second generates the time-independent gauge transformations $\delta A_a(\mathbf{x}) = \partial_a \lambda(\mathbf{x})$. The pair (A_0, π^0) can be dropped altogether, since it is formed by a pure gauge variable and a variable constrained to vanish. The (gauge-invariant) Hamiltonian is $H = 1/8\pi \int d^3\mathbf{x} (E^a E_a + B^a B_a)$, where $B^a = \epsilon^{abc} \partial_b A_c$ is the magnetic field and E^a is easily recognized as the electric field. E^a and B_a are the physical observables.

General Structure of GR Constraints

GR fits into Dirac theory with a certain difficulty. Since the constraints are the generators of the gauge invariances, it is easy to determine their structure in GR. The gauge invariances of GR are given by the coordinate transformations $x^\mu \rightarrow x'^\mu = f^\mu(x)$, where $x = (\mathbf{x}, t)$. Accordingly, we have four primary constraints $\pi^\mu = 0$, analogous to [2], and four secondary constraints $C_\mu(\mathbf{x}) = 0$, analogous to [3]. These are usually separated into the three ‘‘momentum’’ constraints

$$C_a(\mathbf{x}) = 0 \quad [4]$$

which generate fixed-time spatial coordinate transformations and the ‘‘Hamiltonian’’ constraint

$$C(\mathbf{x}) = 0 \quad [5]$$

which generates changes in the t coordinate.

The metric $g_{\mu\nu}(x)$ that represents the gravitational field in Einstein’s original formulation has ten independent components per point. Each first-class constraint indicates that one Lagrangian variable is a gauge degree of freedom. The physical degrees of

freedom of GR are therefore $(10 - 4 - 4) = 2$ per point. In the linearized theory, these are the two degrees of freedom that describe the two polarizations of a gravitational wave of given momentum. Formulations of GR in which there are additional gauge invariances (such as Cartan’s tetrad formulation, see below) have, accordingly, more constraints.

Since the Hamiltonian generates evolution in the Lagrangian evolution parameter t , and since such evolution can be obtained as a gauge transformation, it follows that the Hamiltonian is a constraint in GR. The vanishing of the Hamiltonian is a characteristic feature of general-relativistic systems. The Hamiltonian structure of GR is therefore determined by its phase space and its constraints. The gauge-invariant formulation of the theory is given just by the set $(\Gamma_{\text{ph}}, \omega_{\text{ph}}, \mathcal{A}_{\text{ph}})$ and no Hamiltonian. The physical interpretation of this structure is discussed in the last section.

ADM Formalism

In Einstein’s formulation, the Lagrangian variable of GR is the metric field $g_{\mu\nu}(x, t)$ (here we use the signature $[-, +, +, +]$). Arnowit, Deser, and Misner have introduced the following change of variables:

$$q_{ab} = g_{ab}, \quad N = 1/\sqrt{-g^{00}}, \quad N^a = q^{ab} g_{a0} \quad [6]$$

where q^{ab} is the inverse of the three-dimensional metric q_{ab} , used henceforth to raise and lower space indices $a, b = 1, 2, 3$. This is equivalent to writing the invariant interval in the form

$$ds^2 = -N^2 dt^2 + q_{ab}(dx^a + N^a dt)(dx^b + N^b dt)$$

These variables have an interesting geometric interpretation. Consider a family of spacelike (‘‘ADM’’) surfaces Σ_t defined by $t = \text{constant}$. q_{ab} is the 3-metric induced on the surface. N is called the ‘‘lapse’’ function and N^a is called the ‘‘shift’’ function. Their geometrical interpretation is illustrated in [Figure 2](#).

When written in terms of these variables, the action of GR takes the form

$$S[q_{ab}, N, N^a] = \int d^4x \sqrt{q} N [R + k_{ab} k^{ab} - k^2]$$

where $q = \det q_{ab}$ and R are the determinant and the Ricci scalar of the metric q_{ab} ;

$$k_{ab} = \frac{1}{2N} (\partial_t q_{ab} - D_a N_b - D_b N_a)$$

is the extrinsic curvature of the constant time surface; and D_a is the covariant derivative of q_{ab} . This action is independent of the time derivatives of

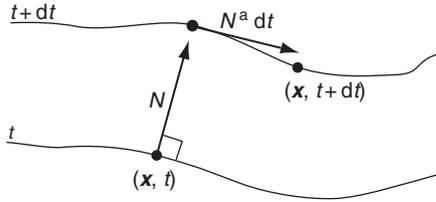


Figure 2 The geometrical interpretation of the lapse $N(\mathbf{x}, t)$ and shift $N^a(\mathbf{x}, t)$ fields. Two ADM surfaces, defined by the values t and $t + dt$, are displayed. $N(\mathbf{x}, t)dt$ is the proper length of the vector joining the two surfaces, normal to the first surface at (\mathbf{x}, t) . This is the proper time lapsed between the two surfaces for an observer at rest on the first surface at (\mathbf{x}, t) . The quantity $d\mathbf{x}^a = N^a(\mathbf{x}, t)dt$ is the shift (the displacement) between the endpoint of this vector and the point $(\mathbf{x}, t + dt)$ having the same spacial coordinates as (\mathbf{x}, t) .

N and N^a . The conjugate momenta π and π_a of these quantities are therefore the primary constraints and the pairs (π, N) and (π_a, N^a) can be taken out of the phase space as for the pair (E^0, A_0) in the Maxwell example. We can therefore take the 3-metric $q_{ab}(\mathbf{x})$ and its conjugate momentum $p^{ab}(\mathbf{x})$ as the canonical variables of GR. The momentum is related to the “velocity” $\partial_t q_{ab}$, by

$$p^{ab} = \sqrt{q}(k^{ab} - kq^{ab})$$

where $k = k_{ab}q^{ab}$.

The secondary constraints [4] and [5] turn out to be

$$C_a = \sqrt{q}D_b \left(\frac{1}{\sqrt{q}} p^b{}_a \right) = 0 \quad [7]$$

and

$$C = \frac{1}{\sqrt{q}} \left(p^{ab} p_{ab} - \frac{1}{2} p^2 \right) - \sqrt{q}R = 0 \quad [8]$$

where $p = p^{ab} q_{ab}$

If the two fields $q_{ab}(\mathbf{x}, t)$ and $p^{ab}(\mathbf{x}, t)$ satisfy the Hamilton equations

$$\frac{\partial q_{ab}(\mathbf{x}, t)}{\partial t} = \{q_{ab}(\mathbf{x}, t), H(t)\} \quad [9]$$

$$\frac{\partial p^{ab}(\mathbf{x}, t)}{\partial t} = \{p^{ab}(\mathbf{x}, t), H(t)\} \quad [10]$$

where

$$H(t) = \int d^3x N(\mathbf{x}, t) C[q_{ab}(\mathbf{x}, t), p^{ab}(\mathbf{x}, t)] \\ + N^a(\mathbf{x}, t) C_a[q_{ab}(\mathbf{x}, t), p^{ab}(\mathbf{x}, t)]$$

with arbitrary functions $N(\mathbf{x}, t), N^a(\mathbf{x}, t)$, then the metric $g_{\mu\nu}(\mathbf{x}, t)$, defined from q_{ab}, N, N^a by eqn [6], is the general solution of the vacuum Einstein equation $\text{Ricci}[g] = 0$. Therefore, these equations provide a Hamiltonian form of the Einstein field equation.

Tetrad Formalism

The tetrad formalism, developed by Cartan, Weyl, and Schwinger, has definite advantages with respect to the metric formalism. It allows the coupling of fermion fields to GR and is, therefore, needed to couple the standard model to GR. In the tetrad formalism, the gravitational field is represented by four covariant fields $e^I_\mu(\mathbf{x})$, where $I, J, \dots = 0, 1, 2, 3$ are flat Lorentz indices raised and lowered with the Minkowski metric $\eta_{IJ} = \text{diag}[-1, +1, +1, +1]$. The relation with the metric formalism is given by

$$g_{\mu\nu} = \eta_{IJ} e^I_\mu e^J_\nu$$

In this formulation, GR has an additional local $\text{SO}(3,1)$ gauge invariance, given by local Lorentz transformations on the I indices. The corresponding canonical formalism is usually defined in a gauge in which $e^i_0 = 0$, where $i, j, \dots = 1, 2, 3$ are flat three-dimensional indices raised and lowered with the $\delta_{ij} = \text{diag}[+1, +1, +1]$. In this gauge, the Lorentz group is reduced to the local $\text{SO}(3)$ group of spatial transformations, and the ADM variable are defined by

$$e^I_\mu = \begin{pmatrix} N & N^i \\ 0 & e^i_a \end{pmatrix} \quad [11]$$

where $N^i = e^i_a N^a$. This is equivalent to writing the invariant interval in the form

$$ds^2 = -N^2 dt^2 + (e_{ai} dx^a + N_i dt)(e^i_b dx^b + N^i dt)$$

The reduced canonical variables can be taken to be the field $e^i_a(\mathbf{x})$ that represents the “triad” of the ADM surface, and its conjugate momentum $p^a_i(\mathbf{x})$. Their relation with the three-dimensional metric variables is given by transforming internal indices into tangent indices with the triad field e^i_a and its inverse e^a_i . In particular,

$$q_{ab} = \delta_{ij} e^i_a e^j_b \quad [12]$$

$$p^{ab} = e^{bi} p^a_i \quad [13]$$

Also, for later reference,

$$k^i_a \equiv e^{ib} k_{ab} = \frac{2}{\det e} (p^i_a - \frac{1}{2} e^i_a p) \quad [14]$$

where $p = e^i_a p^a_i$.

The momentum and Hamiltonian constraints are the same as in the ADM formulation, with q_{ab} and p^{ab} expressed in terms of the triad variables. The additional constraint that generates the internal rotations is

$$G_i = \epsilon_{ijk} e^j_a p^{ak} = 0 \quad [15]$$

Ashtekar Formalism

The Ashtekar formalism simplifies the form of the constraints and casts GR in a form having the same kinematics as Yang–Mills theory. With its variants, it is widely used in nonperturbative quantum gravity, in particular in the loop formulation (*see* Loop Quantum Gravity). It can be obtained from the tetrad canonical formalism by the canonical transformation

$$A_a^i = \frac{1}{2} \epsilon_{jk}^i \omega_a^{jk} + i k_a^i \quad [16]$$

$$E_i^a = \det e e_i^a \quad [17]$$

where $\omega^{ij} = \omega_a^{ij} dx^a$ is the (torsion-free) spin connection of the triad 1-form field $e^i = e_a^i dx^a$, determined by the Cartan equation

$$de^i + \omega_k^j \wedge e^k = 0$$

The “electric” field E is real, while the Sen–Ashtekar connection $A^i = A_a^i dx^a$ is complex and satisfies the reality condition

$$A^i + \overline{A^i} = 2\Gamma^i[e] \quad [18]$$

The connection A^i has a simple geometrical interpretation. It is the pullback $A_{ai} = \omega_{a0i}^{(+)}$ on the $t=0$ ADM surface of the self-dual part

$$\omega_{IJ}^{(+)} = \frac{1}{2} \left(\omega_{IJ} - \frac{i}{2} \epsilon_{IJ}^{KL} \omega_{KL} \right)$$

of the four-dimensional torsion free spin connection ω_{μ}^{IJ} determined by the tetrad field e_{μ}^I .

In terms of these fields, the constraint equations can be written in the form

$$G_i = D_a E_i^a = 0 \quad [19]$$

$$C_a = F_{ab}^i E_i^a = 0 \quad [20]$$

$$C = \epsilon_{ijk} F_{ab}^i E^j E^{kb} = 0 \quad [21]$$

where D_a is the covariant derivative and F_{ab} is the curvature defined by the connection A . The first of these constraints is the nonabelian version of the Gauss law [3]: it is the gauge constraint of Yang–Mills theory. The constraints are polynomial in the canonical variables.

These equations are often written using a basis τ_i in the $\mathfrak{su}(2)$ Lie algebra, and defining the $\mathfrak{su}(2)$ connection $A = A^i \tau_i$ and the $\mathfrak{su}(2)$ -valued vector field $E^a = E^{ai} \tau_i$. In terms of these fields the constraints can be written in the form

$$G = D_a E^a = 0$$

$$C_a = \text{tr}[F_{ab} E^a] = 0$$

$$C = \text{tr}[F_{ab} E^a E^b] = 0$$

where the trace is on $\mathfrak{su}(2)$.

A variant of this formalism commonly used in quantum gravity is obtained by replacing [16] with the Barbero connection

$$A_a^i = \frac{1}{2} \epsilon_{jk}^i \omega_a^{jk} + \gamma k_a^i \quad [22]$$

where γ is an arbitrary complex number, called the Immirzi parameter. In terms of this connection, [21] is replaced by

$$C = \epsilon_{ijk} F_{ab}^i E^j E^{kb} + \frac{1 + \gamma^2}{4} \det e (k_{ab} k^{ab} - k^2) = 0$$

where e_a^i and k_{ab} are given as function of E and A by [22] and [17]. The choice $\gamma = 1$, with the constraint [19]–[21], gives the canonical formulation of Euclidean GR.

All the formulations described extend readily to matter couplings. The structure of the constraints remains the same – with additional constraints corresponding to matter gauge invariances, if any. The GR constraints are modified by the addition of matter terms. In particular, the Hamiltonian constraint C and the momentum constraint C_a are modified by the addition of terms determined by the energy density and the momentum density of the matter, respectively. In the Ashtekar formulation, a fermion field modifies the Gauss law constraint by the addition of a torsion term.

Evolution

In the gauge-invariant canonical structure of GR, there is no explicit time flow generated by a Hamiltonian. If the formalism is utilized just in order to express the Einstein equation in first-order canonical form, this is not a difficulty, because evolution in the coordinate time is generated by the constraints. On the other hand, if we are interested in understanding the structure of states, observables, and evolution of GR, the situation appears to be puzzling. An additional complication arises from the fact that virtually no gauge-invariant observable A_{ph} is known explicitly as a function on the phase space. These issues become especially relevant when the canonical formalism is taken as a starting point for quantization. How is physical evolution represented in canonical GR?

The first relevant observation is that the gauge-invariant phase space Γ_{ph} is better understood as a phase space in the sense of Lagrange: namely as the space Γ of the solutions of the equations of motion modulo gauges, rather than a space of instantaneous states. Recall that in GR the notion of “instantaneous state” is rather unnatural.

In the ADM formulation, for instance, an orbit on the constraint surface of GR can be understood as the ensemble of all possible values that the variables

$(q_{ab}(x), p^{ab}(x))$ can take on arbitrary spacelike ADM surfaces embedded in a given solution of the Einstein equation. Motion along the orbit (which has dimension $4 \times \infty^3$) corresponds to arbitrary deformations of the surface.

Physical applications of classical GR deal with relations between “partial observables.” A partial observable is any variable physical quantity that can be measured, even if its value cannot be determined from the knowledge of the physical state. An example of partial observable in nonrelativistic mechanics is given precisely by the nonrelativistic time t . Partial observables are represented in GR as functions on Γ_0 . A physical state in Γ_{ph} determines an orbit in C , and therefore a set of relations between partial observables (see Figure 1). That is, it determines the possible values that the partial observables can take “when” and “where” other partial observables have given values. All physical predictions of classical GR can be expressed in this form.

One of the partial observables can be selected to play the role of a physical clock time, and evolution can be expressed in terms of such clock time. In general, it is difficult – if not impossible – to find a clock time observable in terms of which evolution is a proper conventional Hamiltonian evolution. Matter couplings partially simplify the task. For instance, if the motion of planet Earth is coupled to GR, then proper time along this motion from a significant event on Earth, which is a partial observable, can be a convenient clock time. In pure gravity, the “York time” defined as the trace of the extrinsic curvature $T_Y = k$, on ADM surfaces where k is spatially constant, has been extensively and effectively used as a clock time in formal analysis of the theory. A Hamiltonian that generates evolution in a given clock time T can be formally obtained by solving the Hamiltonian constraint with respect to a momentum P_T conjugate to T . Such “reparametrizations” of the relative evolution of the partial observables can be useful to analyze equations and to help intuition, but they are by no means necessary to have a well-defined interpretation of the theory.

Another possibility to introduce a preferred time flow is to consider asymptotically flat solutions of the field equations. In this case, one can define a nonvanishing Hamiltonian, given by a boundary integral at spacial infinity. This Hamiltonian generates evolution in an asymptotic Minkowski time. This choice is convenient for describing observations performed from a large distance on isolated gravitational systems. Many general-relativistic physical observations do not belong to this category.

Various other techniques to define a fully generally covariant canonical formalism have been

explored. Among these: definitions of the physical symplectic structure directly on the space of the solutions of the field equations; generalization of the initial and final surfaces to boundaries of compact spacetime regions; construction of “evolving constants of motion,” namely families of gauge-invariant observables depending on a clock time parameter; multisymplectic formalisms that treats space and time derivatives on a more equal footing; and others. Many of these techniques are attempts to overcome the unequal way in which time and space dependence are treated in the conventional Hamiltonian formalism.

GR has deeply modified our understanding of space and time. An extension of the canonical formalism of mechanics, compatible with such a modification, is needed, but consensus on the way (or even the possibility) of formulating a fully satisfactory general-relativistic extension of Hamiltonian mechanics is still lacking.

See also: Asymptotic Structure and Conformal Infinity; Constrained Systems; General Relativity: Overview; Loop Quantum Gravity; Quantum Cosmology; Quantum Geometry and its Applications; Spin Foams; Wheeler–De Witt Theory.

Further Reading

- Arnowitt R, Deser S, and Misner CW (1962) The dynamics of general relativity. In: Witten L (ed.) *Gravitation: An Introduction to Current Research*, p. 227. New York: Wiley.
- Ashtekar A (1991) *Non-Perturbative Canonical Gravity*. Singapore: World Scientific.
- Bergmann P (1989) The canonical formulation of general relativistic theories: the early years, 1930–1959. In: Howard D and Stachel J (eds.) *Einstein and the History of General Relativity*. Boston: Birkhäuser.
- Dirac PAM (1950) Generalized Hamiltonian dynamics. *Canadian Journal of Mathematical Physics* 2: 129–148.
- Dirac PAM (1958) The theory of gravitation in Hamiltonian form. *Proceedings of the Royal Society of London, Series A* 246: 333.
- Dirac PAM (1964) *Lectures on Quantum Mechanics*. New York: Belfer Graduate School of Science, Yeshiva University.
- Gotay MJ, Isenberg J, Marsden JE, and Montgomery R (1998) Momentum maps and classical relativistic fields. Part 1: Covariant field theory. Archives: physics/9801019.
- Hanson A, Regge T, and Teitelboim C (1976) *Constrained Hamiltonian Systems*. Rome: Academia Nazionale dei Lincei.
- Henneaux M and Teitelboim C (1972) *Quantization of Gauge Systems*. Princeton: Princeton University Press.
- Isham CJ (1993) Canonical quantum gravity and the problem of time. In: Ibert LA and Rodriguez MA (eds.) *Recent Problems in Mathematical Physics, Salamanca*, Dordrecht: Kluwer Academic.
- Lagrange JL (1808) *Mémoires de la première classe des sciences mathématiques et physiques*. Paris: Institute de France.
- Rovelli C (2004) *Quantum Gravity*. Cambridge: Cambridge University Press.
- Souriau JM (1969) *Structure des Systemes Dynamiques*. Paris: Dunod.

Capacities Enhanced by Entanglement

P Hayden, McGill University, Montreal, QC, Canada

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Shared entanglement between a sender and receiver can significantly improve the usefulness of a quantum channel for the communication of either classical or quantum data. Superdense coding and teleportation provide the most well-known examples of this improvement; free entanglement doubles the classical capacity of a noiseless quantum channel and makes it possible for a noiseless classical channel to send quantum data. In fact, the entanglement-assisted classical and quantum capacities of a quantum channel are in many senses simpler and better behaved than their unassisted counterparts (Holevo 1998, Schumacher and Westmoreland 1997, Devetak 2005). Most importantly, these capacities can be calculated using simple formulas and finite optimization procedures (Bennett *et al.* 1999, 2002). No such finite procedure is known for either of the unassisted capacities. Moreover, the entanglement-assisted classical and quantum capacities are related by a simple factor of 2. The unassisted capacities, in contrast, have completely different formulas. In fact, the simple factor of 2 generalizes to a statement known as the quantum reverse Shannon theorem, which governs the rate at which one quantum channel can simulate another (Bennett *et al.* 2005). The answer is given by the ratio of the entanglement-assisted capacities.

Notation

Quantum systems will be denoted by A , B , and so on as well as their variants such as A' and \hat{A} . The choice of letter will generally indicate which party holds a given system, with A reserved for the sender, Alice, and B for the receiver, Bob. Given a quantum system C , $C^{\otimes n}$ will often be written as C^n . These symbols will be used to denote both the Hilbert space of the quantum system and the set of density operators on that system. Thus, a quantum channel $\mathcal{N}: A' \rightarrow B$ refers to a trace-preserving, completely positive (TPCP) map from the operators on the Hilbert space of A' to those of B . id^C refers to the identity channel on C . The map $\mathcal{N} \otimes \text{id}^C$ will frequently be abbreviated to \mathcal{N} in order to simplify long expressions. Likewise, the density operator $|\varphi\rangle\langle\varphi|$ of a pure quantum state $|\varphi\rangle$ will be abbreviated to φ . π^C will refer to the maximally

mixed state on C and π_d to the maximally mixed state on a specified d -dimensional quantum system.

For a given quantum state φ^{AB} on the composite system AB , $\varphi^A = \text{tr}_B \varphi^{AB}$ and

$$H(A)_\varphi = H(\varphi^A) = -\text{tr}(\varphi^A \log_2 \varphi^A) \quad [1]$$

is the von Neumann entropy of φ^A , while

$$H(A|B)_\varphi = -I_c(A|B) = H(AB)_\varphi - H(B)_\varphi \quad [2]$$

is its conditional entropy and

$$I(A; B)_\varphi = H(A)_\varphi + H(B)_\varphi - H(AB)_\varphi \quad [3]$$

its mutual information.

Entanglement-Assisted Classical and Quantum Capacities

The entanglement-assisted classical capacity of a quantum channel $\mathcal{N}: A' \rightarrow B$ is the optimal rate at which classical information can be communicated through the channel while in addition making use of an unlimited number of maximally entangled states.

The formal definition proceeds as follows. Alice and Bob are assumed to share nS ebits in the form of a maximally entangled state $|\Phi\rangle^{AB}$ of Schmidt rank 2^{nS} . Conditioned on her message $m \in \{1, 2, \dots, 2^{nR}\}$, Alice will apply an encoding operation $\mathcal{E}_m: \hat{A} \rightarrow A'^n$. Bob's decoding is given by a POVM $\{\Lambda_m\}_{m=1}^{2^{nR}}$ on the composite system $\hat{B}B^n$. The procedure is said to have maximum probability of error ϵ if

$$\max_m \text{tr}[\Lambda_m(\mathcal{N}^{\otimes n} \circ \mathcal{E}_m)(\Phi)] \geq 1 - \epsilon \quad [4]$$

These elements, illustrated in **Figure 1**, consisting of the shared entanglement, as well as the encoding and decoding operations meeting the criterion of eqn [4], are called a $(2^{nR}, 2^{nS}, n, \epsilon)$ entanglement-assisted classical code for the channel \mathcal{N} . A rate R is said to be achievable if there exists a choice of $S \geq 0$ and a sequence of entanglement-assisted classical codes $(2^{nR}, 2^{nS}, n, \epsilon_n)$ with $\epsilon_n \rightarrow 0$. The entanglement-assisted

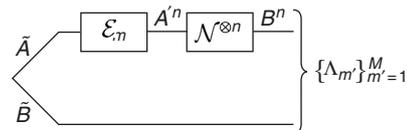


Figure 1 Circuit representation of the elements of an entanglement-assisted classical code for the channel \mathcal{N} . Alice encodes message m by applying the operation \mathcal{E}_m to her half of the shared entanglement. Bob decodes by applying the POVM $\{\Lambda_m\}$ on the output of the channel and his half of the shared entanglement.

classical capacity $C_E(\mathcal{N})$ of \mathcal{N} is defined to be the supremum over all achievable rates.

Theorem 1 (Bennett *et al.* 1999, 2002). *The entanglement-assisted classical capacity C_E of a quantum channel $\mathcal{N}: A' \rightarrow B$ is given by*

$$C_E(\mathcal{N}) = \max_{\sigma} I(A; B)_{\sigma} \quad [5]$$

where the maximization is over states $\sigma^{AB} = \mathcal{N}(\varphi^{AA'})$ arising from the channel by acting on the A' half of any pure state $|\varphi\rangle^{AA'}$.

The theorem bears a strong formal resemblance to Shannon's noisy coding theorem for the classical capacity of a classical noisy channel. There the capacity formula is also given by an optimization of the mutual information, but over joint distributions between the input and output alphabets arising from the action of the channel. Such a joint distribution cannot exist in general for a quantum channel because the no-cloning theorem excludes the possibility of the input and output existing simultaneously. Equation [5] instead refers to a natural substitute for the joint input–output distribution: a quantum state arising from the quantum channel acting on half of an entangled pure state.

Another point worth stressing is that, unlike the known formulas for the unassisted classical and quantum capacities of a quantum channel, eqn [5] refers to only a single use of \mathcal{N} instead of the limit of many uses, $\mathcal{N}^{\otimes n}$. The formula can therefore readily be used to evaluate C_E for any channel of interest.

Consider, for example, the d -dimensional depolarizing channel

$$\mathcal{D}_p(\rho) = (1-p)\rho + p\pi_d \quad [6]$$

that with probability p completely randomizes the input but otherwise leaves the input invariant. For such channels, the maximum is achieved by choosing a maximally entangled state for $|\varphi\rangle^{AA'}$, yielding

$$C_E(\mathcal{D}_p) = 2 \log_2 d - h_d \left(1 - p \frac{d^2 - 1}{d^2} \right) \quad [7]$$

where for any $0 \leq q \leq 1$ and integer $r \geq 1$,

$$h_r(q) = -q \log_2 q - (1-q) \times \log_2 \left(\frac{1-q}{r-1} \right) \quad [8]$$

is the Shannon entropy of the distribution $(q, (1-q)/(r-1), \dots, (1-q)/(r-1))$.

Entanglement assistance also simplifies the relationship between the classical and quantum

capacities of a channel. Proceeding as before to formally define the quantum capacity, Alice and Bob are again assumed to share a maximally entangled state $|\Phi\rangle^{AB}$ of Schmidt rank 2^{nS} . Alice's encoding operation will be a TPCP map $\mathcal{E}: \hat{A}\hat{A} \rightarrow A^n$ acting on an input system \hat{A} and her half of the shared entanglement, \hat{A} . Bob's decoding will likewise be a TPCP map $\mathcal{D}: \hat{B}B^n \rightarrow \hat{B}$ acting on the output of the channel, B^n , and his half of the shared entanglement, \hat{B} . \hat{A} and \hat{B} are assumed to be isomorphic quantum systems of some fixed dimension 2^{nQ} . The procedure is said to have subspace fidelity $1 - \epsilon$ if

$$\langle \varphi | \left(\mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{E} \right) \left(\Phi^{\hat{A}\hat{B}} \otimes \varphi^{\hat{A}} \right) | \varphi \rangle^{\hat{B}} \geq 1 - \epsilon \quad [9]$$

for all $|\varphi\rangle^{\hat{A}} \in \hat{A}$. These elements, illustrated in Figure 2, are together called a $(2^{nQ}, 2^{nS}, n, \epsilon)$ entanglement-assisted quantum code for the channel \mathcal{N} . A rate Q is said to be achievable if there exists a choice of $S \geq 0$ and a sequence of entanglement-assisted quantum codes $(2^{nR}, 2^{nS}, n, \epsilon_n)$ with $\epsilon_n \rightarrow 0$. The entanglement-assisted quantum capacity $Q_E(\mathcal{N})$ of \mathcal{N} is defined to be the supremum over all achievable rates.

There is considerable freedom in the definition of the entanglement-assisted quantum capacity. It could, for example, be defined as the largest amount of maximal entanglement that can be generated using the channel, minus the entanglement consumed during the protocol itself. Alternatively, the fidelity criterion eqn [9] could be strengthened to require that $\mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}$ preserve not only pure states on \hat{A} but any entanglement between \hat{A} and a reference system. All of these variants yield the same capacity formula:

$$Q_E(\mathcal{N}) = \frac{1}{2} C_E(\mathcal{N}) \quad [10]$$

This equivalence is a direct consequence of the existence of the teleportation and superdense coding protocols. When maximal entanglement is available, teleportation converts the ability to send classical data into the ability to send quantum data at half the classical rate. Conversely, by consuming

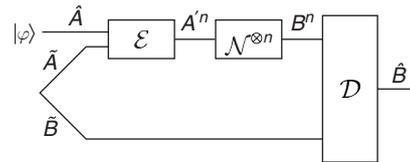


Figure 2 Circuit representation of the elements of an entanglement-assisted quantum code for the channel \mathcal{N} . \mathcal{E} is Alice's encoding operation, which acts on both her input state and her half of the shared entanglement. Bob decodes using a quantum operation \mathcal{D} acting on the output of the channel and his half of the shared entanglement.

maximal entanglement, superdense coding converts the ability to send quantum data into the ability to send classical data at double the quantum rate.

Sketch of Proof

The proof of a capacity theorem can usually be broken into two parts, achievability and optimality. The achievability part demonstrates the existence of a sequence of codes reaching the prescribed rate while the optimality part shows that it is impossible to do better.

The main idea in the achievability proof can be understood by studying the special case where $\varphi^{A'} = \pi^{A'}$. Let $d_{A'} = \dim A'$ and $\{U_j\}_{j=1}^{d_{A'}^2}$ be a set of Weyl operators for A^m . The relevant property of these operators is that averaging over them implements the constant map: for all density operators ρ ,

$$\frac{1}{d_{A'}^2} \sum_{j=1}^{d_{A'}^2} U_j \rho U_j^\dagger = \pi^{A^m} \quad [11]$$

Consider the state σ_j that arises if Alice acts with U_j on the A^m half of a rank- $d_{A'}^2$ maximally entangled state $|\varphi\rangle^{AA^m}$ and then sends the A^m half of the resulting state through \mathcal{N} . (Note that here A^m also plays the role of \tilde{A} .) The entropy of the resulting state is

$$H(\sigma_j) = H(\mathcal{N}((U_j \otimes I_{\tilde{B}})\varphi(U_j^\dagger \otimes I_{\tilde{B}}))) \quad [12]$$

$$= H(\mathcal{N}(\varphi)) \quad [13]$$

since U_j does not change the local density operator on A^m .

On the other hand, if Alice selects a value of j from the uniform distribution, then the resulting average input state to the channel will be

$$\pi^{A^m} \otimes \pi^A = \varphi^{A^m} \otimes \varphi^A \quad [14]$$

and the corresponding average output state will be $\mathcal{N}(\varphi^{A^m}) \otimes \varphi^A$, which has entropy

$$H(\mathcal{N}(\varphi^{A^m})) + H(\varphi^A) \quad [15]$$

Therefore, the Holevo quantity of the ensemble of output states, defined as the entropy of the average state minus the average of the entropies of the individual output states, will be equal to

$$H(\varphi^A) + H(\mathcal{N}(\varphi^{A^m})) - H(\mathcal{N}(\varphi^{AA^m})) \quad [16]$$

This is precisely the quantity $I(A; B)_\sigma$ for the state $\mathcal{N}(\varphi^{AA^m})$ since the channel \mathcal{N} transforms the A^m system into B . Moreover, if Bob is given the A part of the maximally entangled state, then this is the Holevo

quantity of an ensemble of states that can be produced by Alice acting on half of a shared entangled state and then sending her half through the channel. Invoking the Holevo–Schumacher–Westmoreland (HSW) theorem for the classical capacity (Holevo 1998, Schumacher and Westmoreland 1997) therefore completes the proof; using coding, the Holevo quantity is an achievable communication rate.

The proof that eqn [5] is optimal involves a series of entropy manipulations similar to the optimality proofs for the unassisted classical and quantum capacities. From the point of view of quantum information, the truly unusual part of the proof is the demonstration that it is unnecessary to consider multiple copies of \mathcal{N} (Cerf and Adami 1997). Specifically, let

$$f(\mathcal{N}) = \max_{\sigma} I(A; B)_\sigma \quad [17]$$

where the maximization is defined as in Theorem 1. Techniques analogous to those used for the unassisted capacities yield the upper bound

$$C_E(\mathcal{N}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} f(\mathcal{N}^{\otimes n}) \quad [18]$$

Unlike the unassisted case, however, a relatively easy argument shows that

$$f(\mathcal{N}_1 \otimes \mathcal{N}_2) = f(\mathcal{N}_1) + f(\mathcal{N}_2) \quad [19]$$

(The analogous statement is an important conjecture for the classical capacity and is known to be false for the quantum capacity (DiVincenzo *et al.* 1998).) As a result, $C_E(\mathcal{N}) \leq f(\mathcal{N})$, which is the optimality part of Theorem 1.

To see the origin of eqn [19], it will be helpful to invoke Stinespring's theorem to write $\mathcal{N}_j = \text{tr}_{E_j} \mathcal{U}_j^{B_j E_j}$, where $\mathcal{U}_j: A_j' \rightarrow B_j E_j$ is an isometry. Fix a state $|\varphi\rangle^{AA_1' A_2'}$ and let $\sigma = (\mathcal{U}_1 \otimes \mathcal{U}_2)(\varphi)$. Equation [19] follows from the fact that

$$I(A; B_1 B_2)_\sigma \leq I(AB_2 E_2; B_1)_\sigma + I(AB_1 E_1; B_2)_\sigma \quad [20]$$

Simply redefining A to be $AB_2 E_2$ shows that the first term of the right-hand side is upper bounded by $f(\mathcal{N}_1)$. The second term, likewise, is upper bounded by $f(\mathcal{N}_2)$. Equation [20] is itself equivalent to the inequality

$$\begin{aligned} & H(B_1 B_2 | E_1 E_2)_\sigma + H(B_1 B_2)_\sigma \\ & \leq H(B_1 | E_1)_\sigma + H(B_2 | E_2)_\sigma \\ & \quad + H(B_1)_\sigma + H(B_2)_\sigma \end{aligned} \quad [21]$$

The inequality $H(B_1 B_2)_\sigma \leq H(B_1)_\sigma + H(B_2)_\sigma$ holds by the subadditivity of the von Neumann entropy.

Repeated applications of the strong subadditivity inequality, moreover, lead to the inequality

$$H(B_1B_2|E_1E_2)_\sigma \leq H(B_1|E_1)_\sigma + H(B_2|E_2)_\sigma \quad [22]$$

Together, they prove eqn [20] and, thence, eqn [19]. The intuitive meaning of this “single-letterization” is unclear, but regardless, it is interesting to note that the proof involved invoking a pair of purifying environment systems, E_1 and E_2 , and studying the entropy relationships between the true outputs of the channel and the environment’s share.

The Quantum Reverse Shannon Theorem

A strong argument can be made that the entanglement-assisted capacity of a quantum channel is the most important capacity of that channel and that all the other capacities are, in some sense, of less significance. The fact that it is unnecessary to distinguish between the classical and quantum entanglement-assisted capacities because they are related by a factor of 2 is a hint in that direction, as is the simple, single-letter formula for $C_E(\mathcal{N})$.

A more general argument can be made by considering the problem of having one channel simulate another. Indeed, the quantum capacity of a quantum channel is simply the optimal rate at which that channel can simulate the noiseless channel id_2 on a single qubit. Likewise, the classical capacity of a quantum channel is its optimal rate for simulation of a qubit dephasing channel

$$\rho \mapsto |0\rangle\langle 0|\rho|0\rangle\langle 0| + |1\rangle\langle 1|\rho|1\rangle\langle 1| \quad [23]$$

In this spirit, the fact that $C_E(\mathcal{N}) = 2Q_E(\mathcal{N})$ can be re-expressed in the form

$$Q_E(\mathcal{N}) = \frac{C_E(\mathcal{N})}{C_E(\text{id}_2)} \quad [24]$$

Equivalently, when entanglement is free, the optimal rate at which \mathcal{N} can simulate a noiseless qubit channel is given by the ratio between the entanglement-assisted classical capacities of \mathcal{N} and id_2 . The quantum reverse Shannon theorem generalizes this statement to the simulation of arbitrary channels in the presence of free entanglement.

Suppose that Alice and Bob would like to use $\mathcal{N}_1 : A' \rightarrow B$ to simulate another channel $\mathcal{N}_2 : A' \rightarrow B$. Fix an input state $\varphi^{A'}$ and let $|\varphi\rangle^{AA''}$ be a purification of $(\varphi^{A'})^{\otimes n}$. As always, assume that Alice and Bob share a maximally entangled state $|\Phi\rangle^{AB}$ of Schmidt rank 2^{nS} . Alice’s encoding operation will be a TPCP map $\mathcal{E} : \tilde{A}A'' \rightarrow A''$ acting on n copies of the input system A' and her half of the shared entanglement, \tilde{A} . Bob’s

decoding will likewise be a TPCP map $\mathcal{D} : B''\tilde{B} \rightarrow B''$ acting on m copies of the output of the channel, and his half of the shared entanglement, \tilde{B} . This procedure is said to ϵ -simulate $\mathcal{N}_2^{\otimes n}$ on $(\varphi^{A'})^{\otimes n}$ if

$$F\left(\mathcal{N}_2^{\otimes n}(\varphi^{AA''}), (\mathcal{D} \circ \mathcal{N}_1^{\otimes m} \circ \mathcal{E})(\Phi^{\tilde{A}\tilde{B}} \otimes \varphi^{AA''})\right) \geq 1 - \epsilon \quad [25]$$

where F is the mixed state fidelity $F(\rho, \sigma) = (\text{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}})^2$. The entire procedure, illustrated in Figure 3, is said to be a $(2^{nS}, m, n, \epsilon)$ entanglement-assisted simulation of \mathcal{N}_2 by \mathcal{N}_1 . A rate R , measured in copies of \mathcal{N}_2 per copy of \mathcal{N}_1 , is said to be achievable for $\varphi^{A'}$ if there exists a choice of $S \geq 0$ and a sequence of $(2^{nS}, m_n, n, \epsilon_n)$ entanglement-assisted simulations with $n/m_n \rightarrow R$ while $\epsilon_n \rightarrow 0$.

The quantum reverse Shannon theorem states that the entanglement-assisted capacity completely governs the achievable simulation rates.

Theorem 2 (Winter 2004, Bennett *et al.*). *Given two channels $\mathcal{N}_1 : A' \rightarrow B$ and $\mathcal{N}_2 : A' \rightarrow B$, R is an achievable simulation rate for \mathcal{N}_2 by \mathcal{N}_1 and all input states $\varphi^{A'}$ if and only if*

$$R \leq \frac{C_E(\mathcal{N}_1)}{C_E(\mathcal{N}_2)} \quad [26]$$

Note that the form of eqn [26] ensures that the simulation is asymptotically reversible: if a channel \mathcal{N}_1 is used to simulate \mathcal{N}_2 and the simulation is then used to simulate \mathcal{N}_1 again, then the overall rate becomes

$$\frac{C_E(\mathcal{N}_1) C_E(\mathcal{N}_2)}{C_E(\mathcal{N}_2) C_E(\mathcal{N}_1)} = 1 \quad [27]$$

Thus, in the presence of free entanglement and for a known input density operator of the form $(\varphi^{A'})^{\otimes n}$, a single parameter, the entanglement-assisted classical capacity, suffices to completely characterize the asymptotic properties of a quantum channel.

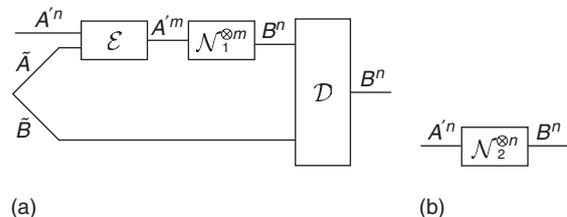


Figure 3 Circuit representation of an entanglement-assisted simulation of \mathcal{N}_2 by \mathcal{N}_1 . (a) The simulation circuit, with Alice’s encoding operation \mathcal{E} acting on n copies of A' and Bob’s decoding operation producing n copies of B . (b) The circuit that the protocol is intended to simulate. As stated, the quantum reverse Shannon theorem allows the simulation circuit to depend on the density operator of the input state restricted to A'' .

Moreover, since two channels that are asymptotically equivalent without free entanglement will surely remain equivalent if free entanglement is permitted, eqn [26] gives essentially the only possible nontrivial, single-parameter asymptotic characterization of quantum channels. This is the sense in which the entanglement-assisted capacity should be regarded as the most important capacity of a quantum channel.

The proof of the quantum reverse Shannon theorem is quite involved, but some of its features can be understood without much work. First, note that by the optimality statement of the entanglement-assisted classical capacity, the desired simulation can exist only if eqn [26] holds. Otherwise, composing the simulation of \mathcal{N}_2 by \mathcal{N}_1 with a sequence of codes achieving $C_E(\mathcal{N}_2)$ would result in a sequence of codes beating the capacity formula for \mathcal{N}_1 .

Similarly, note that one method to simulate a channel \mathcal{N}_1 using \mathcal{N}_2 is to first use \mathcal{N}_2 to simulate the noiseless channel and then use the simulated noiseless channel to simulate \mathcal{N}_1 . Since the achievable rates for the first step are characterized by the entanglement-assisted capacity theorem, proving the achievability part of Theorem 2 reduces to finding protocols for simulating a general noisy quantum channel \mathcal{N}_2 by a noiseless one. That perhaps sounds like a strange goal, but nonetheless is the difficult part of the quantum reverse Shannon theorem.

It is likely that the quantum reverse Shannon theorem can be extended to cover other types of inputs than the known tensor power states $(\varphi^{A'})^{\otimes n}$. The most desirable form of the theorem would be one valid for all possible input density operators on $A'^{\otimes n}$, providing a single simulation procedure dependent only on the channels and not the input state. It is known that without modifying the form of the free entanglement, this most ambitious form of the theorem fails, but it is conjectured that the full-strength theorem does hold provided very large amounts of entanglement are supplied in the form of the so-called embezzling states (van Dam and Hayden 2003).

Relationships between Protocols

There is another sense in which the entanglement-assisted capacity can be viewed as the fundamental capacity of a quantum channel: an efficient protocol for achieving the entanglement-assisted capacity can be converted into protocols achieving the unassisted quantum and classical capacities, or at least very close variants thereof.

An efficient protocol in this case refers to one that does not waste entanglement. Suppose that $\mathcal{N} : A' \rightarrow B$

can be written $\text{tr}_E \mathcal{U}^{BE}$ for some isometry \mathcal{U}^{BE} . Let $|\varphi\rangle^{AA'}$ be a pure state and $|\sigma\rangle^{ABE} = \mathcal{U}^{BE}|\varphi\rangle^{AA'}$ the corresponding purified channel output state. Careful analysis of the entanglement-assisted classical communication protocol achieving the rate $I(A; B)_\sigma$ leads to an entanglement-assisted quantum communication protocol consuming entanglement at the rate $(1/2)I(A; E)_\sigma$ ebits per use of \mathcal{N} and yielding communication at the rate of $(1/2)I(A; B)_\sigma$ qubits per use \mathcal{N} . The protocol achieving this goal is known as the “father” (Devetak *et al.* 2004).

If the entanglement consumed in the father were actually supplied by quantum communication from Alice to Bob, then the net rate of quantum communication produced by the resulting protocol would be $(1/2)I(A; B)_\sigma - (1/2)I(A; E)_\sigma$ qubits from Alice to Bob, that is, the total produced minus the total consumed.

This quantity, how much more information B has about A than E does, can be simplified using an interesting identity. Since $|\sigma\rangle^{ABE}$ is pure,

$$I(A; E)_\sigma = H(A)_\sigma + H(E)_\sigma - H(AE)_\sigma \quad [28]$$

$$= H(A)_\sigma + H(AB)_\sigma - H(B)_\sigma \quad [29]$$

Expanding $I(A; B)_\sigma$ and canceling terms then reveals that

$$\begin{aligned} \frac{1}{2}I(A; B) - \frac{1}{2}I(A; E) &= -H(A|B)_\sigma \\ &= I_c(A)B)_\sigma \end{aligned} \quad [30]$$

where the function I_c is known as the coherent information. After optimizing over input states and multiple channel uses, this is precisely the formula for the unassisted quantum capacity of a quantum channel (Devetak 2005). Thus, the net rate of qubit communication for the protocol derived from the father exactly matches the rates necessary to achieve the unassisted quantum capacity. The only caveat is that the protocol derived from the father uses quantum communication catalytically, meaning that some communication needs to be invested in order to get a gain of $I_c(A)B$. For the unassisted quantum capacity, no investment is necessary. Nonetheless, detailed analysis of the situation reveals that the amount of catalytic communication required can be reduced to an amount sublinear in the number of channel uses, meaning the rate of required investment can be made arbitrarily small. In this sense, the father protocol essentially generates the optimal protocols for the unassisted quantum capacity.

Protocols achieving the unassisted classical capacity can be constructed in a similar way. In this case, one starts from an ensemble $\mathcal{E} = \{p_j, \mathcal{N}(\psi_j^{A'})\}$ of states generated by the channel. Achievability of

the unassisted classical capacity formula follows from achievability of rates of the form

$$\chi(\mathcal{E}) = H\left(\sum_j p_j \mathcal{N}(\psi_j^{A'})\right) - \sum_j p_j H\left(\mathcal{N}(\psi_j^{A'})\right) \quad [31]$$

for arbitrary ensembles of output states. Consider the channel

$$\tilde{\mathcal{N}}(\rho) = \sum_j \langle j|\rho|j\rangle \cdot \mathcal{N}(\psi_j) \quad [32]$$

and input state $|\varphi\rangle^{AA'} = \sum_j \sqrt{p_j} |j\rangle^A |j\rangle^{A'}$. If $\sigma = \tilde{\mathcal{N}}(\varphi)$, then $I(A; B)_\sigma$ is equal to $\chi(\mathcal{E})$. Thus, there are protocols consuming entanglement that achieve the classical communications rate $\chi(\mathcal{E})$ for the modified channel $\tilde{\mathcal{N}}$. Because the channel $\tilde{\mathcal{N}}$ includes an orthonormal measurement which destroys all entanglement between A and B , however, it can be argued that any entanglement used in such a protocol could be replaced by shared randomness, which could then in turn be eliminated by a standard derandomization argument. The net result is a procedure for choosing rate $\chi(\mathcal{E})$ codes for the channel \mathcal{N} consisting of states of the form $\psi_{j_1} \otimes \cdots \otimes \psi_{j_n}$, which is the essence of the achievability proof for the unassisted classical capacity.

This may seem like an unnecessarily cumbersome and even circular approach to the unassisted classical capacity given that the proof sketched above for the entanglement-assisted classical capacity itself invokes the unassisted result in the form of the HSW theorem. The approach becomes more satisfying when one learns that simple and direct proofs of the father protocol exist that completely bypass the HSW theorem (Abeyesinghe *et al.* 2005).

Thus, the entanglement-assisted communication protocols can be easily transformed into their unassisted analogs, confirming the central place of entanglement-assisted communication in quantum information theory.

Acknowledgments

The author is grateful to the inventors of the quantum reverse Shannon theorem for letting him

discuss their results prior to their publication and to Jon Yard for a careful reading of the manuscript. This work has been supported by the Canadian Institute for Advanced Research, the Canada Research Chairs program, and Canada's NSERC.

See also: Capacity for Quantum Information; Channels in Quantum Information Theory; Entanglement; Finite Weyl Systems; Quantum Channels: Classical Capacity; Quantum Entropy.

Further Reading

- Abeyesinghe A, Devetak I, Hayden P, and Winter A (2005) Fully quantum Slepian–Wolf (in preparation).
- Bennett CH, Devetak I, Harrow AW, Shor PW, and Winter A (2005) The quantum Reverse Shannon Theorem (in preparation).
- Bennett CH, Shor PW, Smolin JA, and Thapliyal AV (1999) Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters* 83: 3081 (arXiv.org:quant-ph/9904023).
- Bennett CH, Shor PW, Smolin JA, and Thapliyal AV (2002) Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory* 48(10): 2637 (arXiv.org:quant-ph/0106052).
- Cerf N and Adami C (1997) Von Neumann capacity of noisy quantum channels. *Physical Review A* 56: 3470 (arXiv.org:quant-ph/9609024).
- Devetak I (2005) The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory* 51(1): 44 (arXiv.org/0304127).
- Devetak I, Harrow AW, and Winter A (2004) A family of quantum protocols. *Physical Review Letters* 93: 230504 (arXiv.org:quant-ph/0308044).
- DiVincenzo DP, Smolin JA, and Shor PW (1998) Quantum channel capacity of very noisy channels. *Physical Review A* 57: 830 (arXiv.org:quantph/9706061).
- Holevo AS (1998) The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory* 44: 269–273.
- Schumacher B and Westmoreland MD (1997) Sending classical information via noisy quantum channels. *Physical Review A* 56: 131–138.
- van Dam W and Hayden P (2003) Universal entanglement transformation without communication. *Physical Review A* 67: 060302 (arXiv.org:quant-ph/0201041).
- Winter A (2004) Extrinsic and intrinsic data in quantum measurements: asymptotic convex decomposition of positive operator valued measures. *Communications in Mathematical Physics* 244(1): 157 (arXiv.org:quantph/0109050).

Capacity for Quantum Information

D Kretschmann, Technische Universität Braunschweig, Braunschweig, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Any processing of quantum information, be it storage or transfer, can be represented as a quantum channel: a completely positive and trace-preserving map that transforms states (density matrices) on the sender's end of the channel into states on the receiver's end. Very often, the channel S that sender and receiver (conventionally called Alice and Bob, respectively) would like to implement is not readily available, typically due to detrimental noise effects, limited technology, or insufficient funding. They may then try to simulate S with some other channel T , which they happen to have at their disposal. The quantum channel capacity $Q(T, S)$ of T with respect to S quantifies how well this simulation can be performed, in the limit of long input strings, so that Alice and Bob can take advantage of collective pre- and post-processing (cf. Figure 1). Higher capacities may result if Alice and Bob are allowed to use additional resources in the process, such as classical side channels or a bunch of maximally entangled pairs shared between them.

Quantum capacity thus gives the ultimate benchmarks for the simulation of one quantum channel by another and for the optimal use of auxiliary resources. Together with the compression rate of a quantum source (see Source Coding in Quantum

Information Theory), it lies at the heart of quantum information theory.

In a very typical scenario, Alice and Bob would like to implement the ideal (noiseless) quantum channel $S = \text{id}$: they are interested in sending quantum states undistorted over some distance, or want to store them safely for some period of time, so that all the precious quantum correlations are preserved. The capacity $Q(T) \equiv Q(T, \text{id})$ is then the maximal number of qubit transmissions per use of the channel, taken in the limit of long messages and using collective encoding and decoding schemes asymptotically eliminating all transmission errors. This is what is generally called the *quantum capacity* of the channel T , and it is our main focus in this article. Little is known so far about the quantum capacity for the simulation of other (nonideal) channels (cf. the section "Related capacities").

In remarkable contrast to the classical setting, quantum channel capacities are very much affected by additional resources. This leads to unexpected and fascinating applications such as teleportation and dense coding. But it also results in a bewildering variety of inequivalent channel capacities, which still hold many challenges for future research.

Notation

A quantum channel which transforms input systems on a Hilbert space \mathcal{H}_A into output systems on a (possibly different) Hilbert space \mathcal{H}_B is represented (in Schrödinger picture) by a completely positive and trace-preserving linear map $T: \mathcal{B}_*(\mathcal{H}_A) \rightarrow \mathcal{B}_*(\mathcal{H}_B)$, where $\mathcal{B}_*(\mathcal{H})$ denotes the space of trace class operators on the Hilbert space \mathcal{H} (see Channels in Quantum Information Theory). We write \mathcal{A} instead of $\mathcal{B}_*(\mathcal{H}_A)$ to streamline the presentation, and \mathcal{A}^n for the n -fold tensor product $\mathcal{B}_*(\mathcal{H}_A)^{\otimes n}$.

It is evident that the definition of channel capacity requires the comparison of different quantum channels. A suitable distance measure is the *norm of complete boundedness* (or cb-norm, for short), denoted by $\|\cdot\|_{\text{cb}}$. For two channels T and S , the distance $(1/2)\|T - S\|_{\text{cb}}$ can be defined as the largest difference between the overall probabilities in two statistical quantum experiments differing only by exchanging one use of S by one use of T . These experiments may involve entangling the systems on which the channels act with arbitrary further systems; hence the cb-norm remains a valid distance-measure if the given channel is only part of a larger system. Equivalently, we may set $\|T\|_{\text{cb}} := \sup_n \|T \otimes \text{id}_n\|$, where $\|R\| := \sup_{\|\varrho\|_1 \leq 1} \|R(\varrho)\|_1$

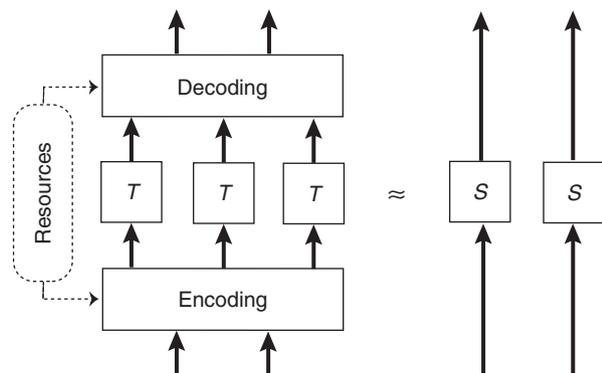


Figure 1 Equipped with collective encoding and decoding operations (and perhaps some auxiliary resources), $n=3$ instances of the channel T simulate $m=2$ instances of the channel S . The transmission rate of the above scheme is $2/3$. Capacity is the largest such rate, in the limit of long messages and optimal encoding and decoding.

denotes the norm of linear operators, and $\|\varrho\|_1 := \text{tr}\sqrt{\varrho^*\varrho}$ is the trace norm on the space of trace-class operators $\mathcal{B}_*(\mathcal{H})$.

We use base two logarithms throughout, and we write $\text{ld}x := \log_2 x$ and $\exp_2 x := 2^x$.

Quantum Channel Capacity

The intuitive concept underlying quantum channel capacity is made rigorous in the following definition:

Definition 1 A positive number R is called *achievable rate* for the quantum channel $T: \mathcal{A} \rightarrow \mathcal{B}$ with respect to the quantum channel $S: \mathcal{A}' \rightarrow \mathcal{B}'$ iff for any pair of integer sequences $(n_\nu)_{\nu \in \mathbb{N}}$ and $(m_\nu)_{\nu \in \mathbb{N}}$ with $\lim_{\nu \rightarrow \infty} n_\nu = \infty$ and $\lim_{\nu \rightarrow \infty} \frac{m_\nu}{n_\nu} \leq R$ we have

$$\liminf_{\nu \rightarrow \infty} \inf_{D,E} \|DT^{\otimes n_\nu} E - S^{\otimes m_\nu}\|_{\text{cb}} = 0 \quad [1]$$

the infimum taken over all encoding channels E and decoding channels D with suitable domain and range. The *channel capacity* $Q(T, S)$ of T with respect to S is defined to be the supremum of all achievable rates. The *quantum capacity* is the special case $Q(T) := Q(T, \text{id}_2)$, with id_2 being the ideal qubit channel.

In this article, we mainly concentrate on channels between finite-dimensional systems. This is enough to bring out the basic ideas. Many of the concepts and results discussed here can be generalized to *Gaussian channels*, which play a central role as building blocks for quantum optical communication lines (Holevo and Werner 2001, Eisert and Wolf).

There is considerable freedom in the definition of quantum channel capacity, at least for ideal reference channels (Kretschmann and Werner 2004). In particular, the encoding channels E in eqn [1] may always be restricted to isometric embeddings.

In addition, it is not necessary to check an infinite number of pairs of sequences $(n_\nu)_{\nu \in \mathbb{N}}$ and $(m_\nu)_{\nu \in \mathbb{N}}$ when testing a given rate R , as Definition 1 would suggest. Instead, it is enough to find one such pair which achieves the rate R infinitely often, $\lim_{\nu \rightarrow \infty} m_\nu/n_\nu = R$.

Without affecting the capacity, the cb-norm $\|T\|_{\text{cb}}$ may be replaced by the unstabilized operator norm $\|T\|$ or by fidelity measures, which are in general much easier to compute. In particular, one might choose the *minimum fidelity*,

$$F(T) := \min_{\|\psi\|=1} \langle \psi | T(|\psi\rangle\langle\psi|) | \psi \rangle \quad [2]$$

or even the *average fidelity*,

$$\bar{F}(T) := \int \langle \psi | T(|\psi\rangle\langle\psi|) | \psi \rangle d\psi \quad [3]$$

Unfortunately, this equivalence is restricted to capacities with noiseless reference channel $S = \text{id}$. In the vicinity of other (nonideal) channels, equivalence of the stabilized and unstabilized error criteria may be lost. Of course, the comparison of channels is ultimately based on the comparison of a state to its image, and here the pure states are the worst case. Hence, the remarkable insensitivity of the quantum capacity to the choice of the error criterion stems from the observation that the comparison between an arbitrary state and a pure state is rather insensitive to the criterion used.

Instead of requiring the error quantity in eqn [1] to approach zero in the large block limit $\nu \rightarrow \infty$, one might feel tempted to impose that the errors vanish completely for some sufficiently large block length, since this is the standard setup in the theory of quantum error correction (see Quantum Error Correction and Fault Tolerance). While it is true that errors can always be assumed to vanish exponentially in eqn [1], requiring perfect correction may completely change the picture: if a channel has some small positive probability for depolarization, the same also holds for its tensor powers, and no such channel allows the perfect transmission of even one qubit. Hence, the capacity for perfect correction will vanish for such channels, while the standard capacity (in accordance with Definition 1) will be close to maximal, $Q(T) \approx 1$. The existence of perfect error-correcting codes thus gives lower bounds on the channel capacity, but is not required for a positive transfer rate.

In the other extreme, one might sometimes feel inclined to tolerate (small) finite errors in the transmission. For some $\varepsilon > 0$, we define $Q_\varepsilon(T)$ exactly like the quantum capacity in Definition 1, but require only that the error quantity in eqn [1] falls below ε for some sufficiently large ν . Obviously, $Q_\varepsilon(T) \geq Q(T)$ for any quantum channel T . We also have $\lim_{\varepsilon \rightarrow 0} Q_\varepsilon(T) = Q(T)$ (Kretschmann and Werner 2004). In the classical setting, even a *strong converse* is known: if $\varepsilon > 0$ is small enough, one cannot achieve bigger rates by allowing small errors, that is, $C_\varepsilon(T) = C(T)$. It is still undecided whether an analogous property holds for the quantum capacity $Q(T)$.

Related Capacities

This article is chiefly concerned with the quantum capacity of a quantum channel. A variety of other

capacities have been derived from [Definition 1](#) by either amending the channel S to be simulated, or allowing Alice and Bob to make use of additional resources. Their interrelations are reviewed in [Bennett et al. \(2004\)](#)

Much interest has been devoted to the hybrid problem of transmitting classical information undistorted over noisy quantum channels. The classical capacity $C(T)$ of a quantum channel T is discussed in the article [Quantum Channels: Classical Capacity](#) of this Encyclopedia. It is obtained by choosing the ideal one-bit channel rather than the one-qubit channel as the standard of reference in [Definition 1](#). Encoding channels E and decoding channels D are then restricted to preparations and measurements, respectively. Since a quantum channel can also be employed to send classical information, we have $C(T) \geq Q(T)$. There are, obviously, examples in which this inequality is strict: the *entanglement-breaking* channel $T(\rho) = \sum_j \langle j|\rho|j\rangle |j\rangle\langle j|$ is composed of a measurement in the orthonormal basis $\{|j\rangle\}_j$, followed by a preparation of the corresponding basis states. It destroys all the entanglement between the sender and a reference system, implying $Q(T) = 0$. Yet all the basis states $|j\rangle$ are transmitted undistorted, which is enough to guarantee that $C(T) = 1$.

[Definition 1](#) also applies to purely classical channels, and thus to the setting of Shannon's information theory. A classical channel T between two d -level systems is completely specified by the $d \times d$ matrix $(T_{xy})_{x,y=1}^d$ of transition probabilities. For these channels the cb-norm difference is just (twice) the maximal error probability:

$$\|\text{id} - T\|_{\text{cb}} = 2 \sup_x \{1 - T_{xx}\}$$

which is the standard error criterium for classical information transfer.

Dense coding and teleportation suggest that entanglement is a powerful resource for information transfer. It doubles the classical channel capacity of a noiseless channel, and it allows to send quantum information over purely classical channels. Surprisingly, the *entanglement-assisted capacities* are often simpler and better behaved than their unassisted counterparts. Unlike the classical and quantum capacities proper, they are relatively easy to calculate using finite optimization procedures, and there has recently been significant progress in understanding the simulation rates for nonideal channels in this scenario (*see Capacities Enhanced by Entanglement*).

The quantum channel capacity is unaffected by entanglement-breaking side channels. In particular, classical forward communication alone cannot

enhance it. However, unlike in the purely classical case, both the quantum and classical channel capacity (but not the entanglement-assisted capacity) may increase under classical feedback.

Elementary Properties

The capacity of a composite channel $T_1 \circ T_2$ cannot be bigger than the capacity of the channel with the smallest bandwidth. This in turn suggests that simulating a concatenated channel is in general easier than simulating any of the individual channels. These relations are known as *bottleneck inequalities*:

$$Q(T_1 \circ T_2, S) \leq \min\{Q(T_1, S), Q(T_2, S)\} \quad [4]$$

$$Q(T, S_1 \circ S_2) \geq \max\{Q(T, S_1), Q(T, S_2)\} \quad [5]$$

Instead of running T_1 and T_2 in succession, we may also run them in parallel. In this case, the capacity can be shown to be *superadditive*,

$$Q(T_1 \otimes T_2, S) \geq Q(T_1, S) + Q(T_2, S) \quad [6]$$

For the standard ideal channels, we even have additivity. The same holds true if both S and one of the channels T_1, T_2 are noiseless, the third channel being arbitrary. However, results on the activation of bound-entangled states seem to suggest that the inequality in [eqn \[6\]](#) may be strict for some channels (*see Entanglement*).

Finally, the *two-step coding inequality* tells us that by using an intermediate channel in the coding process we cannot increase the transmission rate:

$$Q(T_1, T_2) \geq Q(T_1, T_3) Q(T_3, T_2) \quad [7]$$

Applying [eqn \[7\]](#) twice with $T_2 = \text{id}$ and $T_3 = \text{id}$ immediately yields upper and lower bounds on the channel capacity with nonideal reference channel,

$$\frac{Q(T_1)}{Q(T_2)} \geq Q(T_1, T_2) \geq Q(T_1) Q(\text{id}, T_2) \quad [8]$$

The evaluation of the lower bound in [eqn \[8\]](#) then requires efficient protocols for simulating a noisy channel T_2 with a noiseless resource.

There are special cases in which the quantum channel capacity can be evaluated relatively easily, the most relevant one being the noiseless channel id_n , where by the subscript n we denote the dimension of the underlying Hilbert space. In this case, we have

$$Q(\text{id}_n, \text{id}_m) = \frac{\text{ld } n}{\text{ld } m} \quad [9]$$

The lower bound $Q(\text{id}_n, \text{id}_m) \geq \text{ld } n / \text{ld } m$ is immediate from counting dimensions. To establish the upper bound, we use the fact that a noiseless quantum channel cannot simulate itself with a rate

exceeding unity: $Q(\text{id}_m, \text{id}_m) \leq 1$. This is just the upper bound we want to prove for the special case $n = m$, and it can be extended to the general case with the help of the two-step coding inequality [7]: $Q(\text{id}_m, \text{id}_n) Q(\text{id}_n, \text{id}_m) \leq Q(\text{id}_m, \text{id}_m) \leq 1$, implying $Q(\text{id}_n, \text{id}_m) \leq 1/Q(\text{id}_m, \text{id}_n) \leq \text{ld } n / \text{ld } m$, where in the last step we have applied the lower bound with the roles of n and m interchanged.

Combining eqn [9] with the two-step coding inequality [7], we see that for any channel T

$$Q(T, \text{id}_n) = \frac{\text{ld } m}{\text{ld } n} Q(T, \text{id}_m) \quad [10]$$

which shows that quantum channel capacities relative to noiseless channels of different dimensionality only differ by a constant factor. Fixing the dimensionality of the reference channel then only corresponds to a choice of units. Conventionally, the ideal qubit channel id_2 is chosen as a standard of reference, as in Definition 1 above, thereby fixing the unit “bit.”

The upper bound on the capacity of ideal channels can also be obtained from a general upper bound on quantum capacities (Holevo and Werner 2001), which has the virtue of being easily calculated in many situations. It involves the *transposition map*, which we denote by Θ , defined as matrix transposition with respect to some fixed orthonormal basis. The transposition is positive but not completely positive, and thus does not describe a physical channel (see Channels in Quantum Information Theory). We have $\|\Theta\|_{\text{cb}} = d$ for a d -level system. For any channel T and small $\varepsilon > 0$,

$$Q(T) \leq Q_\varepsilon(T) \leq \text{ld } \|T\Theta\|_{\text{cb}} =: Q_\Theta(T) \quad [11]$$

where Q_ε is the finite error capacity introduced in the section “Quantum channel capacity.”

The upper bound $Q_\Theta(T)$ has some remarkable properties, which make it a capacity-like quantity in its own right. For example, it is exactly additive,

$$Q_\Theta(S \otimes T) = Q_\Theta(S) + Q_\Theta(T) \quad [12]$$

for any pair S, T of channels, and it satisfies the bottleneck inequality:

$$Q_\Theta(ST) \leq \min\{Q_\Theta(S), Q_\Theta(T)\}$$

Moreover, it coincides with the quantum capacity on ideal channels, $Q_\Theta(\text{id}_n) = Q(\text{id}_n) = \text{ld } n$, and it vanishes whenever $T\Theta$ is completely positive. In particular, if $\text{id} \otimes T$ maps any entangled state to a state with positive partial transpose, we have $Q_\Theta(T) = 0$.

State–Channel Duality

Quantum capacity is closely related to the *distillable entanglement*, which is the optimal rate m/n at

which n copies of a given bipartite quantum state ρ shared between Alice and Bob can be asymptotically converted into m maximally entangled qubit pairs (see Entanglement). Similar to the quantum capacity, the definition involves the large block limit $n, m \rightarrow \infty$ and an optimization over all conceivable distillation protocols. These may consist of several rounds of local quantum operations and (forward or two-way) classical communication. The one-way and two-way distillable entanglement of ρ will be denoted by $D_1(\rho)$ and $D_2(\rho)$, respectively.

Suppose that Alice and Bob are connected by a quantum channel T and run such a one-way distillation protocol on (many copies of) the state $\rho_T := (T \otimes \text{id})|\Omega\rangle\langle\Omega|$, where $|\Omega\rangle := (1/\sqrt{d_A}) \sum_i |i, i\rangle$ is maximally entangled on $\mathcal{H}_A \otimes \mathcal{H}_A$. If the distillation yields maximally entangled qubits at positive rate R , Alice may apply the standard teleportation scheme to send arbitrary quantum states to Bob undistorted at that same rate R . Like the distillation protocol itself, teleportation requires classical forward communication, which however does not affect the channel capacity (cf. the section “Related capacities”). Thus, $Q(T) \geq D_1(\rho_T)$. If two-way distillation is allowed, we have $Q_2(T) \geq D_2(\rho_T)$ for the capacity $Q_2(T)$ assisted by two-way classical side communication.

Conversely, if Alice and Bob use a bipartite quantum state ρ shared between them as a substitute for the maximally entangled state $|\Omega\rangle$ in the standard teleportation protocol, they will implement some noisy quantum channel T_ρ . If this channel allows to transfer quantum information at nonvanishing rate R , Alice may share maximally entangled states with Bob at that same rate R . Consequently, $D_1(\rho) \geq Q(T_\rho)$ and $D_2(\rho) \geq Q_2(T_\rho)$.

These relations (Bennett *et al.* 1996) allow to bound channel capacities in terms of distillable entanglement and vice versa. If the two maps $T \mapsto \rho_T$ and $\rho \mapsto T_\rho$ are mutually inverse, we even have $D_1(\rho) = Q(T_\rho)$ and $D_2(\rho) = Q_2(T_\rho)$. In this case, the duality $\rho \rightleftharpoons T_\rho$ is the physical implementation of Jamiolkowski’s isomorphism between bipartite states and channels (see Channels in Quantum Information Theory). This has been shown (Horodecki *et al.* 1999) to hold for *isotropic states*, which are invariant under the group of all $\bar{U} \otimes U$ transformations, where \bar{U} is the complex conjugate of the unitary U . The corresponding channels are partly depolarizing.

In general, $T_{\rho_T} \neq T$. However, the so-called *conclusive teleportation* allows us to implement T at least probabilistically, resulting in the relation

$$\frac{1}{d_A^2} Q(T) \leq D_1(\rho_T) \leq Q(T) \quad [13]$$

The duality [13] can be applied to show that both the unassisted and the two-way quantum capacities are continuous in any open set of channels having nonvanishing capacities (Horodecki and Nowakowski 2005).

Coding Theorems

Computing channel capacities straight from Definition 1 is a tricky business. It involves optimization in systems of asymptotically many tensor factors, and can only be performed in special cases, like the noiseless channels in the section “Elementary properties.” Coding theorems aspire to reduce this problem to an optimization over a low-dimensional space. They usually come in two parts: the *converse* provides an upper bound on the channel capacity (typically in terms of some entropic expression), while the *direct* part consists of a coding scheme that attains this bound. By Shannon’s celebrated coding theorem, the classical capacity of a classical noisy channel can be obtained from a maximization of the *mutual information* over all joint input-output distributions.

For the quantum channel capacity, the relevant entropic quantity is the *coherent information*,

$$I_c(T, \varrho) := H(T(\varrho)) - H(T \otimes \text{id}(|\psi_\varrho\rangle\langle\psi_\varrho|)) \quad [14]$$

where H denotes the von Neumann entropy: $H(\varrho) = -\text{tr} \varrho \text{ld} \varrho$, and $|\psi_\varrho\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A'}$ is a purification of the density operator $\varrho \in \mathcal{A}$. The coherent information does not increase under quantum operations, $I_c(S \circ T, \varrho) \leq I_c(T, \varrho)$ for any quantum channel S and state $\varrho \in \mathcal{A}$. This is the *data processing inequality* (Barnum *et al.* 1998), which shows that the regularized coherent information provides an upper bound on the quantum channel capacity: if Alice and Bob have a coding scheme for the channel T with capacity $Q(T)$, n channel uses allow them to share a maximally entangled state of size $\sim \exp_2 n Q(T)$. The coherent information of this state equals $\sim n Q(T)$, and was no larger prior to Bob’s decoding.

Recently, Devetak (2005) developed a coding scheme to show that this bound is in fact attainable. Different proofs were outlined by Lloyd and Shor.

Theorem 1 For every quantum channel T ,

$$Q(T) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{\varrho} I_c(T^{\otimes n}, \varrho) \quad [15]$$

Unlike the classical or quantum mutual information, coherent information is strictly superadditive for some channels (DiVincenzo *et al.* 1998). Hence,

taking the limit $n \rightarrow \infty$ in eqn [15] is indeed required, and in general the evaluation of the capacity formula [15] still demands the solution of asymptotically large variational problems. This should be contrasted with the entanglement-assisted capacities $C_E(T) = 2Q_E(T)$ (where a simple nonregularized coding theorem is known to hold, *see* Capacities Enhanced by Entanglement) and the capacity for classical information $C(T)$ (where additivity is conjectured but not proved, *see* Quantum Channels: Classical Capacity). Even a maximization of the single-shot coherent information $I_c(T, \varrho)$ appears to be a difficult optimization problem, since this quantity is neither convex nor concave and may have multiple local maxima (Shor 2003). Thus, even for simple-looking systems like the qubit depolarizing channel, so far we only have upper and lower bounds on the quantum channel capacity, but do not yet know how to compute its exact value.

We now sketch Devetak’s proof of Theorem 1, assuming only some familiarity with Holevo–Schumacher–Westmoreland (HSW) random codes for the classical channel capacity (*see* Quantum Channels: Classical Capacity). It is easily seen from Stinespring’s dilation theorem (*see* Channels in Quantum Information Theory) that a noiseless quantum channel provides perfect security against eavesdropping. This is one of the characteristic traits of quantum mechanics and lies at the heart of quantum cryptography. In his proof, Devetak showed a way to turn this around and upgrade coding schemes for private classical information to quantum channel codes.

The relation between quantum information transfer over a channel $T: \mathcal{A} \rightarrow \mathcal{B}$ and privacy against eavesdropping is best understood in terms of the *companion channel* $T_\mathcal{E}: \mathcal{A} \rightarrow \mathcal{E}$. $T_\mathcal{E}$ arises from a given Stinespring isometry $V: \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_\mathcal{E}$ of $T \equiv T_B$ by interchanging the roles of the output system \mathcal{B} and the environment \mathcal{E} :

$$T_B(\varrho) = \text{tr}_\mathcal{E} V \varrho V^* \iff T_\mathcal{E}(\varrho) = \text{tr}_B V \varrho V^* \quad [16]$$

The channel $T_\mathcal{E}$ describes the information flow into the environment \mathcal{E} , a system we assume to be under complete control of a potential eavesdropper, Eve say. The setup for private classical information transfer (including the definition of rates and capacity) is then exactly the same as for the classical channel capacity (*see* Quantum Channels: Classical Capacity), but the protocols now have to satisfy the additional requirement that $T_\mathcal{E}$ releases (almost) no information to the environment. This can be achieved by randomizing over $\nu_\mathcal{E} \sim \exp_2 n \chi(T_\mathcal{E}, \{p_i, \varrho_i\})$ code words of a standard HSW code of total size $\sim \exp_2 n \chi(T_B, \{p_i, \varrho_i\})$, where $\{p_i, \varrho_i\}$ is the quantum ensemble from which a set of random code words

$\{\sigma_{k,l}\}_{k=1,l=1}^{\nu_B, \nu_E}$ is generated. The appearance of the *Holevo bound*

$$\chi(T, \{p_i, \varrho_i\}) := H\left(\sum_i p_i T(\varrho_i)\right) - \sum_i p_i H(T(\varrho_i)) \quad [17]$$

in the dimension of both these code spaces can be understood from the size of the relevant typical subspaces (Devetak and Winter 2004).

The randomization guarantees that the remaining $\nu_B \sim \exp_2 n(\chi(T_B) - \chi(T_E))$ code words are almost indistinguishable to Eve:

$$\left\| \frac{1}{\nu_E} \sum_{l=1}^{\nu_E} T_E^{\otimes n}(\sigma_{kl} - \sigma_{jl}) \right\|_1 \leq \varepsilon, \quad \forall j, k = 1, \dots, \nu_B \quad [18]$$

The net transfer rate for private classical information is then $R \sim \chi(T_B) - \chi(T_E)$, which is just the total transfer rate for the channel Alice \rightarrow Bob reduced by the transfer rate Alice \rightarrow Eve.

Remarkably, if $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ is a decomposition of $\varrho \in \mathcal{A}$ into pure states, the private transfer rate exactly equals the coherent information,

$$\begin{aligned} I_c(T_B, \varrho) &= H(T_B(\varrho)) - H(T_E(\varrho)) \\ &= \chi(T_B) - \chi(T_E) \end{aligned} \quad [19]$$

The so-called *entropy exchange*

$$H(T_E(\varrho)) = H(T_B \otimes \text{id}(|\psi_\varrho\rangle\langle\psi_\varrho|))$$

quantifies the extent to which a formerly pure ancilla state becomes mixed via interaction with the signal states. Equation [19] then nicely reflects the intuition that for high-rate quantum information transfer the signal states should not entangle too much with the environment. In fact, for an almost noiseless channel the entropy exchange nearly vanishes, and the optimized coherent information almost attains the maximal value 1, while for nearly depolarizing channels we have $I_c(T_B, \varrho) \approx -H(\varrho) \leq 0$.

So far, we have sketched a protocol for private classical information transfer. Devetak's *coherentification* allows to pass from the transmission of classical messages to the transmission of coherent superpositions. This technique has also been applied to obtain entanglement distillation protocols from secret key distillation, and offers a unified view on the secret classical resources and their quantum counterparts (Devetak and Winter 2004, Devetak *et al.* 2004).

In order to transfer quantum information, Alice will only need to send one half of a maximally entangled state of dimensionality $\sim \exp_2 n I_c(T_B, \varrho)$. As described in the previous section, teleportation then allows her to transfer arbitrary quantum states from a subspace of that size.

Given a set of pure state code words $\{|\varphi_{kl}\rangle\}_{k=1,l=1}^{\nu_B, \nu_E}$ of a private classical information protocol, for entanglement transfer Alice prepares the input state

$$|\Phi\rangle_{\mathcal{A}'\mathcal{A}} = \frac{1}{\sqrt{\nu_B}} \sum_{k=1}^{\nu_B} |k\rangle_{\mathcal{A}'} \otimes \frac{1}{\sqrt{\nu_E}} \sum_{l=1}^{\nu_E} |\varphi_{kl}\rangle_{\mathcal{A}} \quad [20]$$

where \mathcal{A}' denotes a reference system that Alice keeps in her lab. On his share of the resulting output state $|\Phi'\rangle_{\mathcal{A}'\mathcal{B}\mathcal{E}}$ Bob will then employ the corresponding measurement operators $\{M_{kl}\}_{k,l=1}^{\nu_B, \nu_E}$ to implement the coherent measurement

$$V_M |\varphi\rangle_{\mathcal{B}} := \sum_{kl} \sqrt{M_{kl}} |\varphi\rangle_{\mathcal{B}} \otimes |kl\rangle_{\mathcal{B}_1\mathcal{B}_2}$$

which places the measurement outcomes into some reference system $\mathcal{B}_1 \otimes \mathcal{B}_2$. Any measurement which identifies the output with high probability only slightly disturbs the output state, and thus Bob's coherent measurement leaves the total system in an approximation of the state

$$|\Phi''\rangle = \frac{1}{\sqrt{\nu_B \nu_E}} \sum_{k=1, l=1}^{\nu_B, \nu_E} |k\rangle_{\mathcal{A}'} |k\rangle_{\mathcal{B}_1} |l\rangle_{\mathcal{B}_2} |\varphi'_{kl}\rangle_{\mathcal{B}\mathcal{E}} \quad [21]$$

in which Eve and Bob are still entangled. A completely depolarizing channel T_E would directly yield a factorized output state $\mathcal{B} \otimes \mathcal{E}$ here. Although the randomization in eqn [18] does not necessarily result in complete depolarization, there is a controlled unitary operation which Bob may apply to effectively decouple Eve's system, resulting in the output state $\sim (1/\sqrt{\nu_B}) \sum_k |kk\rangle_{\mathcal{A}'\mathcal{B}_1} \otimes \mathcal{E}$, which is the maximally entangled state of size $\nu_B \sim \exp_2 n I_c(T_B, \varrho)$ required for teleportation. The direct part of the capacity theorem then follows by applying the above coding scheme to large blocks and maximizing over (pure) input ensembles, concluding the proof.

Devetak's proof of the coding theorem seems to indicate that the private classical capacity $C_p(T)$ equals the quantum capacity $Q(T)$ for every quantum channel T . However, for the coherentification protocol, we have restricted the private coding schemes to pure state input ensembles, and thus we can only conclude that $Q(T) \leq C_p(T)$. The existence of bound-entangled states with positive one-way distillable secret key rate (Horodecki *et al.* 2005) implies that this inequality can be strict. A general procedure does exist to retrieve (almost) all the information from the output of a noisy quantum channel that releases (almost) no information to the environment. But this requires a stronger form of privacy than eqn [18].

Quantum Channels with Memory

This article has so far been restricted to *memoryless* quantum channels, in which successive channel inputs are acted on independently. Messages of n symbols are then processed by the tensor product channel $T^{\otimes n}$, as in Definition 1 and illustrated in Figure 1. In many real-world applications, the assumption of having uncorrelated noise cannot be justified, and memory effects need to be taken into account. For a quantum channel T with register input \mathcal{A} and register output \mathcal{B} , such effects are conveniently modeled (Bowen and Mancini 2004) by introducing an additional memory system \mathcal{M} , so that now $T: \mathcal{M} \otimes \mathcal{A} \rightarrow \mathcal{B} \otimes \mathcal{M}$ is a completely positive and trace-preserving map with two input systems and two output systems. Long messages with n signal states will then be processed by the concatenated channel $T_n: \mathcal{M} \otimes \mathcal{A}^n \rightarrow \mathcal{B}^n \otimes \mathcal{M}$. In such a concatenation, the memory system is passed on from one channel application to the next, and thus introduces (classical or quantum) correlations between consecutive register inputs.

Remarkably, this relatively simple model can be shown (Kretschmann and Werner 2005) to encompass every reasonable physical process: every stationary channel $S: \mathcal{A}^\infty \rightarrow \mathcal{B}^\infty$ which turns an infinite string of input states (on the quasilocal algebra \mathcal{A}^∞) into an infinite string of output states on \mathcal{B}^∞ and satisfies the *causality* constraint is in fact a concatenated memory channel. Causality here means that the outputs of the stationary channel S at given time t_0 do not depend on inputs at times $t > t_0$. Figure 2 illustrates the *structure theorem* for causal stationary quantum channels. In general, it produces not only the memory channel T with memory algebra \mathcal{M} , but also a map R describing the influence of input states in the remote past. Intuitively, such a map is often not needed, because memory effects decrease in time: the memory channel T is called *forgetful* if outputs at a large time t depend only weakly on the memory initialization at time zero. In fact, memory effects can be

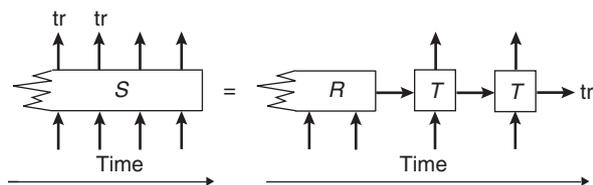


Figure 2 By the structure theorem, a causal automaton S can be decomposed into a chain of concatenated memory channels T plus some input initializer R . Evaluation with the partial trace tr means that the corresponding output is ignored.

shown to die out even exponentially. The set of these channels is open and dense in the set of quantum memory channels. Hence, generic memory channels are forgetful.

The capacity of memory channels is defined in complete analogy to the memoryless case, replacing the n -fold tensor product $T^{\otimes n}$ in Definition 1 by the n -fold concatenation T_n . The coding theorems for (private) classical and quantum information can then be extended from the memoryless case to the very important class of forgetful channels (Kretschmann and Werner 2005).

Nonforgetful channels call for universal coding schemes, which apply irrespective of the initialization of the input memory. Such schemes are presently known only for very special cases.

Acknowledgments

The author thanks the members of the quantum information group at TU Braunschweig for their careful reading of the manuscript and many helpful suggestions. He also gratefully acknowledges the funding from Deutsche Forschungsgemeinschaft (DFG).

See also: Capacities Enhanced by Entanglement; Channels in Quantum Information Theory; Entanglement; Positive Maps on C^* -Algebras; Quantum Channels: Classical Capacity; Quantum Error Correction and Fault Tolerance; Source Coding in Quantum Information Theory.

Further Reading

- Barnum H, Nielsen MA, and Schumacher B (1998) Information transmission through a noisy quantum channel. *Physical Review A* 57: 4153 (quant-ph/9702049).
- Bennett CH, Devetak I, Shor PW, and Smolin JA (2004) Inequalities and separations among assisted capacities of quantum channels, quant-ph/0406086.
- Bennett CH, DiVincenzo DP, Smolin JA, and Wootters WK (1996) Mixed-state entanglement and quantum error correction. *Physical Review A* 54: 3824 (quant-ph/9604024).
- Bowen G and Mancini S (2004) Quantum channels with a finite memory. *Physical Review A* 69: 012306 (quant-ph/0305010).
- Devetak I (2005) The private classical information capacity and quantum information capacity of a quantum channel. *IEEE Transactions on Information Theory* 51: 44 (quant-ph/0304127).
- Devetak I, Harrow AW, and Winter A (2004) A family of quantum protocols. *Physical Review Letters* 93: 230504 (quant-ph/0308044).
- Devetak I and Winter A (2004) Relating quantum privacy and quantum coherence: an operational approach. *Physical Review Letters* 93: 080501 (quant-ph/0307053).
- DiVincenzo DP, Shor PW, and Smolin JA (1998) Quantum channel capacities of very noisy channels. *Physical Review A* 57: 830 (quant-ph/9706061).
- Eisert J and Wolf MM Gaussian quantum channels. In Cerf N, Leuchs G, and Polzik E (eds.) *Quantum Information with*

Continuous Variables of Atoms and Light. London: Imperial College Press (in preparation)(quant-ph/0505151).

Holevo AS and Werner RF (2001) Evaluating capacities of bosonic Gaussian channels. *Physical Review A* 63: 032312 (quant-ph/9912067).

Horodecki M, Horodecki P, and Horodecki R (1999) General teleportation channel, singlet fraction, and quasidistillation. *Physical Review A* 60: 1888 (quant-ph/9807091).

Horodecki P and Nowakowski ML (2005) Simple test for quantum channel capacity, quant-ph/0503070.

Horodecki K, Pankowski L, Horodecki M, and Horodecki P (2005) Low dimensional bound entanglement with one-way distillable cryptographic key, quant-ph/0506203.

Kretschmann D and Werner RF (2004) Tema con variazioni: quantum channel capacity. *New Journal of Physics* 6: 26 (quant-ph/0311037).

Kretschmann D and Werner RF (2005) Quantum channels with memory. *Physical Review A* 72: 062323 (quant-ph/0502106).

Shor PW (2003) Capacities of quantum channels and how to find them. *Mathematical Programming* 97: 311 (quant-ph/0304102).

Capillary Surfaces

R Finn, Stanford University, Stanford, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Historical and Conceptual Background

A capillary surface is the interface separating two fluids that lie adjacent to each other and do not mix. Examples of such surfaces are the upper surface of liquid partially filling a vertical cylinder (capillary tube), the surface of a liquid drop resting in equilibrium on a tabletop (sessile drop) and the surface of a liquid drop hanging from a ceiling (pendent drop); further instances are the surface of a falling raindrop, the bounding surface of the liquid in the fuel tank of a spaceship, and the interface formed by a fluid mass rotating within another fluid. This last example extends to the problem of rotating stars.

Interfaces separating fluids and solids share some of the physical attributes of capillary surfaces, and the study of wetted portions of rigid “support surfaces” becomes essential for describing global behavior of capillary configurations. However, some significant distinctions appear that change the formal structure of the problems, and must be accounted for in the theory.

Phenomena governed by capillarity pervade all of daily life, and most are so familiar as to escape special notice. By contrast, throughout the eighteenth century and presumably earlier, great attention centered on the rise of liquid in a narrow glass circular-cylindrical tube dipped vertically into a liquid reservoir (Figure 1); this striking event had a dramatic impact that confounded intuition. Clarification of the behavior became one of the major problems challenging the scientific world of the time, and was not achieved during that period. The term “capillary,” adapted from the Latin “capillus” for hair, was applied to the phenomenon since it was observed only for tubes with very fine openings; the

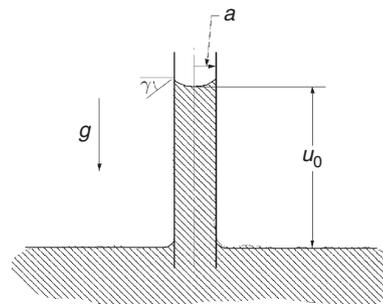


Figure 1 Capillary tube in infinite reservoir, in downward gravity field.

more general usage adopted in the definition above derives from the recognition of a class of phenomena with a common physical basis.

The first recorded observations concerning capillarity seem due to Aristoteles *c.* 350 BC. He wrote that “a broad flat body, even of heavy material, will float on water, however a narrow thin one such as a needle will always sink.” Any reader with access to a needle and a glass of water will have little difficulty refuting the assertion. Remarkably, the error in reasoning seems not to have been pointed out for almost 2000 years, when Galileo addressed the problem in his *Discorsi*, about 1600. The only substantive studies till that time are apparently those of Leonardo da Vinci a hundred years earlier. Leonardo introduced reasoning close in spirit to that of current literature; however, the Calculus was not available to him, and he was not in a position to develop his ideas in quantitative ways.

Young's Contribution

The later discovery of the Calculus provided a driving impetus guiding many new studies during the eighteenth century. But despite the enormity of that weapon, it did not on its own suffice, and initial quantitative success had to await two initiatives

taken by Thomas Young in 1805. Young based his studies on the concept of surface tension that had been introduced by von Segner half a century earlier. Segner hypothesized that every curve on a fluid/fluid interface S experiences on both its sides an orthogonal force σ per unit length, which (for given temperature) depends only on the materials and is directed into the tangent planes on the respective sides. The presence of such forces can be indicated by simple experiments. They become clearly evident in the case of thin (soap) films spanning a frame, in which case there is an easily observed orthogonal pull on the frame, see the section “Dual interpretation of σ : distinction between fluids and solids.” Young made two basic conceptual contributions (Y1, Y2):

Y1. *Relation of pressure jump across a free interface to mean curvature and surface tension.*

Consider a piece of surface S in the shape of a spherical bowl of radius R , separating two immiscible fluid media, as in Figure 2. In equilibrium, any pressure difference δp across S must be balanced by a tension σ on its rim Γ . If S projects to a disk of (small) radius r on the plane tangent to S at the symmetry point, we are led to

$$\pi r^2 \delta p \simeq 2\pi r \sigma \sin \vartheta \tag{1}$$

where ϑ is inclination of S at the rim, relative to the plane. We thus find at the base point

$$\delta p = 2\sigma \frac{d \sin \vartheta}{dr} = 2\sigma \frac{1}{R} \tag{2}$$

Young then went on to consider a general S , without symmetry hypothesis. Letting $1/R_1, 1/R_2$ denote the planar curvatures at a point in S of two normal sections in orthogonal directions, he asserted that

$$\delta p = 2\sigma \frac{1}{2} \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \equiv 2\sigma H \tag{3}$$

where H is the mean curvature of S at the point. Although Young provided no formal justification for this step, we can establish it with the aid of a general formula from differential geometry that was not known in his lifetime:

$$\int_S 2HN \, dS = \oint_r \mathbf{n} \, ds \tag{4}$$

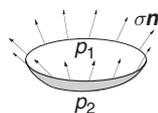


Figure 2 Pressure change across fluid element, balanced by surface tension.

where N is a unit normal on S , and \mathbf{n} is unit conormal (as indicated in Figure 2) on Γ . Multiplying both sides of [4] by σ , the right-hand side becomes the net surface tension force on S . Since that must equal the net balancing pressure force, we obtain

$$\int_S (\delta p - 2\sigma H) N \, dS = 0 \tag{5}$$

Letting the diameter of S tend to zero, the assertion follows.

We emphasize here the implicit assumption above, that σ is a constant depending only on the particular materials, and not on the shape of S . This author knows of no source in which that is clearly established, although experiments and experience provide some *a posteriori* justification. See the further comments under Y2, and later in sections “Gauss’ contribution: the energy method” and “Dual interpretation of σ : distinction between fluids and solids.”

Y2. *The capillary contact angle.*

Young asserted that there are surface tensions for solid/fluid interfaces analogous to those just introduced, and again depending only on the materials. This assertion is erroneous, as was suggested in writings of Bikerman and of others, and more recently established in a definitive example by Finn. Using his premise, Young attempted to characterize the *contact angle* γ made by the fluid surface with a rigid boundary, by requiring that the net tangential component of the three surface tension vectors vanish at the triple interface; this leads to the often employed but incorrect “Young diagram,” see Figure 3, and the relation

$$\cos \gamma = \frac{\sigma_1 - \sigma_2}{\sigma_0} \tag{6}$$

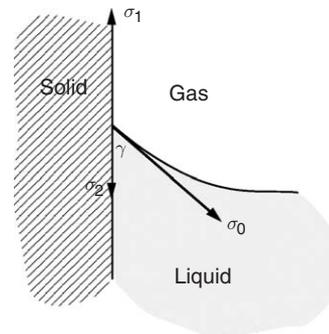


Figure 3 Young diagram; balance of tangential forces. Residual normal force remains.

for $\cos \gamma$ in terms of the magnitudes of the three “surface tensions.” Young concluded that the contact angle depends only on the materials, and in no other way on the conditions of the problem. This basic assertion is by a fortuitous accident correct, as follows from the contribution by Gauss described below; it underlies all modern theory.

Using Y1 and Y2, Young produced the first verifiable prediction for the rise height u_0 in the circular capillary tube of Figure 1. He assumed the interface to be spherical, so that H is constant and $a = \cos \gamma / H$. He assumed vanishing outside pressure. According to classic laws of hydrostatics, $\delta p = \rho g u_0 = 2\sigma H$ by Y1, where ρ is fluid density; there follows the celebrated relation, presented entirely in words in his 1805 article:

$$u_0 \approx \frac{2 \cos \gamma}{\kappa a}, \quad \kappa = \frac{\rho g}{\sigma} \quad [7]$$

Young scorned the mathematical method, and made a point of deriving and publishing his results on capillarity without use of any mathematical symbols. This personal idiosyncrasy causes his publications to be something of a challenge to read.

The Laplace Contribution

In 1806, Laplace published the first analytical expression for the mean curvature of a surface $u(x, y)$, and showed that the expression can be written as a divergence. He obtained the equation

$$\operatorname{div} Tu \equiv 2H, \quad Tu \equiv \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \quad [8]$$

Thus, if H is known from geometrical or physical considerations, as it is for the capillary tube in the example just considered, one finds a second-order (nonlinear) equation for the surface height of any solution as a graph. The equation is elliptic for any function $u(x, y)$ inserted into the coefficients, however not uniformly so; the particular nonuniformity leads to some striking and unusual behavior of its solutions, as we shall see. With the aid of [8], Laplace improved the Young estimate [7] to

$$u_0 \approx \frac{2 \cos \gamma}{\kappa a} - \left[\frac{1}{\cos \gamma} - \frac{2}{3} \left(\frac{1 - \sin^3 \gamma}{\cos^3 \gamma} \right) \right] a \quad [9]$$

Both Young and Laplace proposed their formulas for “narrow tubes”, but neither gave any

quantitative indication of what “narrow” should signify. Note that whenever $0 \leq \gamma < \pi/2$, [9] becomes negative when the nondimensional *Bond Number* $B = \kappa a^2$ exceeds 8; since u is known to be positive in the indicated range for γ , [9] provides no information in that case, whereas [7] is still of some value. Nevertheless, [9] is asymptotically exact and consists of the first two terms of the formal expansion in powers of a ; that was first proved by D Siegel in 1980, almost 200 years following the discovery of the formulas. In 1968, P Concus extended the formal expansion for the height to the entire traverse $0 < r < a$. F Brulois (1981) and independently E Miersemann (1994) proved the expansion to be asymptotic to every order. Explicit bounds for the rise height above and below, making quantitative the notion of “narrow,” were obtained by Finn.

Laplace supplied the first detailed mathematical investigations into the behavior of capillary surfaces, applying his ideas to many specific examples. His underlying motivation apparently derived at least partly from astronomical problems, and he published his contributions in two “Suppléments” to the tenth volume of his *Mécanique Céleste*.

Gauss' Contribution: The Energy Method

Young and Laplace both based their reasonings on force-balance arguments, which at best were unclear and at worst conceptually wrong. In 1830, Gauss took up the problem anew from a variational point of view, using the Johann Bernoulli principle of virtual work. To do so, he attempted to characterize both surface energies and bulk fluid energies in terms of postulated particle attractions and repulsions. In an astonishing 30 pages, he essentially introduced foundations of modern potential theory, of measure theory, and of thermodynamics. He ended up with elaborate expressions that could not readily be applied, and which at least to some extent he did not use. He asserted, for example, that the bulk internal energy would be proportional to volume, which for an incompressible fluid is constant under admissible deformations, and on that basis he ignored the bulk energy term completely. His procedures then led him, in an independent and more convincing way, to the identical equation and boundary condition that had been produced by his predecessors. It must, of course, be remarked that his justification for ignoring the bulk energy term would not be correct for a compressible liquid (see the section “Compressibility”), and it is open to some

question for the central motivating problem of a capillary tube dipped into an infinite liquid bath, in which event there is no volume constraint.

The material that follows is guided by the ideas of Gauss; however, I have found it advantageous to replace his elaborate hypotheses on particle attractions and repulsions by a simpler phenomenological reasoning as to the nature of the energy terms to be expected.

To fix ideas, we consider a semi-infinite cylinder of general section Ω and of homogeneous material, closed at the bottom, situated vertically in a downward gravity field g per unit mass, and partly filled with an incompressible liquid of density ρ covering the bottom (a more exact discussion, taking account of compressibility, is indicated below in the section “Compressibility”). We assume an equilibrium fluid configuration with the liquid bounded above by an ideally thin interface $S:u(x,y)$ (see Figure 4). We distinguish the energy terms that occur:

1. *Surface energy.* This is the energy required to create the surface interface S . We can characterize it by noting that fluid particles within or exterior to the liquid are attracted equally to neighboring particles in all directions; however, at the surface S there is a differential attraction, to particles of the exterior medium (such as air) above, or to the liquid below (see Figure 5). Thus, particles in the interface are pulled orthogonally to S . In general, for a liquid–gas interface, significant work will be done only on the liquid and those particles will be pulled toward the liquid; otherwise, the liquid would evaporate across the interface and disappear. The work done in that (infinitesimal) motion is proportional to the area of S , so that for the surface energy E_S we obtain

$$E_S = \sigma \int_{\Omega} \sqrt{1 + |\nabla u|^2} dx \quad [10]$$

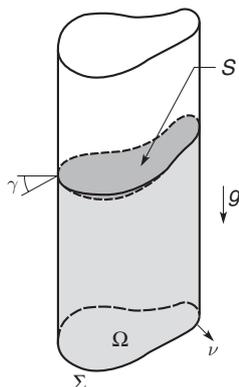


Figure 4 Liquid in cylindrical capillary tube, of general section Ω . Reproduced with permission from the American Institute of Aeronautics and Astronautics.

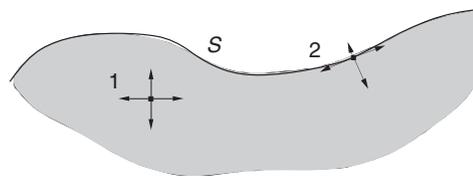


Figure 5 Attractions on a fluid element: (1) interior to the fluid; (2) on the surface interface.

The constant σ has the dimensions of force per unit length, and turns out to be the surface tension of the interface. We note from [10] its dual interpretation as areal energy density on S , arising from formation of that surface. This alternative interpretation lends conceptual support to the supposition that σ is constant on S . See the section “Dual interpretation of σ : distinction between fluids and solids.”

Implicit in the above discussion are deep premises about the nature of the forces acting within the fluid. Essentially these forces must be perceptible only at infinitesimal distances, and grow rapidly with decreasing distance. Forces both of attraction and of repulsion must be present. The recognition of the need for such forces can be traced back to Newton. Quantitative postulates as to their precise nature were introduced by van der Waals in the late nineteenth century, and the topic remains still in active study. Since these forces appear at molecular distance levels, their introduction leads inevitably to questions of statistical mechanics. Additionally, our discussion of work done in forming the surface implicitly assumes a compressible transition layer there, in conflict with our treatment of S as an ideally thin interface bounding an incompressible fluid. In these senses, it is striking that [10] – which is in accord with classical constructions – could be obtained via global qualitative postulates concerning a continuum in static equilibrium, in which the specific nature of the forces is not introduced.

Rayleigh measured the thickness of the surface interface between water and air to be of molecular size, thus providing experimental justification for the procedure adopted.

2. *Wetting energy.* A similar discussion applies at the interface separating the liquid and solid at the cylinder walls; however, this time the net attraction can be in either direction, as particles from neither medium can migrate significantly into the other. For the wetting energy E_W , we write, with Σ the boundary of Ω ,

$$E_W = -\beta\sigma \oint_{\Sigma} u ds \quad [11]$$

We designate β as the *relative adhesion coefficient* of the liquid–gas–solid configuration. We assume that the cylinder walls are of homogeneous material, so that β will be constant. In general, β is a difference of factors that apply on the walls at the two interfaces, with the liquid and with the external medium.

3. *Gravitational energy.* The work done in lifting an amount of liquid $\rho\delta h\delta\Omega$ against the gravity field from the base level to a height h in a vertical tube of small section $\delta\Omega$ is $\rho gh\delta h\delta\Omega$. Thus, the work done in filling that tube up to the surface height u is $(\rho gu^2/2)\delta\Omega$, and the total gravitational energy is

$$E_G = \frac{\rho g}{2} \int_{\Omega} u^2 dx \tag{12}$$

4. *Volume constraint.* In the configuration considered the volume is to be unvaried during admissible deformations; we take account of the constraint by introducing a Lagrange parameter λ , and an additional “energy” term

$$E_V = \lambda\sigma \int_{\Omega} u dx \tag{13}$$

According to the principle of virtual work, the sum E of the above energies must remain unvaried in any deformation that respects all mechanical constraints other than the volume constraint. We choose a deformation $u \rightarrow u + \varepsilon\eta$, with η smooth in the closure of Ω , which determines a functional $E(\varepsilon)$. From $E'(0) = 0$ follows

$$\int_{\Omega} \left\{ \nabla\eta \cdot \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} + \eta(\kappa u + \lambda) \right\} dx - \beta \oint_{\Sigma} \eta ds = 0 \tag{14}$$

from which

$$\int_{\Omega} \eta \{ -\operatorname{div} Tu + (\kappa u + \lambda) \} dx + \oint_{\Sigma} \eta (\nu \cdot Tu - \beta) ds = 0 \tag{15}$$

with $Tu \equiv \nabla u / \sqrt{1 + |\nabla u|^2}$, and with ν the unit exterior normal on Σ . Choosing first η to have compact support in Ω , the boundary term vanishes, and the “fundamental lemma” of the calculus of variations yields

$$\operatorname{div} Tu = \kappa u + \lambda, \quad \kappa = \rho g / \sigma \tag{16}$$

throughout Ω . Thus, the area integral in [15] vanishes for any η . We are therefore free to choose

η as we wish on the boundary, and the fundamental lemma now yields $\nu \cdot Tu = \beta$ on Σ . We now note that for any liquid surface $u(x, y)$ there holds

$$\nu \cdot Tu = \cos \gamma \tag{17}$$

on Σ , where γ is the angle between the cylinder wall and the surface S , measured within the liquid. Since β is assumed to be constant, that is so also for γ . It is a physical constant: the contact angle, that must be measured in an independent experiment, and cannot be prescribed in advance or calculated within the scope of the theory.

The constant β , originally introduced as a general proportionality constant, is now characterized as $\beta = \cos \gamma$. We thus see that a physical surface of the form envisaged is possible only if $-1 \leq \beta \leq 1$. Physically, one expects that if $\beta < -1$ the liquid will separate from the walls, while, if $\beta > 1$, the liquid will spread over the walls as a thin film.

Equation [16] and boundary condition [17] provide a nonlinear second-order equation that is elliptic for any function $u(x, y)$, and also a nonlinear transversality condition on the boundary, for determining the surface interface S . The expression $\operatorname{div} Tu$ is exactly twice the mean curvature of the surface S . If $\kappa \neq 0$ then λ can be eliminated by addition of a constant to u . The problem [16]–[17] for the fluid in a vertical cylindrical capillary tube of general section becomes thus a geometrical one: to find a surface whose mean curvature is a prescribed function of position in space, and which meets the cylindrical boundary walls in a prescribed angle γ .

In the absence of gravity, [16] takes the form

$$\operatorname{div} Tu = 2H \tag{18}$$

for a surface of constant mean curvature H . The constant H is determined by integrating [18] over Ω , and using [17]:

$$2H = \frac{|\Sigma| \cos \gamma}{|\Omega|} \tag{19}$$

where $|\Sigma|$ and $|\Omega|$ denote the respective perimeter and area, and thus H is independent of volume. From the known uniqueness up to an additive constant of the solutions of [18], [17] it follows that the shape of the solution surface is independent of volume. That result holds also for [16], [17] in view of the possibility to eliminate λ from the equation by addition of a constant, and the uniqueness of the solutions of the resulting equation.

Equations [16]–[17] or [18]–[17] are appropriate for determining capillary surfaces that are graphs

$u(x, y)$ over a base domain Ω . More generally, any surface S in 3-space satisfies the equation

$$\Delta x = 2HN \quad [20]$$

where H is its scalar mean curvature and N is a unit normal vector on S . Here Δ is the “intrinsic Laplacian” in the metric of S . This is the appropriate relation to be applied in situations for which the physical surface folds over itself and cannot be expressed globally as a graph. The formal simplicity of [20] is deceptive; the challenges arising from the nonlinearity in the equation can be formidable, and very little general theory is as yet available.

Dual Interpretation of σ : Distinction between Fluids and Solids

We have already remarked the duality in connection with eqn [10] above. It can be made explicit with a simple experiment proposed by Dupré. One makes a rigid frame with a sliding bar of length l , as in Figure 6, and dips the frame into soap solution. On lifting the frame from the solution the opening will be filled with a soap film, and one finds a force $F = 2\sigma l$ on the bar, directed orthogonal to the bar (the factor 2 appears since the film has two sides). The work done in sliding the bar a distance δx is $\delta F = 2\sigma \delta x$, which can also be written $\delta F = 2\sigma \delta A$ with δA an element of area. In this sense, the two interpretations of σ are formally equivalent, for fluid/fluid interfaces.

The equivalence cannot be extended to solid/fluid interfaces. Consider a rigid spherical ball of generic material and radius R , freely floating in an infinite liquid bath in a gravity-free environment, see Figure 7a. It can be shown that the unique symmetric solution to the problem is a horizontal surface, as in the figure. A variational procedure as above shows that if e_0, e_1, e_2 are the interfacial energy densities associated with the three interfaces, then

$$\cos \gamma = \frac{e_1 - e_2}{e_0} \quad [21]$$

in formal analogy with the Young relation [6]. But e_1, e_2 cannot be interpreted as interfacial forces whose net tangential component cancels that of e_0 .

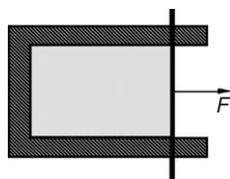


Figure 6 Dupré apparatus for exhibiting surface tension.

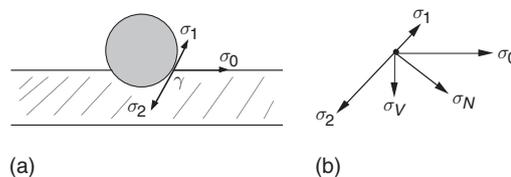


Figure 7 (a) Floating spherical ball; presumed “Young” forces. (b) Normal and vertical components of Young forces; contradiction to presumed equilibrium.

To do so would lead to a net downward force σ_v on the ball (see Figure 7b), contradicting the supposed equilibrium state.

Mathematical and Physical Predictions: Experiments

In the following sections, we study the kinds of behavior imposed on a surface S by the requirement that it appear as solution of one of the indicated equations and boundary conditions. Some of these properties are quite surprising in the context of classically expected behavior of solutions of equations of mathematical physics. The mathematical predictions were, however, corroborated in certain cases experimentally, as we discuss below.

Uniqueness and Nonuniqueness

We begin by considering uniqueness questions. We start with a semi-infinite capillary tube, closed at the bottom, to be partially filled with a prescribed volume of (incompressible) liquid making contact angle γ on the container walls (Figure 8a). If $\kappa \geq 0$, any solution is uniquely determined. That is a quite general theorem, valid for a wide class of domains Ω including all piecewise smooth domains (at the corners of which data of the form [17] cannot be prescribed); formally, data can be omitted on any boundary set of linear Hausdorff measure zero. In this result, no growth conditions need be imposed near the boundary (note that such a statement would be false for solutions of the Laplace equation under Dirichlet boundary conditions).

Next we consider a sessile liquid drop on a horizontal plate (Figure 8b). Again the solution is uniquely determined by the volume and by γ , although the known proof differs greatly from that of the other case.

We now consider a smooth deformation of the base plane, depending on a parameter t , which carries it into the cylinder; that can be done in such a way that the supporting surface is at all times “bowl-shaped,” as in Figure 8c. Since the bowl formation tends to restrict the possible deformations

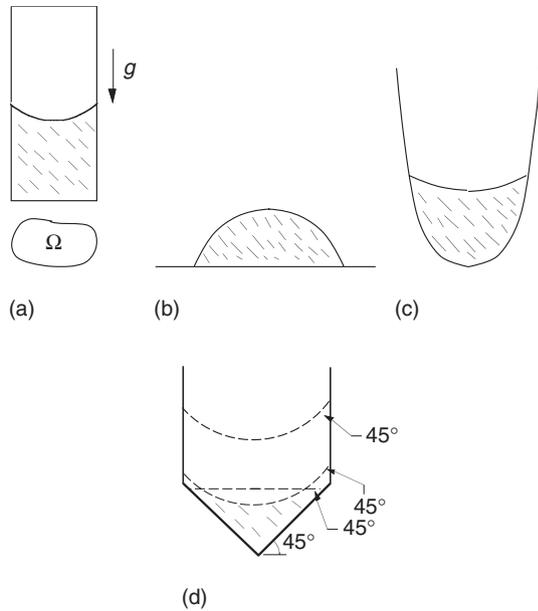


Figure 8 Support configurations: (a) capillary tube, general section; (b) horizontal plate; (c) convex surface appearing during deformation of horizontal plate to capillary tube; and (d) Nonuniqueness of configuration appearing during convex deformation. Reproduced from *Mathematics Intelligencer* 24(3) 2002 21–33 with permission from Springer-Verlag Heidelberg.

of the fluid consistent with smooth contact with the supporting rigid surface, one might expect that the corresponding capillary surface $S(t)$, arising from the identical fluid mass, will for each t be uniquely determined.

That is however not true, even for symmetric configurations. We can see that from the configuration of **Figure 8d**, consisting of a vertical circular cylinder whose base is a 45° cone. We assume a contact angle $\gamma = 45^\circ$ and adjust the radius so that a horizontal surface lying just below the cylinder/cone juncture provides the prescribed volume. This is a formal solution surface. Now fill the configuration with a larger volume, so that the contact line will lie above the juncture. The upper surface will no longer be flat, in view of the 45° contact angle, and takes an appearance as indicated in the figure. Finally, we decrease the fluid volume, keeping all other parameters unchanged. As noted above, the upper surface moves rigidly downward, and it is clear that if the original surface is close enough to the juncture line, then the prescribed volume will be attained before the contact line reaches the juncture. Thus, uniqueness fails.

In this construction as just described, the bounding surface is not smooth; however, one sees easily that the procedure continues to work if the edge and vertex are smoothed locally. In fact, one can carry the procedure to a striking conclusion; by appropriate smoothing, one can construct a bounding surface

admitting an entire continuum of distinct solution interfaces, all with the same contact angle and enclosing the same fluid volume (Gulliver and Hildebrandt; Finn). This can be done for any gravity field. **Figure 9** illustrates seven members of the family of interfaces, in the particular case $\kappa = 0$.

The question immediately arises as to which if any of the continuum of surfaces will be seen in an experiment. In fact, it can be proved that none of the indicated surfaces is mechanically stable (Finn, Concus and Finn, Wentz). Since the indicated family includes all symmetric surfaces that are stationary for the energy functional, we find that any stable stationary configuration must be asymmetric. Thus, we have obtained an example of symmetry breaking, in which all conditions of the problem are symmetric, but for which all physically acceptable solutions are asymmetric.

These results were subjected to computational test by M Callahan using the Surface Evolver software, to experimental test by M Weislogel in a drop tower, and to experimental test by S Lucid in the Mir Space Station. The results of the latter experiment are compared in **Figure 10** with the computer calculations. In both cases, both a local minimizer (potato chip) and a presumed global minimizer (spoon) were observed.

The seven surface interfaces indicated in **Figure 9** all provide the same sum of surface and wetting energy, and bound the same volume of fluid. They all satisfy an eqn [18] with constant H , in accordance with hypotheses of incompressibility and vanishing gravity. Thus, formally, all configurations have identical mechanical energy. The surfaces

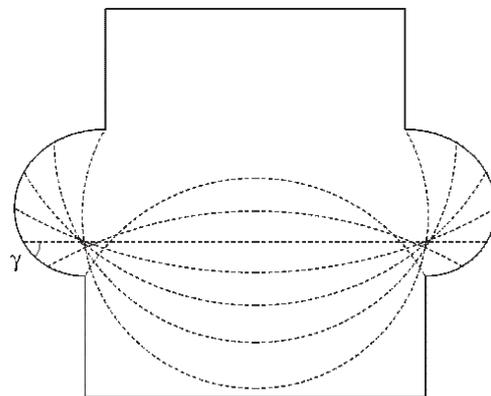


Figure 9 Seven spherical capillary interfaces in an “exotic” container of homogeneous material in zero gravity. All interfaces bound the same volume and have the same sum of free surface and wetting energies. If all pressures above the interfaces are the same, then the pressures below them successively increase as the curvature vectors of the vertical sections change from upwardly to downwardly directed. Reproduced from *Mathematics Intelligencer* 24(3) 2002 21–33 with permission from Springer-Verlag Heidelberg.

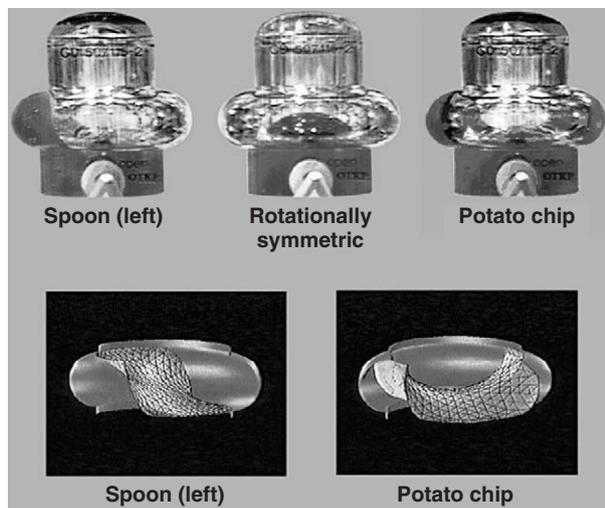


Figure 10 Symmetry breaking in exotic container, $g = 0$. Below: calculated presumed global minimizer (spoon) and local minimizer (potato chip). Above: experiment on Mir: symmetric insertion of fluid (center); spoon (left); potato chip (right). This is a grayscale version of a color figure reproduced from *Journal of Fluid Mechanics*, 224: 383–94, (1991) with permission of Cambridge University Press.

are all spherical caps; however, the radii R of the caps vary considerably. According to Y1 above, the pressure change across each interface is $\Delta p = 2\sigma/R$. Since one may assume the outer region to be a vacuum with zero pressure for all caps, we find that the pressures within the fluids vary greatly among the configurations. One would thus expect that work is done within the fluid in passing from one configuration to another, a circumstance we have excluded by hypothesis when determining the family. From this point of view, the (customary) hypothesis of incompressibility that was used in determining the family is put into significant question; we examine this point in some detail in the section “Compressibility.”

Discontinuous Dependence I

Capillary surfaces can exhibit striking discontinuous dependence on the defining data. As initial example, we consider the behavior of a solution of [18]–[17] at a protruding corner point P of the domain Ω of definition. For simplicity, we assume the corner bounded locally by straight segments, meeting in an opening angle $2\alpha < \pi$, thus forming locally a wedge domain. In anticipation of material to follow, we assume contact angles γ_1 and γ_2 on the respective sides, $0 \leq \gamma_1, \gamma_2 \leq \pi$. One can show that a necessary condition for a solution surface over a domain Ω_δ as in Figure 11 to have a continuous normal vector up to P is that the data point (γ_1, γ_2) lie in the closure of the rectangle R of Figure 12. (This figure includes

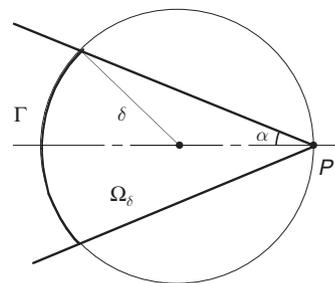


Figure 11 Wedge domain. Reproduced from Finn R “Capillary Surface Interfaces” in *Notices of AMS* 46 No.7 (1999) with permission of the American Mathematical Society.

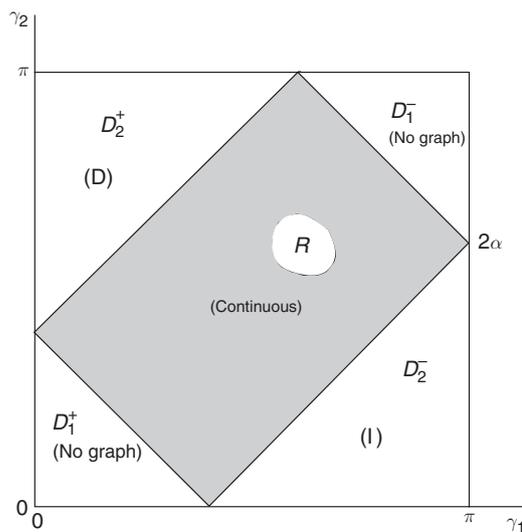


Figure 12 Domain R of data yielding continuous normal to capillary surface in wedge of opening $2\alpha < \pi$. The symbols D and I are clarified in the section “Behavior at a corner point.” Reproduced from “Capillary Wedges Revisited” in *SIAM J. Math. Anal.* 27 No.1 (1996) 56–69 with permission from SIAM.

also additional material anticipating the section “Drops in wedges”).

For data points interior to R , this criterion also suffices for the existence of at least one such solution surface, for any prescribed H ; such surfaces can in fact be produced explicitly as spherical caps (planes if $H = 0$). It remains to discuss what can occur with data arising from the remaining four subregions of the square.

If $(\gamma_1, \gamma_2) \in D_1^\pm$, then there is no solution to [18]–[17] in any neighborhood of the corner point P . On the other hand, an explicit solution for any $H > 0$ can be found as a lower spherical cap on the segment $\gamma_1 + \gamma_2 = \pi - 2\alpha$ that separates D_1^+ from R (see Figure 13, which indicates the equatorial circle). Correspondingly, if $H < 0$ then an explicit solution can be found on the separation line between D_1^- and R . Thus, there is a

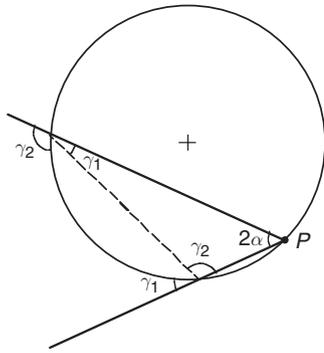


Figure 13 Construction of solution as lower hemisphere; $\gamma_1 + \gamma_2 = \pi - 2\alpha, H > 0$. Reproduced from “Capillary Wedges Revisited” in *SIAM J. Math. Anal.* 27 No.1 (1996) 56–69 with permission from SIAM.

discontinuous change in behavior in crossing from R to either of the D_1 regions.

This behavior was put to experimental test by W Masica, who considered the case $0 < \gamma_1 = \gamma_2 = \gamma < \pi/2$ near the crossing point $\gamma = \gamma_{cr}$ with D_1^+ , for which $\alpha + \gamma_{cr} = \pi/2$. He partially filled a regular hexagonal cylinder of acrylic plastic, successively with two different liquids, making respective contact angles greater or less than γ_{cr} with the plastic. For each liquid, Masica then allowed the cylinder to fall in a 132 m drop tower. **Figure 14** compares the two configurations after about 5 s of free fall. In the case $\gamma > \gamma_{cr}$ he obtained the spherical-cap solution, which in this case covers the entire base domain Ω and appears as an explicit solution of [18]–[17]. When $\gamma < \gamma_{cr}$, the liquid rose to the top of the cylinder near the edges, filling out the edges over the corner points. The surface interface S does not cover Ω , but instead folds back over itself, doubly covering a portion of Ω . Thus, a physical surface appears as it must, but it is not a solution of [18] over Ω .

Discontinuous Dependence II

About 1970, M Miranda raised informally the question, whether a capillary tube Z_0 , whose section

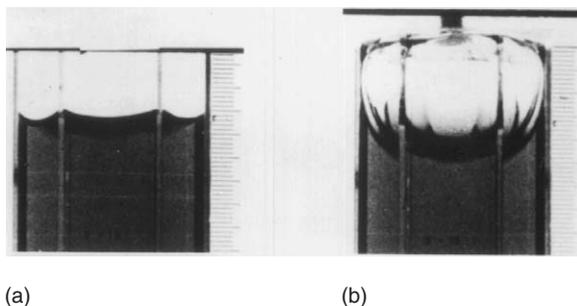


Figure 14 Liquid in hexagonal cylinder, during free fall in drop tower: (a) $\alpha + \gamma > \pi/2$; (b) $\alpha + \gamma < \pi/2$.

Ω_0 lies strictly interior to a section Ω_1 of a tube Z_1 , will raise liquid from an infinite reservoir in a downward directed gravity field to a higher level over Ω_0 than will Z_1 over that subdomain of its section. That is true if both cylinders are circular, and in the intervening years its correctness was established in a number of other cases of particular interest.

Finn and Kosmodem'yanskii, Jr. showed, however, by example that the assertion fails in a large range of cases, and in fact can fail with arbitrarily large height differences, uniformly over Ω_0 . Beyond that, the construction exhibits a strikingly discontinuous change of behavior, under perturbations of a disk as inner domain. Perhaps more remarkably, the assertion can hold with the inner domain a disk, but with discontinuous reversal of behavior as the disk is perturbed to neighboring disks. That was shown in a form of the example given later by Finn, and illustrated in **Figure 15**. Here the outer domain Ω_1 is polygonal, with sides that extend to be tangent to a unit disk Ω_0 , as indicated. The angle γ is to be chosen so that $0 \leq \pi/2 - \gamma \leq \alpha_{min}$, where α_{min} is the smallest of the interior vertex half-angles of Ω_1 . In view of the assumed infinite fluid reservoir, there is no volume constraint, and the governing equation [16] takes the form

$$\text{div } Tu = \kappa u, \quad \kappa = \rho g / \sigma > 0 \quad [22]$$

Taking at first the inner domain to be Ω_0 , it can be shown that for the corresponding solutions u^0 and u^1 of [22], there holds $u^0 > u^1$ over Ω_0 for

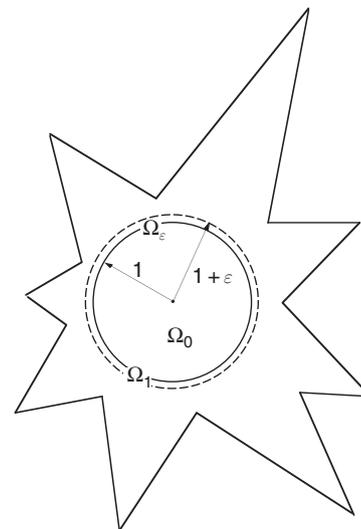


Figure 15 Discontinuous reversal of limiting height behavior. All sides of the polygonal domain Ω_1 are tangent to the unit disk Ω_0 . For the corresponding solution heights u^0 in Ω_0 , u^ϵ in the disk Ω_ϵ of radius $1 + \epsilon$, and u^1 in Ω_1 , there holds $u^1 - u^0 < 0$, for any downward gravity. But $\lim_{\kappa \rightarrow 0} (u^1 - u^\epsilon) = +\infty$, for any $\epsilon > 0$.

any $\kappa > 0$, and thus the Miranda question has a positive answer for that configuration. But if we replace Ω_0 by a concentric disk $\Omega_\varepsilon \subset \Omega_1$ of radius $1 + \varepsilon$, we find

$$\left\{ \inf_{\Omega_\varepsilon} u^1(x; \kappa) - \sup_{\Omega_\varepsilon} u^\varepsilon(x; \kappa) \right\} - \frac{2\varepsilon \cos \gamma}{1 + \varepsilon \kappa} < \frac{1 - \sin \omega}{\cos \gamma} + (1 + \varepsilon) \frac{1 - \sin \gamma}{\cos \gamma} \quad [23]$$

where $\omega = \arccos(\cos \gamma / \sin \alpha)$, and u^ε is the solution of [22], [17] in Ω_ε . Since κ does not appear on the right side of [23], there follows in particular that for any $\varepsilon > 0$, there holds

$$\lim_{\kappa \rightarrow 0} \left\{ \inf_{\Omega_\varepsilon} u^1(x; \kappa) - \sup_{\Omega_\varepsilon} u^\varepsilon(x; \kappa) \right\} = \infty \quad [24]$$

In particular, a negative answer to Miranda’s question appears for all gravity sufficiently small. But as observed above, a positive answer occurs in Ω_0 , for any positive gravity. Thus, the limiting behavior as $\kappa \rightarrow 0$ changes discontinuously, as $\varepsilon \rightarrow 0$. We find that the two limiting procedures cannot be interchanged: for any $x \in \Omega_0$, we obtain

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{\kappa \rightarrow 0} \{u^1(x; \kappa) - u^\varepsilon(x; \kappa)\} &= +\infty. \\ \lim_{\kappa \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \{u^1(x; \kappa) - u^\varepsilon(x; \kappa)\} &\equiv \text{const.} < 0 \end{aligned} \quad [25]$$

Existence Questions I

For the general equation [20] there is an established literature on existence of surfaces containing a prescribed space curve. There is very little literature relating to the capillarity boundary condition that the solution surface S meet a prescribed “support” surface W in a prescribed angle γ . The existence of at least one such surface interior to a prescribed sufficiently smooth closed space domain was proved by Almgren, and then Taylor proved smoothness at the contact curve. These are abstract theorems that are basic for the theory but in general do not provide specific information in particular cases of interest.

Special interest attaches to the nonparametric cases [16] or [18] with boundary condition [17], especially in view of the discontinuous behavior properties described above. These cases were studied in depth by a number of authors, with results that put the above examples into some perspective.

M Emmer proved the existence of a unique solution of [16]–[17] for any compact Ω having Lipschitz boundary with Lipschitz constant L such that $\sqrt{1 + L^2} \cos \gamma < 1 - \varepsilon$ for some $\varepsilon > 0$. Finn and

Gerhardt (F and G) extended this condition, and showed in particular that solutions exist in general in piecewise smooth Ω . This result contrasts with the zero-gravity case [18] discussed in the section “Existence questions II,” for which solutions fail to exist when $\sqrt{1 + L^2} \cos \gamma > 1$ at a protruding corner (see the section “Discontinuous dependence I”). However, in the cases $\sqrt{1 + L^2} \cos \gamma > 1$ studied by F and G the solution $u(x)$ is necessarily unbounded in the corner. This condition is equivalent to $\alpha < |\gamma - \pi/2|$ at the corner. Concus and Finn showed that if $\alpha \geq |\gamma - \pi/2|$ in a neighborhood Ω_δ of a corner with rectilinear sides, as indicated in Figure 11, then the solution $u(x)$ satisfies

$$|u(x; \kappa)| < \frac{2}{\kappa \delta} + \delta \quad [26]$$

independent of α, γ in the range considered. Here it is assumed that [16] is normalized so that $\lambda = 0$; when $\kappa \neq 0$ this can always be achieved by adding a constant to u . On the other hand, if $\alpha < |\gamma - \pi/2|$, then

$$u(x; \kappa) \approx \frac{\cos \vartheta - \sqrt{k^2 - \sin^2 \vartheta}}{k \kappa r} \quad [27]$$

where $k = \sin \alpha / \cos \gamma$ and ϑ is polar angle relative to a bisector at the vertex; hence u becomes unbounded as $O(1/r)$. Thus, the behavior changes discontinuously as the configuration for which $\alpha = |\gamma - \pi/2|$ is crossed.

This prediction was corroborated by T Coburn in a “kitchen sink” experiment in the Medical School at Stanford University. Coburn formed a wedge using two sheets of acrylic plastic, resting on a glass plate, and inserted a drop of distilled water at the base of the wedge. Initially, the wedge was opened sufficiently that $\alpha + \gamma \geq \pi/2$, and he obtained the configuration of Figure 16a, with the maximum height slightly lower than that indicated by [26]. By closing down the angle slightly, the liquid rose to over ten times that height, as shown in Figure 16b. This experiment was later repeated by Weislogel under laboratory conditions; it incidentally establishes the contact angle of water and acrylic plastic in the Earth’s atmosphere as $80^\circ \pm 2^\circ$.

The indicated procedure provides in general a very accurate way to measure contact angles, when the angle is not far from $\pi/2$. For γ near zero or π in the Earth’s gravity field, the discontinuity is confined to a microscopic neighborhood of the vertex, and can be difficult to observe. This technical difficulty was addressed by Fischer and Finn, who introduced “canonical proboscis” domains, the theory of which was further developed by Finn and

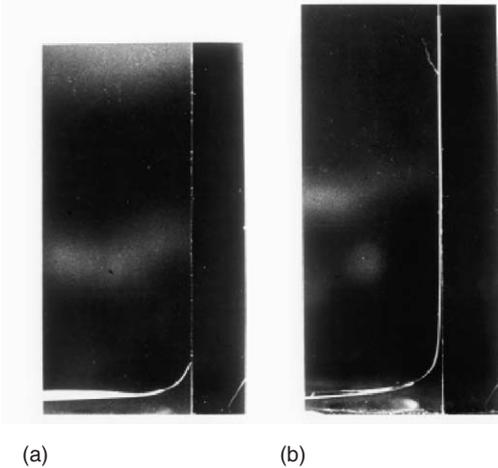


Figure 16 Distilled water in wedges formed by acrylic plastic plates; $g > 0$. (a) $\alpha + \gamma > \pi/2$; (b) $\alpha + \gamma < \pi/2$. Reproduced from P Concus and R Finn, “On Capillary Free Surfaces in a Gravitational Field” in *Acta Math* 132 (1974) 207–223 with permission of Institut Mittag-Leoffler.

Leise and by Finn and Marek. For such domains the change in behavior is not strictly discontinuous, but it is nearly so, and it extends over large portions of the cylinder section, so that it is easily observable. Concus, Finn, and Weislogel conducted space experiments, demonstrating the feasibility of the method as a means for measuring contact angles in general ranges.

In [26]–[27] no growth conditions at the corner are imposed; the estimates hold for every solution defined in Ω_δ and assuming the prescribed data on the side walls, with no data prescribed at the vertex. The formula [27] is the initial term of a formal asymptotic expansion of the solution, in powers of r . Miersemann obtained the complete expansion, asymptotic to every order, when $\alpha < |\gamma - \pi/2|$. He obtained somewhat less complete information in the bounded case [26].

Chen, Finn, and Miersemann provided a form of [27] that is applicable for any data (γ_1, γ_2) on the respective sides of the wedge, that arise from the D_1^\pm regions of Figure 12. Lancaster and Siegel and independently Chen, Finn, and Miersemann showed that if $-2\alpha \leq \gamma_1 + \gamma_2 - \pi \leq 2\alpha$, then every solution is bounded at the vertex. This result holds also for the zero gravity eqn [18].

In the case of [18], Concus and Finn showed that in the D_1^\pm regions no solution exists, regardless of H . Again, this result holds without growth conditions.

From these considerations and from remarks in the section “Discontinuous dependence I” follows that for data in D_2^\pm , all solutions either of [18] or of [16] are bounded but have discontinuous derivatives at the vertex P . Extrapolating from the behavior of

particular computed solutions, Concus and Finn conjectured that all solutions of [18] or of [16] that arise from data in D_2^\pm are discontinuous at P . A number of attempts to prove or to disprove this conjecture have till now been unsuccessful.

An existence theorem for [16]–[17] alternative to that of Emmer was obtained independently by Ural'tseva, using a very different approach. This procedure yielded smoothness estimates up to the boundary, but required a hypothesis of boundary smoothness, so that the result does not mesh with the discontinuous dependence behavior as does that of Emmer. Later versions of the existence result, again under boundary smoothness requirements, were given by Gerhardt, Spruck, and Simon and Spruck. In the procedure introduced by Emmer, the boundary trace is shown to exist only in a very weak sense (which, however, suffices for a uniqueness proof). The later work can be adapted to show that the Emmer solutions are smooth on the smooth parts of $\partial\Omega$.

None of the above procedures provides existence for the zero gravity case [18]. As we shall see in the following section, that is not an accident of the methods, but reflects subtle properties of the equations.

Existence Questions II

We consider here the zero-gravity case [18], over a domain Ω bounded by a piecewise smooth curve Σ , under the boundary condition [17]. Integrating [18] over Ω and using [17], we find $2H|\Omega| = |\Sigma| \cos \gamma$. Let $\Omega^* \subset \Omega$, $\Sigma^* = \Sigma \cap \partial\Omega^*$, $\Gamma = \Omega \cap \partial\Omega^*$. The same procedure over Ω^* , using that $|Tu| < 1$ for any $u(x, y)$, leads to the bound

$$\Phi[\Gamma; \gamma] > 0 \quad [28]$$

where Φ is defined by

$$\Phi[\Gamma; \gamma] \equiv |\Gamma| - |\Sigma^*| \cos \gamma + 2H|\Omega^*| \quad [29]$$

The inequality [28] must hold for any choice of $\Omega^* \subset \Omega$. This provides a necessary condition for existence of a solution to [18]–[17] in Ω . E Giusti showed that when Ω^* is interpreted in a generalized sense as a Caccioppoli set, the condition [28] becomes also sufficient for existence.

It is easy to give specific examples of convex analytic domains Ω , in which subdomains Ω^* can be found such that [28] fails. Thus, the general existence results for [16] do not carry over to [18], regardless of local domain smoothness. Nevertheless, in many cases of interest (e.g., a circular disk or an ellipse that is not too eccentric), solutions of [18]–[17] do exist for any γ and are well behaved. Finn investigated the condition [28] in general by showing the existence of a system of arcs $\{\Gamma\} \subset \Omega$

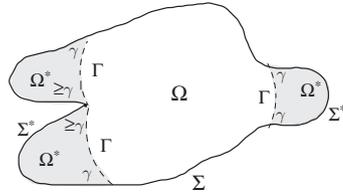


Figure 17 Extremal configuration for the functional Φ .

that minimize Φ . All such arcs are circular of radius $1/2H$, and meet Σ either at smooth points in an angle γ , or else at a reentrant corner point in an angle $\gamma^* \geq \gamma$, measured on the side of Γ opposite to that into which the curvature vector points (Figure 17). All minimizing configurations are bounded by arcs of that form, although not all such configurations minimize. In a typical situation one will encounter only a finite number of such arcs, in which case only a finite number of cases need be examined. If $\Phi > 0$ in each such case, then a solution of [18]–[17] exists for the given Ω and γ . It may occur that no such arcs exist; we then observe that since $\Phi[\emptyset; \gamma] = \Phi[\Sigma; \gamma] = 0$, Φ cannot become nonpositive for any $\Omega^* \subset \Omega$ unless a minimizing Γ can be found in Ω , contradicting the assumed nonexistence of minimizers. Thus, the criterion is then vacuously satisfied, and we conclude that a solution of [18]–[17] exists.

One has, of course, to ask what happens physically in cases for which $\Phi[\Gamma; \gamma] \leq 0$ for some Γ as above. The possible modes of behavior were studied in particular cases by Tam and later by deLazzer, Langbein, Dreyer, and Rath; Finn and Neel characterized the general case. Formally, the fluid rises to infinity throughout domains Ω^* of the form indicated, but with H replaced by a value $H^- < H$; on the opposite side of the circular arcs Γ , the fluid is asymptotic to the vertical cylinders over Γ . In a physical situation, the fluid will rise to the top of the container in a nearly cylindrical region adjacent to a portion of the container walls, approximating the indicated behavior and partially wetting the top of the container. One sees that behavior in Figure 14b, in which the fluid fills out regions adjacent to the corners. An analogous configuration would still be observed if the corners were smoothed locally. If insufficient fluid is available, a portion of the base Ω could become unwetted.

Behavior at a Corner Point

Lancaster and Siegel (L and S) studied the behavior of the limits (which they designate by Ru) of bounded solutions of [16] or of [18] along radial segments

tending to a corner point P of a domain Ω . These limits can exhibit remarkable idiosyncratic behavior. For simplicity of exposition, we restrict ourselves here to rectilinear boundary segments at P , and assume constant boundary angles $\gamma_1, \gamma_2 \neq 0, \pi$ on the two sides. L and S prove first that the limits Ru exist and vary continuously with direction of approach; then they show the existence of “fan” regions of directions adjacent to those of the sides, in which the limits are constant independent of direction, see Figure 18. They obtain that if the opening angle 2α at P satisfies $2\alpha < \pi$, then for data in the rectangle R of Figure 12 the fans overlap (see Figure 18a), so that the solution is necessarily continuous at P . For data in D_2^+ , the solution decreases from the γ_1 side Σ_1 to the γ_2 side Σ_2 (“D” behavior), subject to the Concus–Finn conjecture (see the section “Existence questions I”), with the reverse behavior (“I”) in D_2^- . Concus and Finn showed that if $2\alpha < \pi$ then in D_1^\pm there is no bounded solution of [16]–[17] or [18]–[17] as a graph. For [16]–[17], unbounded solutions do however exist for such data (see the section “Existence questions I”).

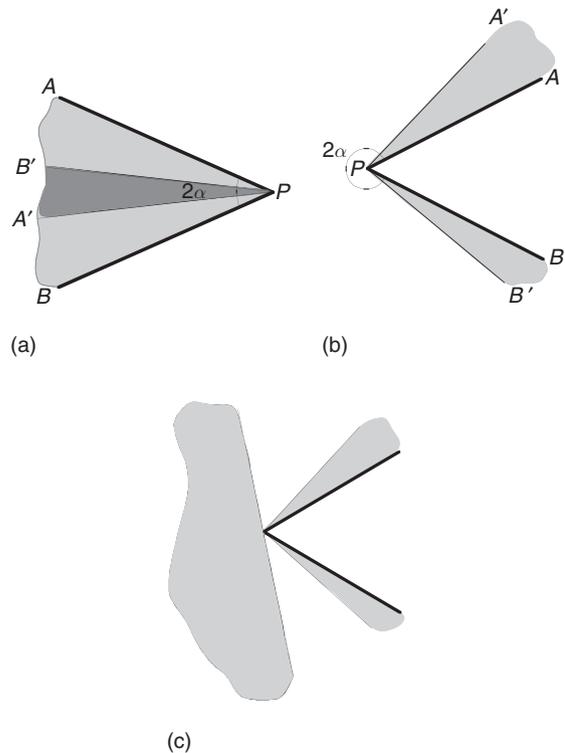


Figure 18 (a) Fan domains APA' and BPB' of constant limiting values; $2\alpha < \pi$ so that the fans overlap when data are in R . (b) $2\alpha > \pi$; case 1. Fans APA' and BPB' of constant radial limits appear. Limiting value changes strictly monotonically as approach direction changes from $A'P$ to $B'P$. (c) $2\alpha > \pi$; case 2. In addition to the two fans adjacent to the sides of the wedge, a half plane of constant radial limits appears.

If $2\alpha > \pi$, then the fans do not overlap, and in fact continuity at P cannot in general be expected. Outside the indicated fan regions adjacent to the wedge sides, the limit values either change strictly monotonically with angle of approach, as in Figure 18b, or else they do so except for approaches within a third, central fan, which covers a full half-space, and interior to which the limiting values again remain constant, see Figure 18c. L and S give an example under which that behavior actually occurs. Remarkably, in the example the prescribed data are the same on both boundary segments. The solution is nevertheless discontinuous at P , with an interval in which the radial limit increases, another interval in which it decreases, two fans of constant limit adjacent to the sides, and a fan of breadth π in-between.

General conditions for continuity at a reentrant corner ($2\alpha > \pi$) have not yet been established. L and S give a sufficient condition, depending on a hypothesis of symmetry. Since no such hypothesis is needed when $2\alpha < \pi$, one might at first expect it to be superfluous. However, Shi and Finn showed that by introducing an asymmetric domain perturbation that in an asymptotic sense can be arbitrarily small, the solution can be made discontinuous at P . That can be done without affecting any other hypotheses of the L and S theorem.

In as yet unpublished work, D Shi characterized all possible behaviors at a reentrant corner, subject to the validity of the Concus–Finn conjecture at a protruding corner. If $\kappa \geq 0$ then all solutions of [16] or of [18] in a neighborhood of P in Ω are bounded at P . The further behavior depends on the particular data, and is indicated in Figure 19. Note the analogy with Figure 12, although the interpretations in the figures differ in detail. Here the symbol I denotes strictly increasing from the side Σ_1 to Σ_2 , except on the fan regions of constant limits; ID denotes constancy on a fan adjacent to Σ_1 , then strictly increasing, then constancy on a fan of opening π , then strictly decreasing, then constancy on a fan adjacent to Σ_2 . D and DI are defined analogously. All cases can be realized in particular configurations.

Drops in Wedges

Closely related to the material just discussed is the question of the possible configurations of a connected drop of liquid placed into a wedge formed by intersecting plates of possibly differing materials, in the absence of gravity. Thus, one has distinct contact angles γ_1, γ_2 on the two plates. Finn and McCuan showed that if $(\gamma_1, \gamma_2) \in R$ then the only

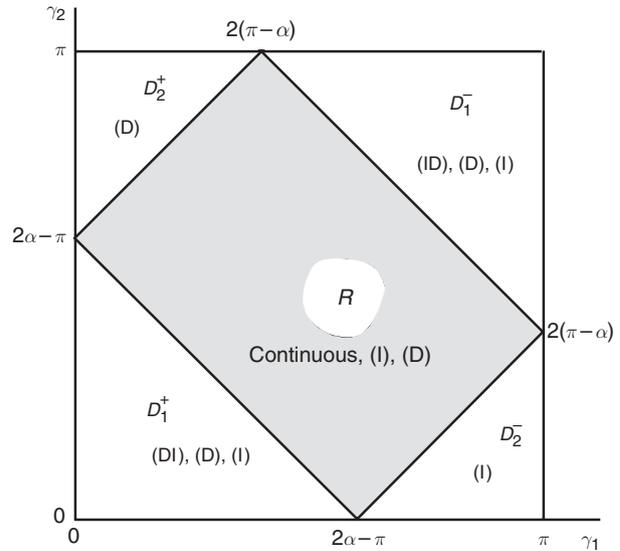


Figure 19 $\pi < 2\alpha < 2\pi$. Possible modes of behavior. Reproduced with permission from the *Pacific Journal of Mathematics*.

possibility is that the drop surface S is part of a sphere. For data in D_1^\pm , no such drop can exist, barring exotically singular behavior at the vertex points where the edge of the wedge meets S .

For data in D_2^\pm the situation is less clear. Concus, Finn, and McCuan (CFM) showed that local behavior exhibiting such data is indeed possible; however, they conjectured that such behavior cannot occur for simple drops. In conjunction with the above results, they were led to the conjecture that the free surface S of any liquid drop in a planar wedge, that meets the wedge in exactly two vertices and the wedge faces in constant contact angles γ_1, γ_2 , is necessarily spherical. Here it is supposed only that $0 \leq \gamma_1, \gamma_2 \leq \pi$.

The behavior of a drop of prescribed volume, as the data move from the midpoint of R to the D regions along parallels to the sides of R , is displayed in Figure 20. As one moves into the D_2^\pm regions, the drop detaches from one side of the wedge and becomes a spherical cap resting on a single planar surface, in accord with the above conjecture. As D_1^- is approached, the liquid becomes a drop of very large radius that fills out a long thin region in the wedge, and disappears to infinity as the boundary of R is crossed. However, as D_1^+ is entered, the configuration transforms smoothly into a spherical liquid bridge, connecting the two faces of the wedge without contacting the wedge line.

Stability Questions

A number of authors, for example, Langbein, Vogel, Finn and Vogel, Steen, and Zhou, have studied the

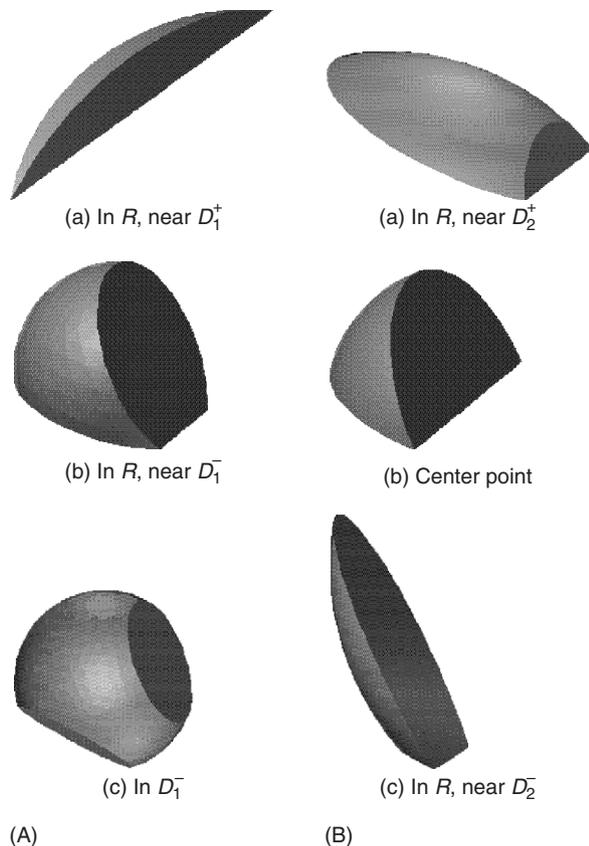


Figure 20 (A) Drop configurations in wedge with opening angle $2\alpha = 50^\circ$, for three data positions on the line $\gamma_1 = \gamma_2 = \gamma$ (a) $\gamma = 70^\circ$ (in R , near D_1^+); (b) $\gamma = 90^\circ$ (in R , near D_1^-); (c) $\gamma = 110^\circ$ (in D_1^-). The first two cases yield edge blobs, the third a spherical tube that does not contact the edge line. (B) Drop configurations in a wedge of opening angle $2\alpha = 50^\circ$, for three data choices in R , on the line $\gamma_1 = \pi - \gamma_2 = \gamma$; (a) $\gamma = 70^\circ$ (near D_2^+); (b) $\gamma = 90^\circ$ (center of R); (c) $\gamma = 35^\circ$ (near D_2^-). As D_2^\pm is entered, original boundary conditions can no longer be satisfied by spherical drop, but configuration changes smoothly into drop on single plane, with prescribed data for that plane. Reproduced with permission from Concus P, Finn R and McCuan J (2001) Liquid bridges, edge blobs, and Scherk-type capillary surfaces. *Indiana University Mathematics Journal* 50: 411–441.

stability of liquid drops trapped between parallel plates, forming an annular liquid bridge joining the plates under the capillarity boundary condition of prescribed contact angles γ_1, γ_2 on the respective plates. These studies consider the effects of disturbances within the fluid, assuming the plates are rigid and perfectly parallel. CFM show that from the point of view of physical prediction, the results of these studies may be open to some question. Specifically, they show that unless the drop is initially of spherical form, then infinitesimal tilting of one of the plates always results in a discontinuous transition of the drop form. Depending on the particular data, the transition can be to a spherical drop; however, it can also occur that the tilting

causes the entire fluid to disappear to infinity in the wedge.

CFM proved that if a connected liquid mass with spherical outer surface S cuts off areas $|W_1|, |W_2|$ from plates Π_1, Π_2 which it meets in angles γ_1, γ_2 , as in Figure 20, then

$$-\sum_1^2 |W_j| \cos \gamma_j + |S| = \frac{3|V|}{R} \quad [30]$$

where $|S|$ denotes area of the spherical free surface interface, $|V|$ the enclosed volume, and R the radius. An immediate consequence is that the mechanical energy E of the configuration is

$$E = \frac{3\sigma|V|}{R} \quad [31]$$

where σ is surface tension. Using this result, they show that if a spherical liquid mass meets two wedge faces in angles γ_1, γ_2 in the absence of gravity, then the configuration has smaller mechanical energy than does any connected liquid mass of the same volume that meets only one of the faces in the contact angle for that face. In turn, the drop on a single face has smaller energy than does a spherical ball of the same volume that meets no face. Note that in all zero-gravity cases for which stability relative to plate tilting can be expected, the liquid mass must be spherical.

Compressibility

Until very recently, all literature on capillarity was based on a hypothesis that the body of the fluid is incompressible. Indeed, from the point of view of macroscopic mechanical measurements, most liquids are nearly incompressible. But all liquids are also to some extent compressible, and this property was even conceptually essential in our characterization in the section “Gauss’ contribution: the energy method” of the surface energy, even for the nominally incompressible case. It is as yet unclear to what extent the compressibility properties of the bulk liquid will influence the physical predictions of the theory. In this connection, see the remarks at the end of the section “Uniqueness and nonuniqueness.”

The Equations I

Finn derived two possible equations extending [16] and [17], arising from different modelings. Both characterize equilibrium points as stationary points for the mechanical energy, and both are based on a hypothesized pressure–density relation $\rho = \rho_0 + \chi(p - p_0)$. The first equation takes account of the change in density with height, arising from

the gravity field. For a container consisting of a semi-infinite vertical cylinder, closed at the bottom, one obtains

$$\operatorname{div} Tu = \frac{\rho_0 g}{\sigma} u + \chi g(1 - \cos \omega) + \lambda \quad [32]$$

where ω is the angle between the upward directed surface normal and the vertical axis, and λ is to be determined by a volume constraint. Athanassenas and Finn proved that for a general smooth domain Ω , prescribed γ , and prescribed fluid mass M subject to the restriction

$$M < \rho_0 |\Omega| / \chi g \quad [33]$$

there exists exactly one solution of [32] achieving the boundary data γ .

The condition [33] is necessary for existence with the prescribed mass.

The methods used for this theorem do not permit regularity conditions to be relaxed to allow domains with corner points. An approximation procedure yields an existence theorem for such cases, however the uniqueness proof then fails; it can be replaced by a weaker result, estimating the difference between two eventual solutions: Let u, v , be solutions of [32] in a piecewise smooth domain Ω , and suppose $v \cdot Tu \leq v \cdot Tv$ on $\Sigma = \partial\Omega$ except at the corner points, where no data are prescribed. Then

$$u \leq v + \chi \sigma / \rho_0 \quad [34]$$

throughout Ω .

Note that in this result, no growth condition is imposed at the corner points. It can happen that both u and v are unbounded at a corner point; nevertheless, [34] holds uniformly over Ω .

The solutions of [32] emulate many of the characteristics of solutions of [16]. Notably, there is again a dichotomy of behavior, depending on opening angle 2α at a corner point, with all solutions either bounded, or unbounded with growth like $1/r$.

The Equations II

If in addition to taking account of the change of density with height, one accounts for the energy change due to expansion or contraction of volume elements with changing density, one is led to the equation

$$\operatorname{div} Tu = \frac{\rho_0 - \chi p_0}{\sigma \chi} (e^{\chi g u} - 1) + \chi g(1 - \cos \omega) + \Lambda \quad [35]$$

Here the changes from the incompressible case are much more significant than for [32]. In order to ensure stable behavior of solutions, it seems appropriate to impose the condition $\rho_0 > \chi p_0$. The general

existence theorem above can no longer be expected; it is possible to give explicit examples of analytic domains, and constant data γ , for which no solution of the problem exists. Thus, even in a large downward gravity field, the solutions can emulate the behavior of solutions of [18]. That can happen, however, only for data γ exceeding $\pi/2$. The condition [33] is again necessary for existence.

For eqn [34], Λ cannot be eliminated by addition of a constant to the solution, and its determination creates a new level of difficulty toward solution of the physical existence question. Athanassenas and Finn proved unique existence of solutions of [35], [17] for a capillary tube of general smooth section Ω dipped into an infinite liquid bath (which corresponds to $\Lambda = 0$), when $0 \leq \gamma \leq \pi/2$. If $\gamma > \pi/2$ then solutions do not always exist; it can happen that the surface moves down to the bottom of the tube, regardless of the depth of immersion. Under a hypothesis of radial symmetry, Finn and Luli were able to prove the existence of solutions with prescribed mass in a semi-infinite cylinder closed at the bottom, in the range $0 \leq \gamma < \pi$, and uniqueness if $0 \leq \gamma \leq \pi/2$. Note that in this case, values $\gamma > \pi/2$ are not excluded. For large enough mass, the surface will always cover the base of the tube.

Closing Remarks

This brief survey is intended only as a general indication of the current state of the theory; much material of interest could not be included. Nor have we addressed hysteresis effects on contact angle. Detailed references to the material discussed and also to further information can be found in the articles listed below. More recent publications can be located by following links in MathSciNet or Zentralblatt.

Acknowledgmnt

I owe a special debt of thanks to my colleague Paul Concus, who read the material in detail and provided many effectual suggestions, leading to a much-improved exposition.

See also: Compressible Flows: Mathematical Theory; Interfaces and Multicomponent Fluids; Newtonian Fluids and Thermohydraulics.

Further Reading

References for text material and for further reading are cited in the expository articles:

- Finn R (2002a) *Milan Journal of Mathematics* 70: 1–23.
- Finn R (2002b) *Mathematical Intelligencer* 24: 21–33.

Cartan Model see Equivariant Cohomology and the Cartan Model

Cauchy Problem for Burgers-Type Equations

G M Henkin, Université P.-M. Curie, Paris VI,
Paris, France

© 2006 Elsevier Ltd. All rights reserved.

Burgers Type Equations

We consider here two types of equations: the scalar partial differential equations (PDEs) of the form

$$\frac{\partial f}{\partial t} + \varphi(f) \frac{\partial f}{\partial x} = \varepsilon \frac{\partial^2 f}{\partial x^2}, \quad \varepsilon > 0 \quad [1]$$

$f = f(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$, and the scalar difference-differential equations of the form

$$\frac{\partial F}{\partial t} + \varphi(F) \frac{F(x, t) - F(x - \varepsilon, t)}{\varepsilon} = 0, \quad \varepsilon > 0 \quad [2]$$

$F = F(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$.

Equation [1] for the case of linear $f \mapsto \varphi(f)$ was called as Burgers equation by Hopf (1950), who justified this by the remark: “equation was first

$$\frac{\partial f}{\partial t} + f \frac{\partial f}{\partial x} = \varepsilon \frac{\partial^2 f}{\partial x^2}$$

introduced by J. M. Burgers (1940) as a simplest model to the differential equations of fluid flow”. In fact, eqn [1] for linear $\varphi(f)$ was introduced earlier in 1915 by Bateman. Equation [1] for general $\varphi(f)$ appeared later in very different models, for example, in the model for displacement of oil by water, in a model of road traffic, etc.

For $\varphi(f) = a + b \cdot f$, Hopf and Cole have studied [1] basing on the substitution

$$f = -\frac{1}{b} \left(a + \varepsilon \frac{\partial g}{\partial x} / g \right)$$

reducing [1] to the heat equation

$$\frac{\partial y}{\partial t} = \varepsilon \frac{\partial^2 g}{\partial x^2}$$

This transformation (often called as the Hopf-Cole transform) appeared for the first time in 1906 in the book of Forsyth “Theory of differential equations.”

Equation [2] first appeared for $\varphi(F) = a + b \cdot F$, $\varepsilon = 1$, $x = n \in \mathbb{Z}$, in Levi, Ragnisco, Bruchi (1983) as a semidiscrete equation reducible to the linear equation

$$\frac{dG_n(t)}{dt} = a(G_{n-1}(t) - G_n(t))$$

by the substitution

$$F(n, t) = -\frac{a}{b} \left(\frac{G_n(t) - G_{n-1}(t)}{G_n(t)} \right)$$

Equation [2] for general $\varphi(F)$ was introduced by Henkin, Polterovich (1991) for the description of a Schumpeterian evolution of industry. For any $\varepsilon > 0$, one can consider [2] as the family of difference-differential equations, depending on the parameter $\theta = \{x/\varepsilon\} \in [0, 1)$, where $\{x/\varepsilon\}$ denotes the fractional part of x/ε . For physical applications of [1] (see Gelfand (1959), Landan and Lifschitz (1968), Lax (1973)), the inviscid case ($\varepsilon = +0$) is the most interesting. But, for some special physical models and for some social and biological applications (see Henkin, Polterovich (1991), Serre (1999)), the interesting case concerns eqn [2] with $\varepsilon = 1$ and $x \in \mathbb{Z}$.

The results considered in this article concern mainly the Cauchy problem for eqns [1] and [2] with initial data $f(x, 0)$, $F(x, 0)$ satisfying the conditions

$$\begin{aligned} f(x, 0) &\rightarrow \alpha^\pm, \quad x \rightarrow \pm\infty \\ \int_{-\infty}^0 |f(x, 0) - \alpha^-| dx & \\ + \int_0^{\infty} |\alpha^+ - f(x, 0)| dx &< \infty \end{aligned} \quad [3]$$

and correspondingly

$$\begin{aligned} F(k\varepsilon + \theta\varepsilon) &\rightarrow \alpha^\pm, \quad k \rightarrow \pm\infty \\ \sum_{k=-\infty}^0 |F(k\varepsilon + \theta\varepsilon, 0) - \alpha^-| & \\ + \sum_{k=0}^{\infty} |\alpha^+ - F(k\varepsilon + \theta\varepsilon, 0)| &< \infty \end{aligned} \quad [4]$$

where $\alpha^- \leq \alpha^+$, $\theta \in [0, 1)$ and the mapping $\theta \mapsto \{F(k\varepsilon + \theta\varepsilon, 0) - \alpha^{\text{sgn } k}, k \in \mathbb{Z}\} \in l^1$ is smooth.

The standard classical questions concerning Cauchy problems [1], [3] and [2], [4], namely those relating to existence, unicity, regularity, and conservation laws are well established (see Oleinik (1959), and Serre (1999)). This section formulates only those which are essential for the study of asymptotic behavior of solutions $f(x, t)$ and $F(x, t)$, when $t \rightarrow \infty$ or $\varepsilon \rightarrow 0$, and of the relation between vanishing viscosity and difference scheme approximations for inviscid Burgers type equations.

One can see that asymptotic behavior of solutions of [2], [4] when $\varepsilon \rightarrow +0$ is not the same as the asymptotic behavior of [1], [3] when $\varepsilon \rightarrow +0$, in spite of fact that in the limiting case $\varepsilon = +0$ both [1] and [2] look identical. It can be explained by the fact that eqn [2] can be interpreted as a semidiscrete approximation of the nonconservative (nonphysical) equation

$$\frac{\partial F}{\partial t} + \varphi(F) \frac{\partial F}{\partial x} = \frac{\varepsilon}{2} \varphi(F) \frac{\partial^2 F}{\partial x^2}$$

However, the problem [2], [4] can be naturally transformed into conservative (physical) initial problem. Indeed, the substitution

$$f = \int_0^F \frac{dy}{\varphi(y)}$$

(under condition of integrability of $1/\varphi(y)$) transforms [2] into the equation

$$\frac{\partial f(x, t)}{\partial t} + \frac{\psi(f(x, t)) - \psi(f(x - \varepsilon, t))}{\varepsilon} = 0 \quad [5]$$

where $\psi'(f) = \varphi(F)$. Equation [5] is the so-called monotone one-sided semidiscrete approximation of conservative viscous equation,

$$\frac{\partial f}{\partial t} + \varphi(F) \frac{\partial f}{\partial x} = \frac{\varepsilon}{2} \frac{\partial}{\partial x} \left(\varphi(F) \frac{\partial f}{\partial x} \right) \quad [6]$$

where

$$f(x, 0) \rightarrow \int_0^{\alpha^\pm} \frac{dy}{\varphi(y)}, \quad x \rightarrow \pm\infty$$

The results of finite-difference approximations for nonlinear conservation laws (see A. Harten, J. Hyman, P. Lax (1976)) explain both the similarity of behavior of [6] and [5] as well as some difference in the behavior of [1] and [2].

For further exposition the following assumption is useful:

Assumption 1 Let φ in [1], [2] be a positive and continuously differentiable function on the interval $[\alpha^-, \alpha^+]$. Let φ' have only isolated zeros.

From references one can deduce the following general properties of Cauchy problems [1], [3] and [2], [4].

Theorem 0 Under Assumption 1, we have:

- (i) There exists a unique (weak) solution $f(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$ of the problem [1], [3]; this solution is necessarily smooth for $t > 0$; besides, it satisfies the following conservation laws for $t > 0$:

$$f(x, t) \rightarrow \alpha^-, \quad x \rightarrow -\infty$$

$$f(x, t) \rightarrow \alpha^+, \quad x \rightarrow +\infty$$

$$\begin{aligned} \frac{d}{dt} \left[\int_0^\infty (\alpha^+ - f(x, t)) dx - \int_{-\infty}^0 (f(x, t) - \alpha^-) dx \right] \\ = \int_{\alpha^-}^{\alpha^+} \varphi(y) dy \end{aligned}$$

Moreover, if the initial value $f(x, 0)$ is nondecreasing as a function of x , then solution $f(x, t)$ is nondecreasing as a function of x for all $t \geq 0$.

- (ii) There exists a unique solution $F(x, t)$ $x \in \mathbb{R}$, $t \in \mathbb{R}_+$ of the problem [2], [4]; this solution is smooth for $t > 0$; besides, it satisfies the following conservation laws for $t > 0$ and $\theta \in [0, 1)$:

$$F(k\varepsilon + \theta\varepsilon, t) \rightarrow \alpha^-, \quad k \rightarrow -\infty$$

$$F(k\varepsilon + \theta\varepsilon, t) \rightarrow \alpha^+, \quad k \rightarrow +\infty$$

$$\begin{aligned} \frac{d}{dt} \left[\sum_{k=1}^\infty \int_{F(k\varepsilon + \theta\varepsilon, t)}^{\alpha^+} \frac{dy}{\varphi(y)} - \sum_{k=-\infty}^0 \int_{\alpha^-}^{F(k\varepsilon + \theta\varepsilon, t)} \frac{dy}{\varphi(y)} \right] \\ = \alpha^+ - \alpha^- \end{aligned}$$

Moreover, if for some $\theta \in [0, 1)$ the $F(k\varepsilon + \theta\varepsilon, 0)$ is nondecreasing as a function of $k \in \mathbb{Z}$ then solution $F(k\varepsilon + \theta\varepsilon, t)$ is also nondecreasing as a function of $k \in \mathbb{Z}$ for all $t \geq 0$ and the same θ .

Gelfand's Problem and Iljin-Oleinik Theorem

The main results considered in this article are related to the following problem, formulated explicitly by Gelfand (1959): to find the asymptotic ($t \rightarrow \infty$) of the solution $f(x, t)$ of the eqn [1] with the initial condition

$$f(x, 0) = \begin{cases} \alpha^\pm, & \text{if } \pm x > \pm x^\pm \\ f^0(x), & \text{if } x \in [x^-, x^+] \end{cases} \quad [7]$$

where $\alpha^- \leq \alpha^+$.

Gelfand found a solution to this problem for the inviscid case $\varepsilon = +0$ with initial conditions $f(x, 0) = \alpha^-$ if $x < 0$, and $f(x, 0) = \alpha^+$ if $x \geq 0$ (see below), and remarked that it would be interesting to prove that the main term of the asymptotic ($t \rightarrow \infty$) of $f(x, t)$ satisfying [1], [7] coincides with the solution of [1], [7] for $\varepsilon = +0$.

Gelfand’s problem admits natural extension for eqn [2] with the initial conditions

$$\begin{aligned} F(x, 0) &= \alpha^\pm, & \text{if } \pm x > \pm x^\pm \\ F(x, 0) &= F^0(x), & \text{if } x \in [x^-, x^+] \end{aligned} \quad [8]$$

Let us introduce, for $u \in [\alpha^-, \alpha^+]$, the function $\psi(u) = -\int_{\alpha^-}^u \varphi(y)dy$. Let the function $\hat{\psi}(u)$, $u \in [\alpha^-, \alpha^+]$, be upper bound of the convex set

$$\{(u, v): v \leq \psi(u), u \in [\alpha^-, \alpha^+]\}$$

By Assumption 1, the set $s = \{u \in [\alpha^-, \alpha^+]: \psi(u) < \hat{\psi}(u)\}$ is the finite union of intervals, $s = (\alpha^-, \beta_0) \cup (\alpha_1, \beta_1) \cup \dots \cup (\alpha_L, \alpha^+)$, where $\alpha^- = \alpha_0 \leq \beta_0 \leq \alpha_1 < \beta_1 \dots \leq \alpha_L \leq \beta_L = \alpha^+$.

Let us define the function $\hat{f}(x, t)$ by

$$\hat{f}(x, t) = \begin{cases} \alpha^-, & \text{if } x < \varphi(\alpha^-) \cdot t \\ (\hat{\psi}')^{(-1)}(x/t), & \text{if } \varphi(\alpha^-) \cdot t \leq x \leq \varphi(\alpha^+) \cdot t \\ \alpha^+, & \text{if } x > \varphi(\alpha^+) \cdot t \end{cases}$$

where in the case $\hat{\psi}'(u) \equiv \xi_l$, $u \in (\alpha_l, \beta_l)$, $l = 0, 1, \dots, L$; also, by definition, $(\hat{\psi}')^{(-1)}(\xi_l) = [\alpha_l, \beta_l]$.

Theorem 1 (Gelfand) *The solution $f(x, t)$ of the problem [1], [7] for the case $\varepsilon = +0$ and initial conditions $f(x, 0) = \alpha^\pm$, if $\pm x > 0$, has the explicit form: $f(x, t) = \hat{f}(x, t)$.*

The analogous statement is valid also for the problem [2], [8] if, in the construction above, one takes

$$\Psi(u) = \int_0^u \frac{dy}{\varphi(y)}$$

instead of $\psi(u)$, $u \in [\alpha^-, \alpha^+]$.

The Gelfand problem for [1], [3] and [1], [7] with monotonic $\varphi(f)$ was solved by Iljin and Oleinik (1960). In the case $\alpha^- = \alpha^+$, the solution of this problem follows from an earlier work of Lax (1957). For the case of linear $\varphi(f)$, the solution of this problem follows from an earlier work of Hopf (1950).

For semidiscrete initial problems [2], [4] and [2], [8], the analog of the asymptotic results of Hopf and Iljin–Oleinik have been obtained and applied by Henkin and Polterovich (1991).

The case of increasing $\varphi(f)$ has been studied in detail. In this case, for both initial problems [1], [3] and [2], [4], there is uniform convergence of solutions $f(x, t)$ and $F(x, t)$ to the so-called rarefaction profile

$$g(x/t) = \begin{cases} \alpha^\pm, & \pm x > \varphi(\alpha^\pm)t \\ \varphi^{(-1)}(x/t), & x \in [\varphi(\alpha^-) \cdot t, \varphi(\alpha^+) \cdot t] \end{cases}$$

$t \rightarrow \infty$ (see Iljin and Oleinik (1960) and Henkin and Polterovich (1991)). More precise result in this case about convergence to the so-called

N-wave has been obtained by Dafermos (1977) and Liu (1978).

For the case of a general $\varphi(f)$, in particular, for the case of nonincreasing $\varphi(f)$, we need the notion of shock profile. Following Serre (1999), three definitions can be introduced.

Definition The initial problem [1], [3] (correspondingly, [2], [4]) admits (α^-, α^+) -shock profile ($\alpha^- < \alpha^+$) if there exists a traveling-wave solution of this equation, that is, of the form $f = \tilde{f}(x - ct)$ (correspondingly, $F = \tilde{F}(x - Ct)$), such that $\tilde{f}(x) \rightarrow \alpha^\pm$ when $x \rightarrow \pm\infty$ (correspondingly, $\tilde{F}(x) \rightarrow \alpha^\pm$ when $x \rightarrow \pm\infty$).

From the results of Gelfand (1959) and Oleinik (1959), it follows that initial problem [1], [3] admits (α^-, α^+) -shock profile iff

$$\begin{aligned} c &= \frac{1}{\alpha^+ - \alpha^-} \int_{\alpha^-}^{\alpha^+} \varphi(y)dy \\ &< \frac{1}{u - \alpha^-} \int_{\alpha^-}^u \varphi(y)dy, \quad \forall u \in (\alpha^-, \alpha^+) \end{aligned} \quad [9]$$

From the results of Henkin and Polterovich (1991) and Belenky (1990), it follows that initial problem [2], [4] admits (α^-, α^+) -shock profile iff

$$\begin{aligned} \frac{1}{C} &= \frac{1}{\alpha^+ - \alpha^-} \int_{\alpha^-}^{\alpha^+} \frac{dy}{\varphi(y)} \\ &> \frac{1}{u - u^-} \int_{\alpha^-}^u \frac{dy}{\varphi(y)}, \quad \forall u \in (\alpha^-, \alpha^+) \end{aligned} \quad [10]$$

In the case $\varepsilon = +0$, the equality in [9] and [10] is called the Rankine–Hugoniot condition, the inequality in [9] and [10] is called the entropy condition (or the Gelfand–Oleinik condition).

Definition For initial problem [1], [3] (correspondingly, [2], [4]) admitting (α^-, α^+) -shock profile and for $\varepsilon = +0$, we will call by shock waves the weak solutions of [1], [3] (correspondingly, [2], [5], [4]) of the form

$$\begin{aligned} \tilde{f}^{\alpha^\pm}(x - ct) &= \alpha^\pm, & \text{if } \pm x \geq \pm ct \\ \tilde{F}^{\alpha^\pm}(x - Ct) &= \alpha^\pm, & \text{if } \pm x \geq \pm Ct \end{aligned}$$

where c, C satisfy Rankine–Hugoniot and entropy conditions [9], [10].

Definition The (α^-, α^+) -shock profile for [1] (correspondingly, for [2]) is called strict if in addition to [9], [10] we have the Lax (1954) condition:

$$\varphi(\alpha^+) < c < \varphi(\alpha^-) \quad [11]$$

and correspondingly

$$\varphi(\alpha^+) < C < \varphi(\alpha^-) \quad [12]$$

The (α^-, α^+) -shock profile for [1] or [2] is called semicharacteristic if one of the inequalities in [11] or [12] is strict and the other is an equality. This profile is called characteristic if both inequalities in [11] or [12] are equalities.

One can check (Iljin and Oleinik 1960, Henkin and Polterovich 1991) that if in addition to Assumption 1 the function φ on $[\alpha^-, \alpha^+]$ is nonconstant and nonincreasing then eqn [1] (correspondingly, [2]) admits a strict (α^-, α^+) -shock profile.

The main result of Iljin–Oleinik (1960) for eqn [1] and analogous statement of Henkin and Polterovich (1991) for eqn [2] can be presented as follows.

Theorem 2

- (i) Let the initial problem [1], [3] admit a strict (α^-, α^+) -shock profile \tilde{f} . Let $f(x, t), x \in \mathbb{R}, t \in \mathbb{R}_+$, be a solution of [1], [3]. Then there exists $d_0 \in \mathbb{R}$

$$\sup_{x \in \mathbb{R}} |f(x, t) - \tilde{f}(x - ct - d_0)| \rightarrow 0, \quad t \rightarrow \infty \quad [13]$$

The value of d_0 is determined uniquely by relation

$$\int_{-\infty}^{\infty} \{f(x, 0) - \tilde{f}(x - d_0)\} dx = 0$$

- (ii) Let the initial problem [2], [4] admit a strict (α^-, α^+) -shock profile \tilde{F} . Let $F(x, t), x \in \mathbb{R}, t \in \mathbb{R}_+$ be a solution of [2], [4]. Then there exists continuous function $D_0(\theta), \theta \in [0, 1]$, such that

$$\sup_{x \in \mathbb{R}} |F(x, t) - \tilde{F}(x - Ct - D_0(\{x/\varepsilon\}))| \rightarrow 0, \quad t \rightarrow \infty \quad [14]$$

The function $D_0(\theta), \theta \in [0, 1]$, is determined uniquely from relation

$$\sum_{k=-\infty}^{\infty} \{\Phi(F(n, 0)) - \Phi(\tilde{F}(n - D_0))\} = 0$$

where

$$\Phi(F) = \int_F^A \frac{dy}{\varphi(y)}, \quad F < A, \tilde{F} < A$$

- (iii) If in conditions (i) and (ii), we take $\varepsilon = +0$ then there exist d_0, D_0 such that $\forall \delta > 0$, we have

$$\begin{aligned} & \sup_{x \geq ct + d_0 + \delta} |\alpha^+ - f(x, t)| \\ & + \sup_{x \leq ct + d_0 - \delta} |\alpha^- - f(x, t)| \rightarrow 0, \quad t \rightarrow \infty \\ & \sup_{x \geq Ct + D_0 + \delta} |\alpha^+ - F(x, t)| \\ & + \sup_{x \leq Ct + D_0 - \delta} |\alpha^- - F(x, t)| \rightarrow 0, \quad t \rightarrow \infty \end{aligned} \quad [15]$$

The values of d_0 and D_0 are determined by

$$\begin{aligned} & \int_{-\infty}^{d_0} (f(x, 0) - \alpha^-) dx + \int_{d_0}^{\infty} (f(x, 0) - \alpha^+) dx = 0 \\ & \int_{-\infty}^{D_0} (F(x, 0) - \alpha^-) dx + \int_{D_0}^{\infty} (F(x, 0) - \alpha^+) dx = 0 \end{aligned}$$

Remarks

- (i) The statements of Theorem 2 give a positive answer to Gelfand’s question for the case of initial problem [1], [3] and [2], [4], admitting strict shock profiles.
- (ii) For linear $\varphi(f) = a + bf, a > 0, a + b\alpha^+ > 0, b < 0$, the traveling waves \tilde{f}, \tilde{F} for [1], [3] and [2], [4] can be found explicitly:

$$\begin{aligned} \tilde{f} &= \alpha^- + \frac{\alpha^+ - \alpha^-}{1 + \exp\{-p(x - ct)\}} \\ c &= a + \frac{b}{2}(\alpha^+ + \alpha^-), \quad p = \frac{(\alpha^- - \alpha^+)b}{2\varepsilon} \\ \tilde{F} &= \alpha^- + \frac{\alpha^+ - \alpha^-}{1 + \exp\{-P(x - Ct)\}} \\ C &= b \left/ \ln \frac{a + b\alpha^+}{a + b\alpha^-} \right., \quad P = \frac{\gamma}{\varepsilon} \ln \frac{a + b\alpha^-}{a + b\alpha^+} \end{aligned}$$

where

$$b\gamma = (\alpha + b\alpha^-) \left(1 - \left(\frac{a + b\alpha^+}{a + b\alpha^-} \right)^\gamma \right)$$

- (iii) For initial problems [1], [7] and [2], [8], $\alpha^+ > \alpha^-$, the asymptotic convergence statements [13]–[15] admit the precise asymptotic estimates (see Iljin and Oleinik (1960) for [1], [7]:

$$\sup_{x \in \mathbb{R}} |f(x, t) - \tilde{f}(x - ct - d_0)| = O(e^{-\gamma t}) \quad [16]$$

$\gamma > 0, \varepsilon > 0$

$$\sup_{x \in \mathbb{R}} |F(x, t) - \tilde{F}(x - Ct - D_0(\{x/\varepsilon\}))| = O(e^{-\gamma t}) \quad [17]$$

$\gamma > 0, \varepsilon > 0$

$$\begin{aligned} f(x, t) &= \alpha^\pm \quad \text{for } \pm x > \pm(ct + d_0) \\ t &\geq t_0, \varepsilon = +0 \\ F(x, t) &= \alpha^\pm \quad \text{for } \pm x > \pm(Ct + D_0) \\ t &\geq t_0, \varepsilon = +0 \end{aligned} \quad [18]$$

Theorem 2(i) is proved basing on the following idea. Let f satisfy the initial problem [1], [3] and let

$\tilde{f}(x - ct + d_0)$ be (α^-, α^+) -shock profile for [1], satisfying condition [13]. Put

$$\delta(x, t) = \int_{-\infty}^x \{f(y, t) - \tilde{f}(y - ct - d_0)\} dy$$

The function $\delta(x, t)$ satisfies the nonlinear parabolic equation

$$\frac{\partial \delta}{\partial t} + \varphi(\kappa \tilde{f} + (1 - \kappa)f) \frac{\partial \delta}{\partial x} = \varepsilon \frac{\partial^2 \delta}{\partial x^2}$$

where $\kappa(x, t)$ is some smooth function of (x, t) with values in $[0, 1]$.

Besides, by conservation law of Theorem 0(i), we have $\delta(x, t) \rightarrow 0, x \rightarrow \pm\infty, \forall t \geq 0$.

Estimates basing on maximum principle and appropriate comparison statements give that $\delta(x, t) \Rightarrow 0, x \in \mathbb{R}, t \rightarrow \infty$. It implies that

$$f(x, t) - \tilde{f}(x - ct - d_0) \Rightarrow 0, \quad x \in \mathbb{R}, t \rightarrow \infty$$

Theorem 2(ii) is proved in a similar way. Let $F(n, t)$ satisfy the initial problem [2], [4] with $x = n \in \mathbb{Z}, \varepsilon = 1, \theta = \{x\} = 0$, and let $\tilde{F}(n - Ct - D_0)$ be (α^-, α^+) -shock profile for [2], satisfying condition [14]. Put

$$\Delta(n, t) = \sum_{-\infty}^n \{\Phi(F(n, t)) - \Phi(\tilde{F}(n - Ct - D_0))\}$$

Then function $\Delta(n, t)$ satisfies the semidiscrete parabolic equation

$$\begin{aligned} \frac{d\Delta(n, t)}{dt} &= \varphi(\Phi^{(-1)}(\kappa\Phi(F) \\ &+ (1 - \kappa)\Phi(\tilde{F}))) (\Delta(n - 1, t) - \Delta(n, t)) \end{aligned}$$

where $\kappa(n, t)$ is some function with values in $[0, 1]$.

Besides, by conservation law of Theorem 0(ii), we have

$$\Delta(n, t) \rightarrow 0, \quad n \rightarrow \pm\infty, \forall t \geq 0$$

Estimates, basing on generalized maximum principle and comparison statements, give that $\Delta(n, t) \Rightarrow 0, n \in \mathbb{Z}, t \rightarrow \infty$. It implies that

$$F(n, t) - \tilde{F}(n - Ct + D_0) \Rightarrow 0, \quad n \in \mathbb{Z}, t \rightarrow \infty$$

Remark For the cases of nonstrict shock profiles (characteristic or semicharacteristic) the statements of Theorem 2 are not valid. The reason is that, under initial conditions [3], [4] for any d_0 and D_0 , we have

$$\int_{-\infty}^{\infty} \{f(x, 0) - \tilde{f}(x - d_0)\} dx = \infty$$

and, correspondingly,

$$\sum_{-\infty}^{\infty} \{\Phi(\tilde{F}(k\varepsilon + \theta\varepsilon - D_0)) - \Phi(F(k\varepsilon + \theta\varepsilon, 0))\} = \infty$$

So, the crucial argument, related to conservation law, does not hold.

One can extend the important Theorems 2(i), 2(ii) for the case of nonstrict shock profiles in two different ways: by changing conditions of these theorems or by changing conclusions of these theorems.

The first method (started by Mei, Matsumura, and Nishihara in 1994) was completed by the following L^1 -asymptotic stability result (Serre 2004).

Theorem 3 (Freistühler–Serre). *Let eqns [1], [2] admit (α^-, α^+) -shock profiles and \tilde{f}, \tilde{F} – the corresponding train-wave solutions of [1], [2]. Let $f(x, t), F(n, t), x \in \mathbb{R}, n \in \mathbb{Z}, t \in \mathbb{R}_+$ be solutions of eqns [1], [2] with such initial conditions that*

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x, 0) - \tilde{f}(x)| dx &< \infty \\ \sum_{-\infty}^{\infty} |F(n, 0) - \tilde{F}(n)| &< \infty \end{aligned}$$

Then

$$\int_{-\infty}^{\infty} |f(x, t) - \tilde{f}(x - ct - d_0)| dx \rightarrow 0$$

and, correspondingly,

$$\sum_{-\infty}^{\infty} |F(n, t) - \tilde{F}(n - Ct - D_0)| \rightarrow 0, \quad t \rightarrow \infty$$

where constants d_0 and D_0 are calculated from the same relations as in Theorem 2.

Remark For the inviscid case $\varepsilon = +0$, the statement of Theorem 3 is still valid for equations admitting strict shock profiles, but generally is not valid for equations admitting only nonstrict shock profiles (see Serre (2004)).

The second method permits, keeping initial conditions [3], [4], to localize the positions of viscous shock waves for generalized Burgers equations (see the next section).

Asymptotic Behavior of Solutions of Generalized Burgers Equations

The main current interest and the main difficulty in the study of Gelfand’s problem for generalized Burgers equations consist in the following question formulated explicitly for initial problem [1], [3] by Liu *et al.* (1998): “In the Cauchy problem there is

the question of determining the location of viscous shock waves". A similar question and related conjecture were formulated by Henkin and Potterovich (1999) for the initial problem [2], [4].

For solving this problem, it is important to solve it first for the Burgers type equations admitting nonstrict shock profiles.

Theorem 4 (Henkin–Shananin–Tumanov).

(i) Let the initial problem [1], [3] admit the nonstrict (α^-, α^+) -shock profile [9] and $\tilde{f}(x - ct)$ be a corresponding traveling-wave solution. Let

$$\begin{aligned} \varphi'(\alpha^-) \neq 0, & \quad \text{if } \varphi(\alpha^-) = c \\ \varphi'(\alpha^+) \neq 0, & \quad \text{if } \varphi(\alpha^+) = c \end{aligned}$$

Let $f(x, t)$ be a solution of [1], [3]. Then there exist constants γ_0 and d_0 such that

$$\sup_{x \in \mathbb{R}} |f(x, t) - \tilde{f}(x - ct - \epsilon\gamma_0 \ln t - d_0)| \rightarrow 0, \quad t \rightarrow \infty$$

where

$$\begin{aligned} & (\alpha^+ - \alpha^-) \cdot \gamma_0 \\ & = \begin{cases} -1/\varphi'(\alpha^+), & \text{if } \varphi(\alpha^-) > c = \varphi(\alpha^+) \\ 1/\varphi'(\alpha^-), & \text{if } \varphi(\alpha^-) = c > \varphi(\alpha^+) \\ 1/\varphi'(\alpha^-) - 1/\varphi'(\alpha^+), & \text{if } \varphi(\alpha^-) = c = \varphi(\alpha^+) \end{cases} \end{aligned}$$

(ii) Let the initial problem [2], [4] with $\epsilon = 1$ admit the nonstrict (α^-, α^+) -shock profile [10] and $\tilde{F}(n - Ct)$ be a corresponding traveling-wave solution. Let

$$\begin{aligned} \varphi'(\alpha^-) \neq 0, & \quad \text{if } \varphi(\alpha^-) = C \\ \varphi'(\alpha^+) \neq 0, & \quad \text{if } \varphi(\alpha^+) = C \end{aligned}$$

Let $F(n, t)$ be a solution of [2], [4]. Let

$$\Delta F(n, 0) \stackrel{\text{def}}{=} F(n, 0) - F(n - 1, 0) \geq 0$$

Then there exist constants Γ_0 and D_0 such that

$$\sup_{n \in \mathbb{Z}} |F(n, t) - \tilde{F}(n - Ct - \Gamma_0 \ln t - D_0)| \rightarrow 0, \quad t \rightarrow \infty$$

where

$$\begin{aligned} & (\alpha^+ - \alpha^-) \cdot \Gamma_0 \\ & = \begin{cases} -C/(2\varphi'(\alpha^+)), & \text{if } \varphi(\alpha^-) > C = \varphi(\alpha^+) \\ C/(2\varphi'(\alpha^-)), & \text{if } \varphi(\alpha^-) = C > \varphi(\alpha^+) \\ (C/2)[-1/\varphi'(\alpha^+) \\ +1/\varphi'(\alpha^-)], & \text{if } \varphi(\alpha^-) = C = \varphi(\alpha^+) \end{cases} \end{aligned}$$

Remarks

(i) One could think that nonstrict shock profiles as in Theorem 4 can appear only in exceptional cases. But Proposition 2 and Theorem 5 below

show, on the contrary, that characteristic shock profiles and, as a consequence, the behavior of initial problems [1], [3] and [2], [4] as in Theorem 4 are rather a rule than an exception.

(ii) The statement of Theorem 4(i) (and also of Theorem 5(i)) below) disprove the Gelfand hope that the main term of asymptotic ($t \rightarrow \infty$) of $f(x, t)$, satisfying [1], [7], coincides with the solution of [1], [7] for $\epsilon = +0$ with the same initial condition. Indeed, in conditions of Theorem 4, we have $\varphi(\alpha^-) = c$ or $\varphi(\alpha^+) = c$, but $\varphi'(\alpha^-) \neq \varphi'(\alpha^+)$; then for any $\epsilon > 0$ the traveling wave $\tilde{f}(x - ct - \epsilon\gamma_0 \ln t - d_0)$ for [1], [3], concentrated near the point $x_\epsilon(t) = ct + \epsilon\gamma_0 \ln t + d_0$, moves away ($t \rightarrow \infty$) from the shockwave for [1], [7] for $\epsilon = +0$, concentrated near the point $x_0(t) = ct + o(\ln t)$, where $o(\ln t)/\ln t \rightarrow 0, t \rightarrow \infty$.

(iii) Theorem 4 (and also Theorem 5 below) also illustrate another interesting phenomenon: for the case $\varphi'(\alpha^-) \neq \varphi'(\alpha^+)$, one has asymptotic convergence of the solution of [1], [3] (correspondingly of [2], [4]) to the traveling wave $\tilde{f}(x - ct - \epsilon\gamma_0 \ln t - d_0)$ (correspondingly $\tilde{F}(x - Ct - \epsilon\Gamma_0 \ln t - D_0)$), which does not satisfy eqn [1] or correspondingly eqn [2]. Such a phenomenon was first discovered by Liu and Yu (1997) in the special boundary-value problem for the classical Burgers equations, if $u(x, t)$ satisfies the following conditions:

$$\begin{aligned} \text{if } u_t + u \cdot u_x = u_{xx}, \quad u(0, t) = 1, \quad u(\infty, t) = -1, \\ u(x, 0) = -th \frac{x}{2}, \text{ then} \end{aligned}$$

$$|u(x, t) + th \frac{1}{2}(x - \ln(1 + t))| \rightarrow 0, \quad t \rightarrow \infty, \quad x \geq 0$$

Theorem 4 is proved in basing on the following idea. Let $f(x, t)$ satisfy [1], [3] and $F(n, t)$ satisfy [2], [4]. Let $\tilde{f}(x - ct)$ be the traveling wave for [1], [3] and $\tilde{F}(n - Ct)$ be the traveling wave for [2], [4]. Suppose that $\varphi(\alpha^-) > c = C = \varphi(\alpha^+)$. Let $d_A(t)$ and $D_A(t)$, $A > 0$ be functions such that

$$\int_{ct - A\sqrt{t}}^{ct + A\sqrt{t}} \{f(x, t) - \tilde{f}(x - ct - d_A(t))\} dx = 0 \quad [19]$$

and, correspondingly,

$$\begin{aligned} & \sum_{k=[Ct - A\sqrt{t}] }^{[Ct + A\sqrt{t}]} \{ \Phi(F(k, t)) - \Phi(\tilde{F}(k - Ct - D_A(t))) \} \\ & + (Ct + A\sqrt{t} - [Ct + A\sqrt{t}])(\Phi(F(Ct + A\sqrt{t}) + 1, t)) \\ & - \Phi(\tilde{F}([Ct + A\sqrt{t}] + 1 - Ct + D_A(t))) = 0 \end{aligned}$$

[20]

The relations [9], [20] can be called “localized conservation law.” The proof contains two difficult parts.

The first part consists in proving that for $A > 2\sqrt{C}$ (correspondingly, $A > 2\sqrt{C}$) the following asymptotics are valid:

$$d_A(t) = \frac{\epsilon \cdot \ln t}{(\alpha^- - \alpha^+) \varphi'(\alpha^+)} + d^0 + o(1), \quad t \rightarrow \infty$$

$$D_A(t) = \frac{C \ln t}{2(\alpha^- - \alpha^+) \varphi'(\alpha^+)} + D^0 + o(1), \quad t \rightarrow \infty$$
[21]

where d^0, D^0 are independent of A .

The second part gives the following convergence statements:

$$\sup_{x \in [ct - A\sqrt{t}, ct + A\sqrt{t}]} \left| \int_{ct - A\sqrt{t}}^x \{f(y, t) - \tilde{f}(y - ct - d_A(t))\} dy \right| \rightarrow 0, \quad t \rightarrow \infty$$

$$\sup_{x \in [Ct - A\sqrt{t}, Ct + A\sqrt{t}]} \left| \sum_{k=[Ct - A\sqrt{t}]}^n \{\Phi(F(k, t)) - \Phi(\tilde{F}(k - Ct - D_A(t)))\} \right| \rightarrow 0, \quad t \rightarrow \infty$$

The precise *a priori* estimates of local solutions of [1], [2] play an important role in the proof. An example of such an estimate, also useful for further results, is given below.

Proposition 1 *Let, in eqn [2], $C = \varphi(0) > 0, \epsilon = 1, 0 \leq \varphi'(0) < \gamma_0, \bar{x} \stackrel{\text{def}}{=} (x - Ct)/\sqrt{Ct}$. Let the function $F(x, t)$, defined in the domain $\Omega_0 = \{(x, t): a_1 < \bar{x} < a_2\}, a_2 > 0$, satisfy eqn [2],*

$$\Delta F(x, t) \stackrel{\text{def}}{=} F(x, t) - F(x - 1, t) \geq 0$$

$$|F(x, t)| \leq \frac{\Gamma}{\sqrt{Ct}}, \quad (x, t) \in \Omega_0, \quad t \geq t_0$$

Then

$$\Delta F(x, t) \leq \frac{B \cdot \Gamma}{Ct}, \quad (x, t) \in \Omega_0, \quad t \geq t_0$$

where

$$B = B_0 \left[a_2 + \left(\frac{1}{d} + \frac{\gamma_0 \Gamma}{C} \right) (1 + \ln(1 + a_2)) \right]$$

$$d = \min(\bar{x} - a_1, a_2 - \bar{x})$$

and B_0 is an absolute constant.

It is interesting to compare *a priori* estimate of Proposition 1 with some similar (but less precise) estimates in the theory of classical quasilinear parabolic equations (Ladyzhenskaya *et al.* 1968).

We will formulate now the general conjecture concerning asymptotic behavior of solutions of

initial problems [1], [3] and [2], [4] and some partial results which confirm this conjecture. To simplify formulation we admit the following.

Assumption 2 Let $\hat{\psi}(u)$ and $\hat{\Psi}(u)$ be upper bounds of the convex hulls for the graphs of

$$\psi(u) = - \int_{\alpha^-}^u \varphi(y) dy$$

and

$$\Psi(u) = \int_{\alpha^-}^u \frac{dy}{\varphi(y)}$$

respectively, with $u \in [\alpha^-, \alpha^+]$. We suppose that

$$s = \{u \in [\alpha^-, \alpha^+]: \psi(u) < \hat{\psi}(u)\}$$

$$= (\alpha^-, \beta_0) \cup (\alpha_1, \beta_1) \cup \dots \cup (\alpha_L, \alpha^+)$$

where

$$\alpha^- = \alpha_0 < \beta_0 < \alpha_1 < \beta_1 < \dots < \alpha_L < \beta_L = \alpha^+$$

or, correspondingly,

$$S = \{u \in [\alpha^-, \alpha^+]: \Psi(u) < \hat{\Psi}(u)\}$$

$$= (\alpha^-, b_0) \cup (a_1, b_1) \cup \dots \cup (a_M, \alpha^+)$$

where

$$\alpha^- = a_0 < b_0 < a_1 < b_1 < \dots < a_M < b_M = \alpha^+$$

In addition, we suppose that $\varphi'(\alpha_l) \neq 0, \varphi'(\beta_l) \neq 0, l=0, 1, \dots, L$ or, correspondingly, $\varphi'(a_m) \neq 0, \varphi'(b_m) \neq 0, m=0, 1, \dots, M$.

Proposition 2 (Weinberger 1990, Henkin and Polterovich 1999). *Under Assumptions 1, 2, one has:*

(i) *If $u \in [\alpha^-, \alpha^+] \setminus s$ and, correspondingly, $u \in [\alpha^-, \alpha^+] \setminus S$, then following functions are well defined:*

$$g_l \left(\frac{x}{t} \right) = \begin{cases} \beta_l, & \text{if } x < \varphi(\beta_l) \cdot t \\ \varphi^{(-1)}(x/t), & \text{if } \varphi(\beta_l) \cdot t \leq x \\ & \leq \varphi(\alpha_{l+1}) \cdot t \\ \alpha_{l+1}, & \text{if } x > \varphi(\alpha_{l+1}) \cdot t, \\ & l = 0, 1, \dots, L \end{cases}$$

and, correspondingly,

$$G_l \left(\frac{x}{t} \right) = \begin{cases} b_m, & \text{if } x < \varphi(b_m) \cdot t \\ \varphi^{(-1)}(x/t), & \text{if } \varphi(b_m) \cdot t \leq x \\ & \leq \varphi(a_{m+1}) \cdot t \\ a_m, & \text{if } x > \varphi(a_{m+1}) \cdot t, \\ & m = 0, 1, \dots, M \end{cases}$$

(ii) *For any interval $(\alpha_l, \beta_l) \subset s$ and, correspondingly, $(a_m, b_m) \subset S$ there exist traveling waves $\tilde{f}_l(x - ct)$ for [1] with overfall (α_l, β_l) and,*

correspondingly, $\tilde{F}_m(x - C_m t)$ for [2] with overfall (a_m, b_m) , where

$$c_l = \frac{1}{\beta_l - \alpha_l} \int_{\alpha_l}^{\beta_l} \varphi(y) dy$$

$$c_l = \varphi(\beta_l), \quad l = 0, \dots, L - 1$$

$$c_l = \varphi(\alpha_l), \quad l = 1, \dots, L$$

and, correspondingly,

$$C_m^{-1} = \frac{1}{b_m - a_m} \int_{a_m}^{b_m} \frac{dy}{\varphi(y)}$$

$$C_m = \varphi(b_m), \quad m = 0, \dots, M - 1$$

$$C_m = \varphi(a_m), \quad m = 1, \dots, M$$

Conjecture (Henkin and Polterovich 1994, 1999, Henkin and Shanenin 2004). Let

$$\tilde{f}(x, t, \gamma_0, \dots, \gamma_L)$$

$$= \sum_{l=0}^L \tilde{f}_l(x - c_l t - \varepsilon \gamma_l(t)) + \sum_{l=0}^{L-1} g_l\left(\frac{x}{t}\right) - \sum_{l=0}^{L-1} \beta_l$$

$$- \sum_{l=1}^L \alpha_l, \quad L \geq 1$$

$$\tilde{F}(n\varepsilon, t, \Gamma_0, \dots, \Gamma_M)$$

$$= \sum_{m=0}^M \tilde{F}_m(n\varepsilon - C_m t - \varepsilon \Gamma_m(t)) + \sum_{m=0}^{M-1} G_m\left(\frac{n\varepsilon}{t}\right)$$

$$- \sum_{m=0}^{M-1} b_m - \sum_{m=1}^M a_m, \quad M \geq 1$$

Then under Assumptions 1, 2, the following statements are valid:

- (i) For any solution $f(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$, of initial problem [1], [3], there exist shift-functions $\gamma_l(t)$:

$$\gamma_l^- \ln t + O(1) \leq \gamma_l(t) \leq \gamma_l^+ \ln t + O(1)$$

$$0 \leq \gamma_l^- \leq \gamma_l^+ < \infty, \quad l = 0, 1, \dots, L$$

such that

$$\sup_{x \in \mathbb{R}} |f(x, t) - \tilde{f}(x, t, \gamma_0, \gamma_1, \dots, \gamma_L)| \rightarrow 0,$$

$$t \rightarrow \infty$$

- (ii) Moreover, in (i) one can take

$$\gamma_l^- = \gamma_l^+$$

$$= \frac{\varepsilon}{(\beta_l - \alpha_l)}$$

$$\times \begin{cases} -\frac{1}{\varphi'(\beta_l)}, & \text{if } l = 0 < L, \varphi(\alpha_0) \neq \varphi(\beta_0) \\ \frac{1}{\varphi'(\alpha_l)} - \frac{1}{\varphi'(\beta_l)}, & \text{if } 0 < l < L \\ \frac{1}{\varphi'(\alpha_l)}, & \text{if } l = L > 0, \varphi(\alpha_L) \neq \varphi(\beta_L) \end{cases}$$

- (iii) For any solution $F(n\varepsilon, t)$, $n \in \mathbb{Z}$, $t \in \mathbb{R}_+$, of initial problem [2], [4], there exist shift-functions $\Gamma_m(t)$:

$$\Gamma_m^- \ln t + O(1) \leq \Gamma_m(t) \leq \Gamma_m^+ \ln t + O(1)$$

$$0 \leq \Gamma_m^- \leq \Gamma_m^+ < \infty, \quad l = 0, 1, \dots, L$$

such that

$$\sup_{n \in \mathbb{Z}} |F(n\varepsilon, t) - \tilde{F}(n\varepsilon, t, \Gamma_0, \Gamma_1, \dots, \Gamma_M)| \rightarrow 0,$$

$$t \rightarrow \infty$$

- (iv) Moreover, in (iii) one can take

$$\Gamma_m^- = \Gamma_m^+$$

$$= \frac{C_m}{(b_m - a_m)}$$

$$\times \begin{cases} -\frac{1}{\varphi'(b_m)}, & \text{if } m = 0 < M, \varphi(a_0) \neq \varphi(b_0) \\ \frac{1}{\varphi'(a_m)} - \frac{1}{\varphi'(b_m)}, & \text{if } 0 < m < M \\ \frac{1}{\varphi'(a_m)}, & \text{if } m = M > 0, \varphi(a_M) \neq \varphi(b_M) \end{cases}$$

The main result confirming formulated conjectures is the following.

Theorem 5 (Henkin and Shanenin). *Conjecture (i) for $L = 1$ and corresponding conjecture (iii) for $M = 1$ are true, that is, for solution of initial problem [1], [3] there exist shift functions $\gamma_l(t) = O(\ln t)$ such that for $t \rightarrow \infty$ we have*

$$f(x, t) \mapsto \begin{cases} \tilde{f}_0(x - c_0 t - \varepsilon \gamma_0(t)), & \text{if } x \leq c_0 t \\ \varphi^{(-1)}(x/t), & \text{if } c_0 t \leq x \leq c_1 t \\ \tilde{f}_1(x - c_1 t - \varepsilon \gamma_1(t)), & \text{if } x \geq c_1 t \end{cases}$$

and for solution of initial problem [2], [4] there exist shift functions $\Gamma_m(t) = O(\ln t)$ such that for $t \rightarrow \infty$ we have

$$F(n\varepsilon, t) \mapsto \begin{cases} \tilde{F}_0(n\varepsilon - C_0 t - \varepsilon \Gamma_0(t)), & \text{if } n\varepsilon \leq C_0 t \\ \varphi^{(-1)}(n\varepsilon/t), & \text{if } C_0 t \leq n\varepsilon \\ \leq C_1 t \\ \tilde{F}_1(n\varepsilon - C_1 t - \varepsilon \Gamma_1(t)), & \text{if } n\varepsilon \geq C_1 t \end{cases}$$

The proof of Theorem 5 is of the same nature as the proof of Theorem 4.

Remarks

- (i) The proof of stronger Conjectures (ii) and (iv) for $L = 1$ or $M = 1$ are in preparation.
- (ii) The numerical results, Rykova and Spivak (preprint, 2004), confirm conjecture (iii) for $M = 2$.
- (iii) The results of Weinberger (1990) and Henkin and Polterovich (1999) confirm convergence statements of Conjectures (i), (iii) for all L and M , but only on the intervals of rarefaction

profiles: $x \in [\varphi(\beta_l)t, \varphi(\alpha_{l+1})t]$ or, correspondingly, $x \in [\varphi(b_m)t, \varphi(a_{m+1})t]$, $t > 0$.

The problem of finding asymptotics ($t \rightarrow \infty$) of solutions of (viscous) conservation laws has been posed originally not only for generalized Burgers equations but also for systems of conservation laws in one spatial variable (see Gelfand (1959)). In this direction many important results on existence and asymptotic stability of viscous shock profiles (continuous and discrete) have been obtained and applied (see Benzoni-Gavage (2004), Lax (1973), Serre (1999), Zumbrun and Howard (1998) and references therein). The results of type of Theorems 4,5 have not yet been obtained for systems of conservation laws.

It is also very interesting to study asymptotic behavior of scalar (viscous) conservation laws in several spatial variables (continuous or discrete), basing on the asymptotic properties of Burgers type equations. In this direction there have been several important results and problems (see Bauman and Phillips (1986), Henkin and Polterovich (1991), Hoff and Zumbrun (2000), Serre (1999), Weinberger (1990), and references therein).

Further Reading

- Bauman P and Phillips D (1986b) Large-time behavior of solutions to a scalar conservation law in several space dimensions. *Transactions of the American Mathematical Society* 298: 401–419.
- Belenky V (1990) Diagram of growth of a monotonic function and a problem of their reconstruction by the diagram. Preprint, CEMI Academy of Science, Moscow, 1–44 (in Russian).
- Benzoni-Gavage S (2002a) Stability of semi-discrete shock profiles by means of an Evans function in infinite dimension. *J. Dyn. Diff. Equations* 14: 613–674.
- Burgers JM (1940) Application of a model system to illustrate some points of the statistical theory of free turbulence. *Proc. Acad. Sci. Amsterdam* 43: 2–12.
- Dafermos CM (1977) Characteristics in hyperbolic conservation laws. A study of structure and the asymptotic behavior of solutions. In: Knops RJ (ed.) *Nonlinear Analysis and Mechanics: Heriot-Watt Symposium*, vol. 17, pp. 1–58. Research Notes in Mathematics, London: Pitman.
- Gelfand IM (1959) Some problems in the theory of quasilinear equations. *Usp. Mat. Nauk* 14: 87–158 (in Russian). ((1963) *American Mathematical Society Translations* 33).
- Harten A, Hyman JM, and Lax PD (1976) On finite-difference approximations and entropy conditions for shocks. *Communications in Pure and Applied Mathematics* 29: 297–322.
- Henkin GM and Polterovich VM (1991) Schumpeterian dynamics as a nonlinear wave theory. *Journal of Mathematical Economics* 20: 551–590.
- Henkin GM and Polterovich VM (1999) A difference-differential analogue of the Burgers equation and some models of economic development. *Discrete and Continuous Dynamical Systems* 5: 697–728.
- Henkin GM and Shananin AA (2004) Asymptotic behavior of solutions of the Cauchy problem for Burgers type equations. *Journal Mathématiques Pure et Appliquée* 83: 1457–1500.
- Henkin GM, Shananin AA, and Tumanov AE (2005) Estimates for solutions of Burgers type equations and some applications. *Journal Mathématiques Pure et Appliquée* 84: 717–752.
- Hoff D and Zumbrun K (2000) Asymptotic behavior of multi-dimensional viscous shock fronts. *Indiana University Mathematical Journal* 49: 427–474.
- Hopf E (1950) The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Communications in Pure and Applied Mathematics* 3: 201–230.
- Ilijin AM and Oleinik OA (1960) Asymptotic behavior of the solutions of the Cauchy problem for some quasilinear equations for large values of time. *Mat. Sbornik* 51: 191–216 (in Russian).
- Ladyzhenskaya OA, Solonnikov VA, and Ural'ceva NN (1968) *Linear and Quasilinear Equations of Parabolic Type*. Amer. Math.Soc.Transl. Monogr. vol. 23. Providence, RI.
- Landau LD and Lifschitz EM (1968) *Fluid Mechanics*. Elmsford, NY: Pergamon.
- Lax PD (1954) Weak solutions of nonlinear hyperbolic equation and their numerical computation. *Communications in Pure and Applied Mathematics* 7: 159–193.
- Lax PD (1957) Hyperbolic systems of conservation laws, II. *Communications in Pure and Applied Mathematics* 10: 537–566.
- Lax PD, (1973) Hyperbolic systems of conservation laws and the mathematical theory of shock waves. Conference Board of the Mathematical Science, Monograph 11. SIAM.
- Levi D, Ragnisco O, and Brushi M (1983) Continuous and discrete matrix Burgers Hierarchies. *Nuovo Cimento* 74: 33–51.
- Liu T-P (1978) Invariants and asymptotic behavior of solutions of a conservation law. *Proceedings of American Mathematical Society* 71: 227–231.
- Liu T-P, Matsumura A, and Nishihara K (1998) Behaviors of solutions for the Burgers equation with boundary corresponding to rarefaction waves. *SIAM Journal of Mathematical Analysis* 29: 293–308.
- Liu T-P and Yu S-H (1997) Propagation of stationary viscous Burgers shock under the effect of boundary. *Archives for Rational and Mechanical Analysis* 139: 57–92.
- Oleinik OA (1959) Uniqueness and stability of the generalized solution of the Cauchy problem for a quasi-linear equation. *Usp. Mat. Nauk* 14: 165–170. ((1963) *American Mathematical Society Translations* 33).
- Serre D (1999) *Systems of Conservation Laws, I*. Cambridge: Cambridge University Press.
- Serre D (2004) L^1 -stability of nonlinear waves in scalar conservation laws. In: Dafermos C and Feireisl E (eds.) *Handbook of Differential Equations*, pp. 473–553. Elsevier.
- Weinberger HF (1990) Long-time behavior for a regularized scalar conservation law in the absence of genuine nonlinearity. *Annales de L'institut Henri Poincaré (C) Analyse Nonlineaire*.
- Zumbrun K and Howard D (1998) Poinwise semigroup methods and stability of viscous shock waves. *Indiana University Mathematical Journal* 47: 63–185.

Cellular Automata

M Bruschi, Università di Roma “La Sapienza”, Rome, Italy

F Musso, Università “Roma Tre”, Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

What is a Cellular Automaton?

Cellular automata (CAs) were first introduced by J von Neumann in his investigation of “complexity,” following an inspired suggestion by S Ulam. But in the last 50 years they have been investigated and used in a number of fields; widely different terminologies have been used by researchers that now it is difficult even to give a precise general definition of a CA. Thus, some definitions and approximations are in order.

First a broad definition:

1. have a number of cells (boxes);
2. at any (discrete) time step, any cell can present itself in a certain “state” among a finite number of different states;
3. the state of any cell can change (evolve) from a time step to the subsequent time step; and
4. there is a rule (evolution law, EL) which determines this transition.

Note that the number of cells can be finite or infinite; the cells can be arranged on a line, on a surface, in the ordinary three-dimensional (3D) space, or possibly in a hyperspace (in any case, the cells can be numbered); the different states of a cell can be denoted by integer numbers but, in different contexts of application of CAs, different imaginative pictures have also been used (e.g., different colors, dead and living cells, number of balls in a box, etc.); the evolution of a CA proceeds in finite time steps (time is also discrete); the EL, provided that it is effective on any possible configuration of a given CA (computability), is otherwise completely arbitrary (indeed, there are not only deterministic and probabilistic ELs, but also those that “evolve” in time – following a meta-EL, which in turn can be deterministic or probabilistic).

Consider some examples of CAs.

Example 1 (CA1) Consider a linear array of seven boxes (cells; one can number them $c(i)$, $i = 1, 2, \dots, 7$). Each box can be empty or it can contain a ball (so there are just two states for each cell). Given a configuration of this CA at time t , what happens at time $t + 1$ (EL)?

- (i) the state of the first box $c(1)$ never changes;
- (ii) for each other box $c(i)$, $i = 2, 3, \dots, 7$;

- (iia) if the box is empty and the box on its left is empty then put a ball in the box;
- (iib) if there is a ball in the box and also there is a ball in the box on its left then empty the box.

An example of the evolution of such a rather trivial CA is given in [Figure 1](#).

A more precise notation can now be established.

First, let us denote the state of a cell at time t by a “state function,” say S . According to the point (iib) above, the number of possible states is arbitrary but finite: denote this number by the positive integer M ($M > 1$). Then S takes values on a finite field, say $\mathbb{Z}_M = \mathbb{Z}/M\mathbb{Z} = \{0, 1, 2, \dots, M - 1\}$ (in plain words, we have denoted the M states for the CA by the first M non-negative integers). Different cells can be labeled with a progressive number: $c(n)$, $n = n_1, n_1 + 1, \dots, n_2 - 1, n_2$; possibly, in case of an infinite number of cells, one has $n_1 \rightarrow -\infty$ and/or $n_2 \rightarrow +\infty$. In the case of $n_1 = -\infty$, $n_2 = \infty$, one speaks of a unidimensional CA. Of course, the field S depends on n as well as on time (remember that, for a CA, “time” is a discrete variable: $t = 0, 1, 2, \dots$). The field $S(n, t)$ describes completely the CA. If the EL is deterministic, then one can determine (compute) $S(n, t)$ step by step for $t > 0$ from the initial configuration $S(n, 0)$ (initial datum, ID). Consider only static ELs, namely those that do not change in time. A further distinction can be made: there are ELs such that the future state of the generic cell, $S(n, t + 1)$, depends on the whole current configuration of the CA (these are called nonlocal ELs) and there are ELs for which $S(n, t + 1)$ depends only on

	$c(1)$	$c(2)$	$c(3)$	$c(4)$	$c(5)$	$c(6)$	$c(7)$
$t=0$			●	●		●	●
$t=1$		●	●			●	
$t=2$		●			●	●	
$t=3$		●		●	●		
$t=4$		●		●			●
$t=5$		●		●		●	●
$t=6$		●		●		●	
$t=7$		●		●		●	

Figure 1 A seven time-step evolution of CA1 starting from a given ID ($t=0$). Note that a stable configuration has been reached at $t=6$.

the current state of a finite number, say N , of cells (local ELs):

$$\{S(n + k_i, t)\}, i = 1, 2, \dots, N, k_i \in \mathbb{Z} \implies S(n, t + 1) \quad [1]$$

Note that, in principle, the set of cells that determine, according to the EL, the future state of the generic cell n , could depend on n , namely one can have $N = N(n)$, as well $k_i = k_i(n), i = 1, 2, \dots, N(n)$ (see CA2 below). In any case, such a set of cells is called the interaction set (IS). Moreover, the distance from the cell n of the farthest cell in the IS is called the range R (of the interaction): $R = \max(|k_i|)$. If $IS \equiv \{c(n - R), c(n - R + 1), \dots, c(n), \dots, c(n + R - 1), c(n + R)\}$, then this IS is called a neighborhood of range R . It is, moreover, clear that, for unidimensional CA, there exists at least one infinite subset of cells that have the same state. If there is only one such subset, then it is called the vacuum set and the state of its cells is called vacuum state: let V denote the value of this state ($0 \leq V < M, S(n, t) \xrightarrow{n \rightarrow \pm\infty} V$).

Example 2 (CA2) An example of CA with n -dependent IS ($M = 2, R = 3, V = 0$). This is the EL: the cell $c(n)$ changes its state ($0 \rightarrow 1, 1 \rightarrow 0$) iff

- (i) n is even and at least one of the two cells on its left is not in the vacuum state;
- (ii) n is odd and one or three of the three cells on its right are not in the vacuum state.

An example of the evolution of such a CA is given in **Figure 2**.

Usually, only a subclass of ELs is considered for which the phenomenon of vacuum excitation cannot occur. Namely, during the evolution of the CA, an infinite subset of the vacuum set cannot change its state in just one time step. In other words: if the set of cells starting from the first cell and ending with the last one for which

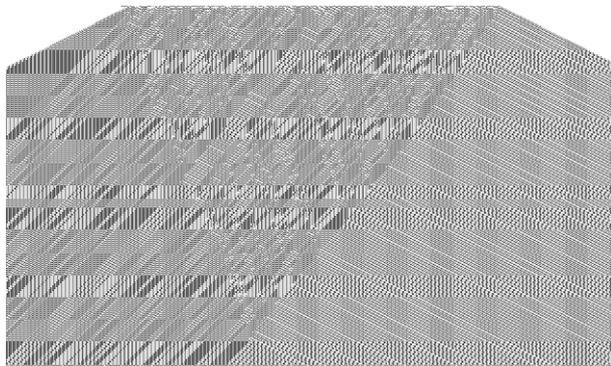


Figure 2 Three hundred and eighty time steps of CA2, starting from a random chosen initial configuration. Note the left-right asymmetry due to the asymmetry of its IS and EL.

$S(n, t) \neq V$ be called population set (PS), then PS is a finite set at each time.

Of course, one can easily devise an EL for which this is not true; nevertheless, the EL itself is still valid (computable), for instance,

Example 3 (CA3) This is an unidimensional CA, namely there are infinite cells on a line ($n \in \mathbb{Z}$). The cells have M states and $V = 0$; the EL reads:

the state of each cell cycles in the set of available states ($0 \rightarrow 1, 1 \rightarrow 2, \dots, M - 2 \rightarrow M - 1, M - 1 \rightarrow 0$)

Note that the range R is zero, there is a vacuum excitation; nevertheless, the EL is effective.

Deterministic, static, and local ELs that do not give rise to vacuum excitation are called normal ELs (NELs).

Since M, N are finite for an NEL, one can give the NEL itself as a table, considering every possible configuration of the IS and specifying the outcome for each configuration (note that there are M^N possible configurations).

Example 4 (CA4) $n \in \mathbb{Z}, M = 2, V = 0, IS \equiv \{c(n), c(n - 1), c(n + 2)\}, N = 3, R = 2$. The EL is:

$$\begin{matrix} S(n, t) & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ S(n - 1, t) & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ S(n + 2, t) & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ S(n, t + 1) & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \end{matrix} \quad [2]$$

An example of the evolution of such a CA is given in **Figure 3**.

However, these NELs can also be given in an alternative representation (more useful in view of the extensions of the concept of CA itself, see below). Namely, an NEL can be given as a discrete-time EL for the state function $S(n, t)$ in the finite field $\mathbb{Z}_M = \{0, 1, 2, \dots, M - 1\}$.

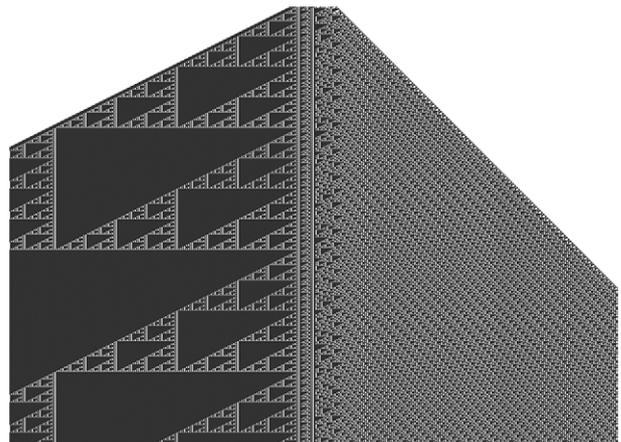


Figure 3 Four hundred and sixty-one time steps of CA4, starting from a random chosen PS of 50 cells.

For example, the NEL above for CA4 can be expressed as follows:

$$S(n, t + 1) \stackrel{2}{=} S(n - 1, t) + S(n, t) + S(n + 2, t) + S(n, t)S(n + 2, t) + S(n - 1, t)S(n, t)S(n + 2, t) \quad [3]$$

Here and in the following, the symbol $\stackrel{M}{=}$ denotes a congruence mod M .

Another example is the following.

Example 5 (CA5) $n \in \mathbb{Z}, M = 3, N = 3, V = 0, R = 1, IS \equiv \{c(n - 1), c(n), c(n + 1)\}$. The NEL is:

$$S(n, t + 1) \stackrel{3}{=} S(n - 1, t) + S(n, t) + S(n + 1, t) + 2S(n - 1, t)S(n + 1, t) \quad [4]$$

An example of the evolution of such a CA is given in **Figure 4**.

Classification of ELs

Considering a CA with given $M > 1, N \geq 1$, the number L of possible deterministic, static ELs is

$$L(M, N) = M^{(M^N)} \quad [5]$$

Of course, this number can be very large for relatively small values of M and N also. Nevertheless, it is a finite positive integer, so that, for given M, N , one could denote every EL by an integer number and investigate the typical behavior of each EL. A considerable reduction of this number is obtained if one limits attention to totalistic ELs, namely to those whose outcome depends only on the global configuration of the IS, often just on

$$\sigma(n, t) = \sum_{i=1}^N S(n + k_i); \quad i = 1, 2, \dots, N, k_i \in \mathbb{Z} \quad [6]$$

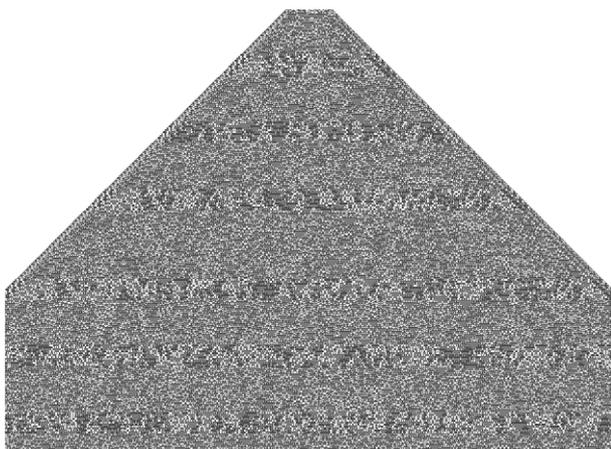


Figure 4 Four hundred and sixty-one time steps of CA5, starting from a random chosen PS of 50 cells.

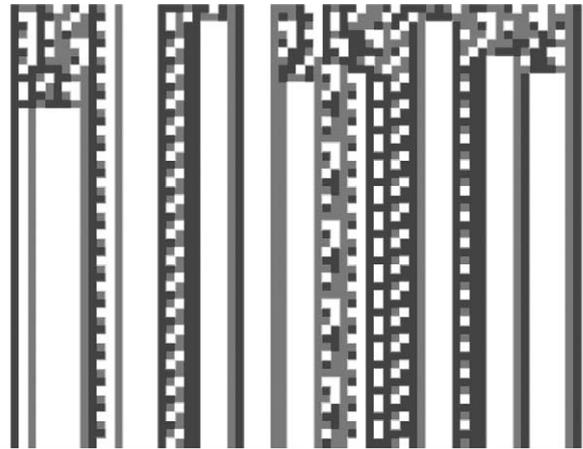


Figure 5 A class-1 CA: every ID rapidly evolves to periodic structures; $M = 3, V = 0, R = 2$, EL: $S(n, t + 1) \stackrel{3}{=} S(n, t) + S(n - 1, t)S(n + 2, t)$.

Deep and extensive computer investigations have been exploited for unidimensional CAs with small values of M, N . Surprisingly enough, it seems that the typical behavior of all these CAs can be (roughly and heuristically) classified in just four classes (Wolfram 2002):

- Class 1 (simple): possibly after a complicated transient, simple patterns emerge.
- Class 2 (fractal): possibly after a transient, overall regular nested structures are obtained.
- Class 3 (chaotic): complicated but seemingly random behavior.
- Class 4 (complex): possibly after a transient, localized structures emerge that interact in complex ways.

Due to the looseness of the above definitions, perhaps a better way to distinguish between classes is to train one’s eye. Consider some examples of CAs for each class: the typical behavior of class-1 CA is shown in **Figures 5 and 6**, of class-2 CA in **Figures 7 and 8**, of class-3 CA in **Figures 4 and 9**, of class-4 CA in **Figures 10 and 11**. Note, however, that often one has “mixed type” CA: for example, CA4 is of class 1 on the right and of class 2 on the left (see **Figure 3**); **Figure 12** exhibits a CA where the typical behaviors of classes 2 and 3 are superimposed.

Extensions

The concept of a CA is so simple that many extensions of the above-sketched definition of a CA can be easily devised. A (nonexhaustive) survey of such extensions follows.

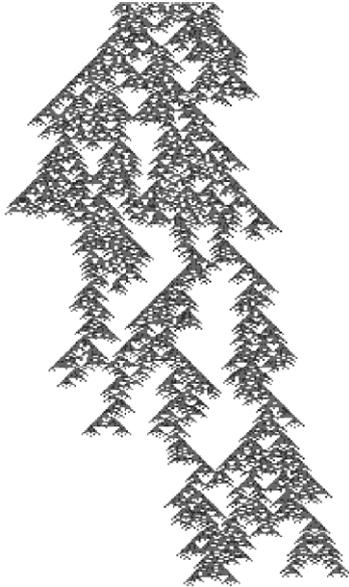


Figure 6 A class-1 CA, a random ID vanishes after 337 time steps, $M=5, V=0, R=2$, EL: $S(n, t+1) \stackrel{\cong}{=} S(n-1, t)S(n-2, t) + S(n+1, t)S(n+2, t) + S(n-1, t)S(n+1, t) + S(n-2, t)S(n+2, t)$.

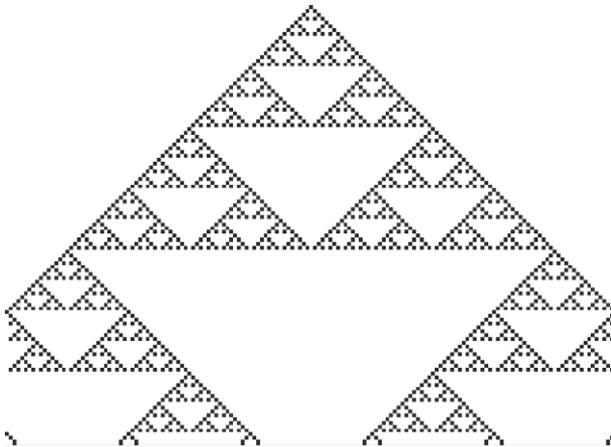


Figure 7 A class-2 CA: Sierpinski triangles appear; $M=2, V=0, R=1$, EL: $S(n, t+1) \stackrel{\cong}{=} S(n-1, t) + S(n+1, t)$.

Vector CA

In this extension, the state function $S(n, t)$ is considered as a “vector,” namely $S(n, t) \equiv (S_1(n, t), S_2(n, t), \dots, S_L(n, t))$, L being a positive integer. Each component $S_l(n, t) (l=1, 2, \dots, L)$ takes values in a finite field, say $\mathbb{Z}_{M_l} = \{0, 1, 2, \dots, M_l - 1\}$, and evolves, according to some EL, interacting with the other components. Of course, one can give separately the time evolution for each component; however, it is also possible to give a global representation of a vector CA, introducing a global

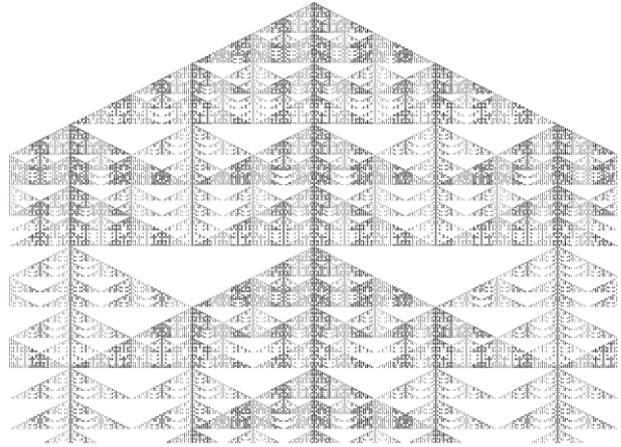


Figure 8 A class-2 CA: a double fractal structure appears; $M=4, V=0, R=2$, EL: $S(n, t+1) \stackrel{\cong}{=} S(n-2, t) + S(n, t) + S(n+2, t)$.

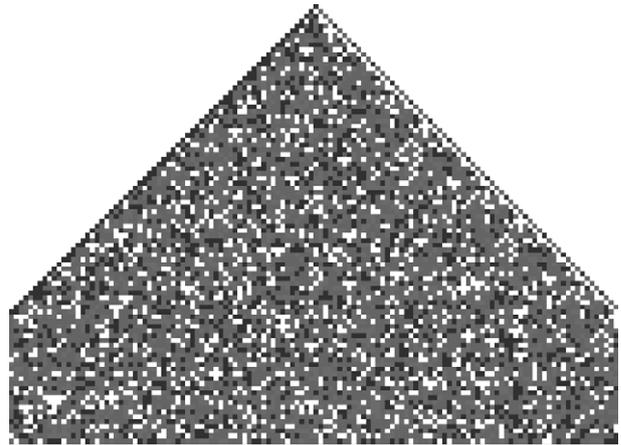


Figure 9 A class-3 CA: $M=5, V=0, R=2$, EL: $S(n, t+1) \stackrel{\cong}{=} 2S(n-1, t) + S(n+1, t) + S(n, t)(S(n+1, t) + S(n+2, t)) + S(n-1, t)S(n+1, t)$.

function $\tilde{S}(n, t)$ that takes values in the finite field $\mathbb{Z}_M, M = M_1 M_2 \dots M_L$; for example,

$$\tilde{S}(n, t) = S_L(n, t) + \sum_{l=1}^{L-1} \left(S_l(n, t) \prod_{k>l}^L M_k \right) \quad [7]$$

Thus, in a sense, vector CAs are still usual CAs with a complicated EL.

Example 6 (CA6) A two-component vector CA:

$$S_1(n, t+1) \stackrel{M_1}{=} S_1(n, t)S_1(n+1, t) + (M_1 - 1)S_2(n-1, t)S_2(n, t) + c_1 \quad [8]$$

$$S_2(n, t+1) \stackrel{M_2}{=} S_1(n-1, t)S_2(n, t) + S_1(n, t)S_2(n+1, t) + c_2 \quad [9]$$

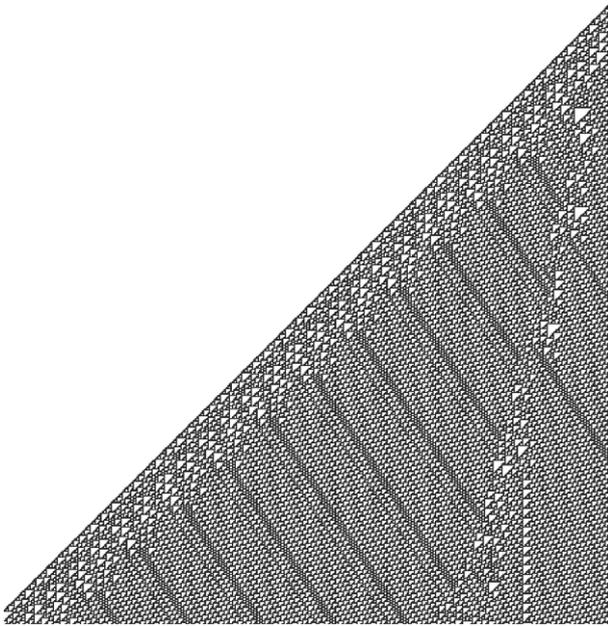


Figure 10 A class-4 CA (Wolfram CA 110): $M=2, V=0, R=1$, EL: $S(n, t+1) \stackrel{2}{=} S(n, t) + S(n+1, t) + S(n, t)S(n+1, t) + S(n-1, t)S(n, t)S(n+1, t)$.

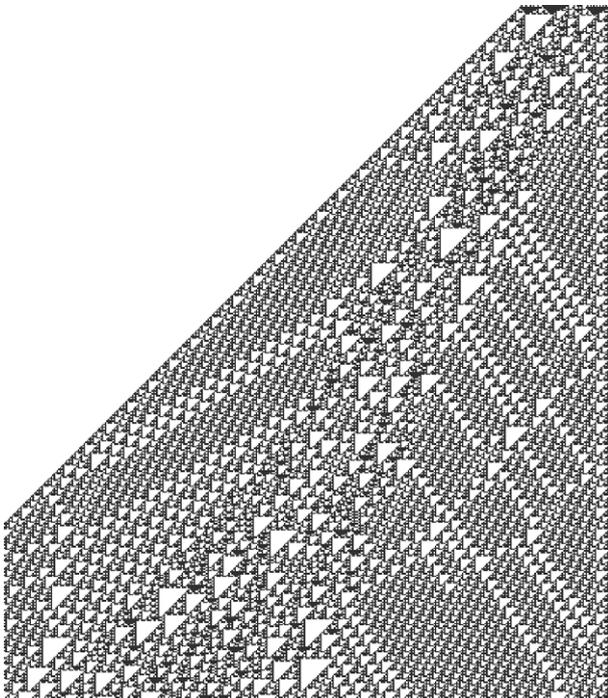


Figure 11 A class-4 CA. Note the interacting moving structures on the left and on the right; note also the apparently chaotic behavior in the center; $M=2, V=0, R=2$, EL: $S(n, t+1) \stackrel{2}{=} S(n, t) + S(n+1, t) + S(n-1, t)S(n+2, t)$.

The global behavior of this CA can be expressed, for example, through the global state function

$$\tilde{S}(n, t) = M_2 S_1(n, t) + S_2(n, t) \quad [10]$$

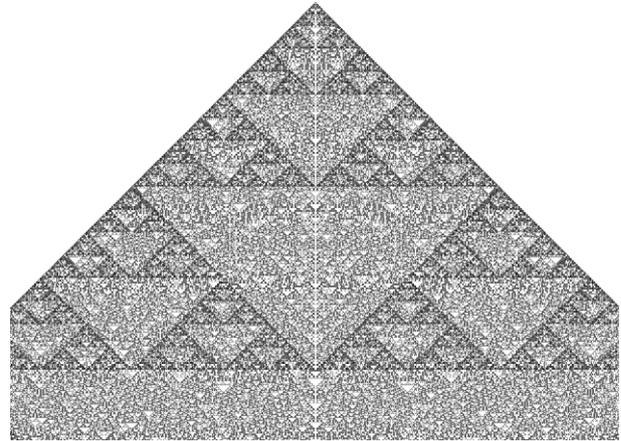


Figure 12 A mixed-class CA: a fractalic structure is superimposed on a chaotic one; $M=4, V=0, R=2$, EL: $S(n, t+1) \stackrel{2}{=} S(n, t)(S(n-2, t) + S(n+2, t)) + S(n-1, t)S(n+1, t)$.

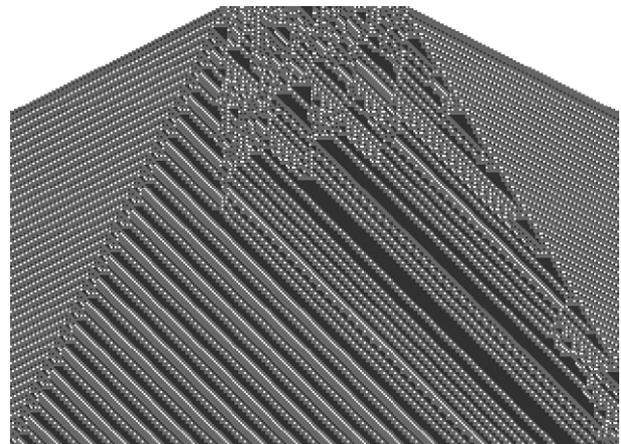


Figure 13 Global behavior of the vector CA6.

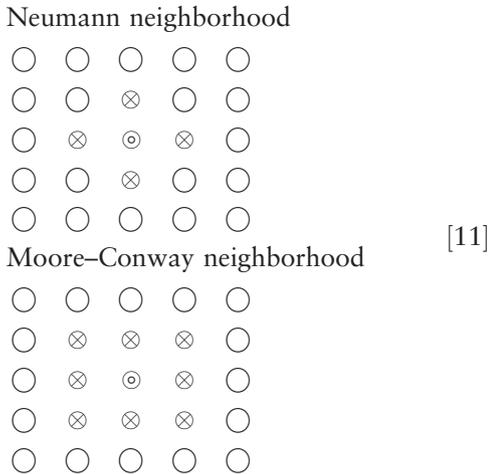
Obviously, $\tilde{S} \in \mathbb{Z}_M$ with $M = M_1 M_2$. **Figure 13** represents the global behavior of this CA with $M_1 = 2, M_2 = 3, c_1 = 1, c_2 = 1, V = 0$.

Note that this CA can be considered as an extension of the celebrated quadratic map.

Multidimensional CA

Up to now we have considered CAs with finite number of cells (finite CAs) or with an infinite number of cells arranged on a line (unidimensional CAs). Now we consider CAs with cells arranged on a surface, usually a plane (bidimensional CAs), or on 3-space (tridimensional CAs), or even on a hyperspace (multidimensional CAs). In any case, if the number of cells is finite, the evolution of such CAs, according to an NEL, must end up to a final cycle: this is due to the finiteness of the “phase space” (thus, these CAs should be classified as class 1; however, note that, if the “phase space” is large enough, the dynamics of

such CAs can still be very rich). Usually, one considers an infinite number of cells tessellating the whole s -space, $s=2,3,\dots$ (e.g., squares or hexagons on the plane, cubic cells in 3-space). The changes in the previous notation and definitions are plain: for example, for a bidimensional CA, the state function depends now on two discrete “space” variables ($S(n_1, n_2, t), n_1 \in \mathbb{Z}, n_2 \in \mathbb{Z}$); furthermore, there is a greater freedom in choosing a neighborhood of range R . Two most-used neighborhoods of range 1 are shown below:



The most famous (and interesting) bidimensional CA is “Life”, introduced by J H Conway, which is discussed next.

Example 7 (CA “Life”; Moore–Conway neighborhood, $V=0, M=2$). A cell in the vacuum state 0 is called “dead”; a cell in the state 1 is called “alive.” The EL is as follows:

- (i) If a cell is dead at time t , it comes alive at time $t + 1$ if and only if exactly three of its eight neighbors are alive at time t (reproduction).
- (ii) If a cell is alive at time t , it dies at time $t + 1$ if and only if fewer than two (loneliness) or more than three (overcrowding) neighbors are alive at time t .

Clearly, this is a totalistic NEL. Now considering the explicit form of σ (see [6]):

$$\sigma(n_1, n_2, t) = -S(n_1, n_2, t) + \sum_{k_1=-1}^1 \sum_{k_2=-1}^1 S(n_1 + k_1, n_2 + k_2, t) \quad [12]$$

the above EL can be simply expressed as:

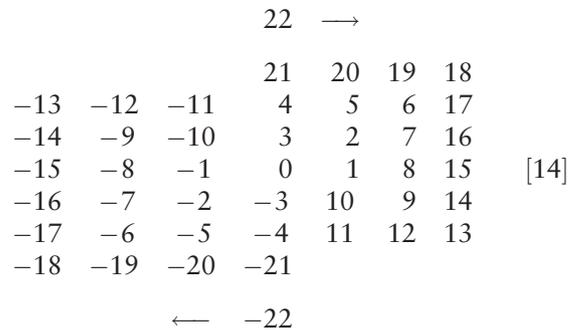
$$S(n_1, n_2, t + 1) = \delta_{3,\sigma} + \delta_{2,\sigma} S(n_1, n_2, t) \quad [13]$$

where $\delta_{3,\sigma}$ is the Kroenecker symbol.

Life is a class-4 CA; it exhibits a rich variety of interesting structures: stable structures, oscillators

(periodic structures), gliders and ships (moving structures), emitters and absorbers (namely, structures that, after a time period, reconstitute themselves, but meanwhile they have emitted or adsorbed moving structures). These structures are essential to prove that Life can be used to construct a universal Turing machine (see below). One can get a rough idea of such “richness” from Figure 14.

As in the previous case of vector CA, one could object that also multidimensional CAs are not true extensions of the unidimensional CAs. Indeed, since the whole set of cells is still a countable set, one could number the cells with just a discrete “space” variable (say $n \in \mathbb{Z}$). For example, in the case of a square tessellation of the plane, we could enumerate the cells in the plane starting from the origin as follows:



Thus, any multidimensional CA could in principle be viewed as a unidimensional one. Of course, one has to pay a price for this: ISs and ELs that are simple for a multidimensional CA become cumbersome for its unidimensional version and vice versa.

Higher Time Derivatives

Up to now, we have considered CAs whose evolved state $S(t + 1)$ depends only on the state $S(t)$, namely the state of the CA itself at the previous time step. In other words the EL involves just the first (discrete) time derivative (1_CA). One can easily extend all the previous definitions to consider higher-order discrete time derivatives (K _CA). Of course, the ID and the IS for such a CA involve the state of the CA at K subsequent time steps.

An example of a unidimensional 2_CA is given below.

Example 8 (CA7) $M=3, V=0, R=1$. The EL is:

$$S(n, t + 1) \stackrel{\text{3}}{=} S(n - 1, t) + S(n, t - 1) + S(n + 1, t) \quad [15]$$

An example of the evolution of such a CA is given in Figure 15.

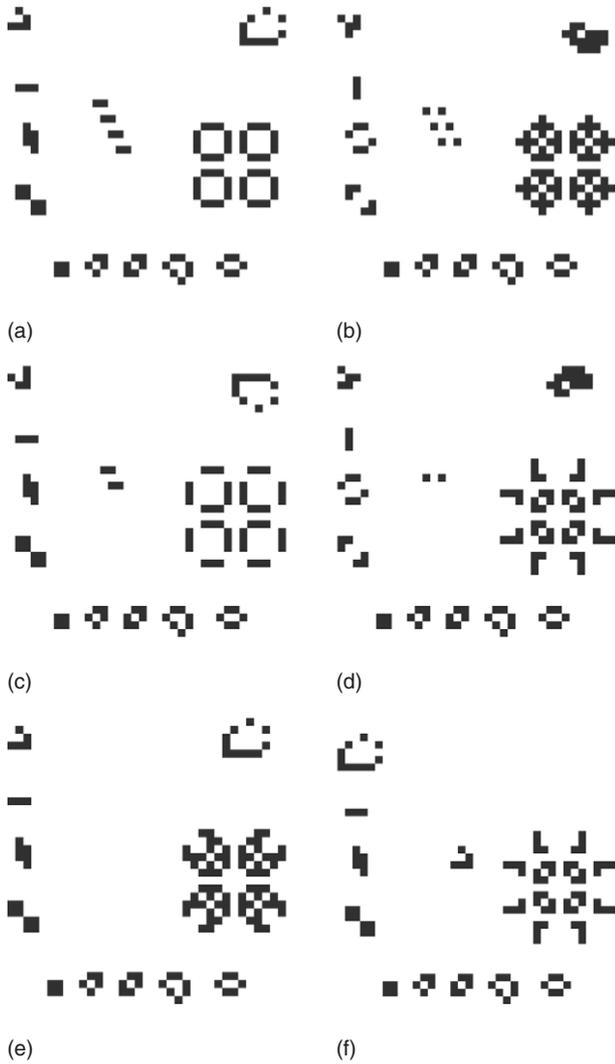


Figure 14 CA “Life”: (a) Time 0. Near the lower border, five stable structures (from the left to the right: a “block”, a “boat”, a “ship”, a “loaf”, a “beehive”); near the left border three “blinkers” (period-2 oscillators); near the right corner, a symmetric structure that, in one time step, evolves into a “pulsar” (a period-3 oscillator), on the left-up corner a “glider” (a moving structure); on the right-up corner a “medium weight spaceship” (another moving structure); in the center, a configuration that vanishes in a few time steps. (b) Time 1. The structures on the lower border are unchanged, the blinkers, the glider, and the space ship are in an intermediate state, on the right border, the pulsar starts to pulse. (c) Time 2. The three blinkers on the left border are again in their original configurations (periodic structure with period 2), the pulsar, the glider and the spaceship are in another intermediate state. (d) Time 3. The pulsar is in its second state, the glider and the spaceship in their third, the structure in the center is going to vanish. (e) Time 4. The pulsar has completed its pulsation (period-3 oscillator, see [Figure 14b](#)); the structure in the center has vanished, the glider and the spaceship have recovered their original configurations (see [Figure 14a](#)) but meanwhile they have moved of a cell in four time steps ($1/4$ of the highest velocity attainable by a moving structure in a CA of range 1). The glider is moving downward and to the right, the space ship in horizontal to the left. (f) Time 60. The space ship has almost completed its crossing, the glider has reached the center and it is in a collision route with the pulsar.

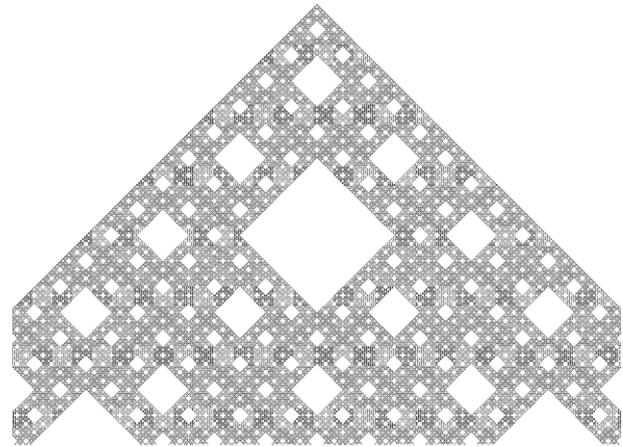


Figure 15 CA7, clearly a class-2 CA.

It is plain that taking a suitable continuum limit of a K -CA one gets a partial differential equation of order K for the evolution. However, there are also special and interesting CAs, called “filter” CAs, that in a suitable continuum limit end up in integral evolution equations. For a filter unidimensional CA, the evolved state at the cell n , $S(n, t + 1)$, depends also on the (already) evolved states of the cells on its left (or right): for example, an NEL of the type

$$\begin{aligned}
 S(n, t + 1) &\stackrel{M}{=} F(S(n + k_i, t), S(n - \tilde{k}_j, t + 1)) \\
 i &= 1, 2, \dots, N; \quad k_i \in \mathbb{Z} \\
 j &= 1, 2, \dots, \tilde{N}; \quad \tilde{k}_j \in \mathbb{N}
 \end{aligned}
 \tag{16}$$

is still valid (computable). Extensions to K -CAs or vector CAs or multidimensional CA are plain. Very often filter CAs exhibit a class-4 behavior with particle-like structures moving and interacting in a complex way; see the following example and examples in the next section.

Example 9 (CA8) $M = 2, V = 0, R = 2$. The EL is:

$$\begin{aligned}
 S(n, t + 1) &\stackrel{2}{=} S(n - 1, t - 1)S(n - 2, t) \\
 &+ S(n, t) + S(n + 1, t)S(n + 2, t)
 \end{aligned}
 \tag{17}$$

An example of the evolution of such a CA is given in [Figure 16](#).

Invertible CA

For most of the ELs there is a loss of information in the course of the evolution (see, e.g., [Figures 5 and 6](#)). Indeed, different definitions of “CA entropy” have been introduced to measure the “randomness” in the behavior of a given CA. However, since CAs are important in physical

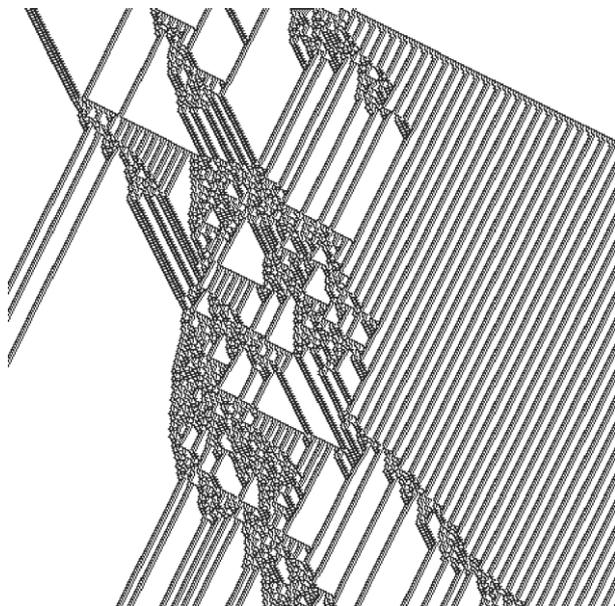


Figure 16 CA8, a “filter” CA. Note the emerging of particle-like structures moving to the left and to the right and interacting in complex ways.

modeling as well as in cryptography and data compression, there is great interest in a special subclass of CAs which are “invertible” (time reversible). Namely, for an “invertible” CA following a given EL and starting from an arbitrary ID, there exists an “inverse” EL such that one can recover the ID from the evolved states. Invertible CAs can be easily devised in the case of K -CA ($K > 1$). For example, if $K = 2, 3 \dots$, one can consider ELs of the form

$$S(n, t + 1) \stackrel{M}{=} S(n, t - K + 1) + F(S(n + k_i^j, t - j)) \quad [18a]$$

where

$$\begin{aligned} i &= 1, 2, \dots, N^j; \quad k_i^j \in \mathbb{Z} \\ j &= 0, 1, 2, \dots, K - 2 \end{aligned} \quad [18b]$$

and F is an arbitrary polynomial function.

It is then clear that the inverse EL reads

$$\begin{aligned} \tilde{S}(n, \tilde{t} + 1) &\stackrel{M}{=} \tilde{S}(n, \tilde{t} - K + 1) \\ &+ (M - 1)F(\tilde{S}(n + k_i^j, \tilde{t} + j - K + 2)) \end{aligned} \quad [19]$$

Indeed, if an arbitrary ID evolves according to the EL [18], then applying the inverse EL [19] to K subsequent evolved states (taken in reversed order), eventually the original ID is recovered (in reversed order) (see the following example).

Example 10 (CA9) A 6-CA: $M = 2, V = 0, R = 1$. The EL is:

$$\begin{aligned} S(n, t + 1) &\stackrel{2}{=} S(n, t - 5) + S(n, t - 3) + S(n + 1, t - 2) \\ &+ S(n - 1, t - 1) \\ &+ S(n, t - 2)S(n + 1, t - 2) \\ &+ S(n, t)S(n - 1, t) \end{aligned} \quad [20]$$

The inverse EL, according to [19], reads (Figure 17)

$$\begin{aligned} \tilde{S}(n, \tilde{t} + 1) &\stackrel{2}{=} \tilde{S}(n, \tilde{t} - 5) + \tilde{S}(n, \tilde{t} - 1) + \tilde{S}(n + 1, \tilde{t} - 2) \\ &+ \tilde{S}(n - 1, \tilde{t} - 3) \\ &+ \tilde{S}(n, \tilde{t} - 2)\tilde{S}(n + 1, \tilde{t} - 2) \\ &+ \tilde{S}(n, \tilde{t} - 4)\tilde{S}(n - 1, \tilde{t} - 4) \end{aligned} \quad [21]$$



(a)



(b)

Figure 17 CA9, a 6-CA: (a) a 50 time-step evolution from a peculiar ID; (b) a 50 time-step evolution of the inverse EL, starting from the last six configurations of Figure 17a (taken in inverse order); the ID of Figure 17a is recovered (in inverse order).

Of course, more complicated invertible ELs can be devised. Invertible ELs can be also easily devised for “filter” CA, for example, if an NEL for a “filter” CA reads

$$S(n, t + 1) \stackrel{M}{=} S(n, t) + F(S(n + k_i, t), S(n - \tilde{k}_j, t + 1)) \quad [22]$$

where k_i and \tilde{k}_j are positive integers ($i = 1, 2, \dots, N; j = 1, 2, \dots, \tilde{N}$) and F is an arbitrary (polynomial) function, then it is invertible and the inverse NEL reads

$$\tilde{S}(n, \tilde{t} + 1) \stackrel{M}{=} \tilde{S}(n, \tilde{t}) + (M - 1) \times F(\tilde{S}(n + k_i, \tilde{t} + 1), \tilde{S}(n - \tilde{k}_j, \tilde{t})) \quad [23]$$

Note that [22] is computable starting from $n = -\infty$, whereas [23] is computable starting from $n = +\infty$.

Example 11 (CA10) A 1.5-CA, $M = 2, V = 0, R = 3$. The EL is:

$$S(n, t + 1) \stackrel{2}{=} S(n, t) + S(n - 3, t + 1)S(n - 2, t + 1) + S(n + 2, t)S(n + 3, t) + S(n - 2, t + 1)S(n - 1, t + 1) + S(n + 1, t)S(n + 2, t) \quad [24]$$

Note that this EL is of the form [22]; therefore, it is invertible (see Figure 18a). According to [23], the inverse EL reads:

$$\tilde{S}(n, \tilde{t} + 1) \stackrel{2}{=} S(n, \tilde{t}) + \tilde{S}(n + 3, \tilde{t} + 1)\tilde{S}(n + 2, \tilde{t} + 1) + \tilde{S}(n - 2, \tilde{t})\tilde{S}(n - 3, \tilde{t}) + \tilde{S}(n + 2, \tilde{t} + 1)\tilde{S}(n + 1, \tilde{t} + 1) + \tilde{S}(n - 1, \tilde{t})\tilde{S}(n - 2, \tilde{t}) \quad [25]$$

This CA exhibits a very rich dynamics: any complex ID rapidly decays in a great variety of coherent particle-like structures, steady or moving to the right or

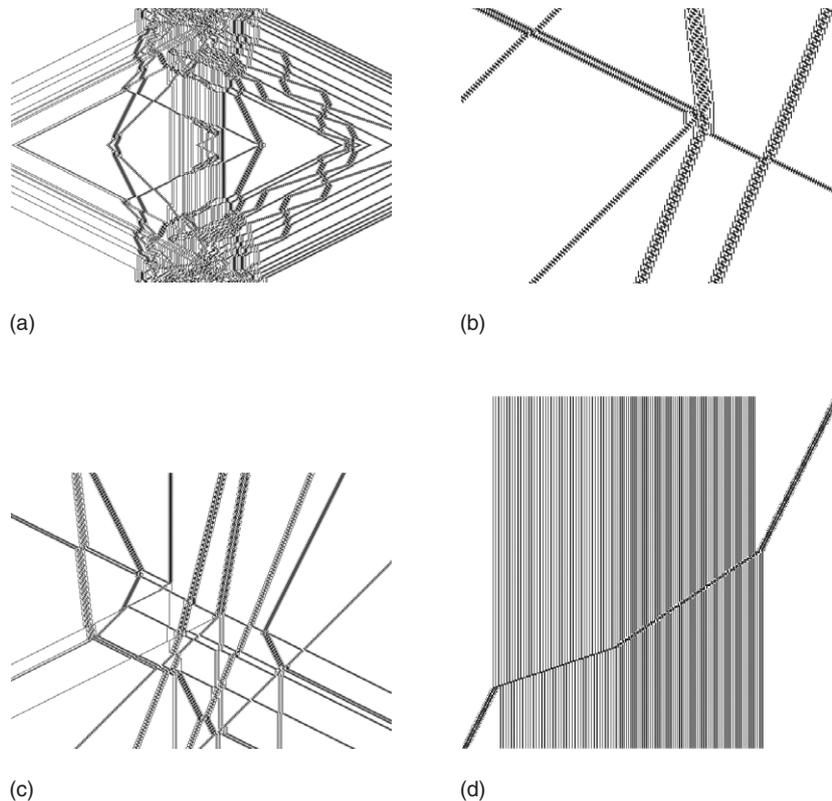


Figure 18 CA10: (a) 230 time-step evolution, then the inverse EL is applied for 230 further time step in order to recover the initial configuration. (b) Collisions between different kinds of particle-like coherent moving structures. The last collision (on the right) is a solitonic one: the interaction produces just a phase shift, preserving number, shape, and velocities of the involved “particles.” (c) “Particles” moving with different velocities and interacting in complex ways (solitonic collisions, particle creations and annihilations). (d) A particle goes through a nonhomogeneous medium and undergoes refraction by the medium itself.

to the left with different velocities. The interactions between different particles may be solitonic (the particles emerge unchanged but shifted) or annihilation–creation phenomena can occur (see [Figures 18a–d](#)).

Applications of CAs

CAs as Universal Constructors and Turing Machines

In the 1950s, von Neumann, who contributed to the development of the first computer (ENIAC), decided to work out a mathematical theory of automata. Such a theory was finalized to give an answer to the following question: is it possible to build an automaton such that it allows universal computation (i.e., it embodies a universal Turing machine) and, moreover, it is able to build (in order of decreasing generality)

1. an arbitrary automata (universal constructor);
2. a copy of itself (self-reproducing); and
3. an automaton that is itself a universal Turing machine (constructor)?

The last question von Neumann had intention to address was if in the process of automata self-reproduction (if possible) a process of evolution could take place, that is, if a simpler automaton could generate a more complex one.

In the beginning, the idea of von Neumann was to describe, using mathematical axioms, an automaton moving inside a warehouse and selecting various elementary spare parts (e.g., “muscles,” switches, rigid girders) and then assembling them into a new automaton. While this original idea was very realistic, it was also very difficult to pursue, so that von Neumann, following a suggestion by Ulam, decided to consider his questions in the more abstract framework of CAs.

The particular CA he considered is an infinite square CA with 29 possible states. The transition rule is dependent upon the cell to update and its north, east, south, and west neighbor cell (the von Neumann neighborhood). Among the 29 possible states there is one state that is “quiescent” (the vacuum state).

von Neumann proved the existence of a configuration of $\sim 50\,000$ cells immersed in a sea of quiescent states that embodies a universal Turing machine and that is a universal constructor. An infinite one-dimensional “tape” is used to store a description of the automaton to build. The universal constructor reads the description on the tape, develops a “constructing arm” that builds the configuration described on the tape in an unoccupied part of the cellular space, makes a copy of the tape and finally attaches it to the newly built automaton and retracts

the constructing arm. When on the tape, it stores a description of the universal constructor itself, then it self-reproduces. The total size of the self-reproducing automaton amounts to $\sim 200\,000$ cells. (Some computer simulations of von Neumann self-reproducing automaton are available on the web.)

Since von Neumann’s CA is a very complex one, it led researchers to think that a CA able to simulate a universal Turing machine should also be quite complex. The perspective changed completely after the introduction of CA Life. Conway was looking for a simple CA with a possible rich dynamics; however, it was subsequently realized that Life was much more complicated than anyone could have thought. Finally, thanks to the development of faster computers that allowed visualization of the evolution of quite large populations and through the contribution of a large number of researchers, it was proved that a universal Turing machine could be embedded in Life.

The discovery that even a simple CA such as Life could incorporate a universal Turing machine led to the question whether it could be possible to build a universal Turing machine inside a simple one-dimensional CA. This is indeed the case: up to now, the simplest CA capable of universal computation is the W110 CA (see [Figure 10](#)), as proved recently by Cook after a conjecture formulated by Wolfram in 1985.

CAs for Computer Simulations

One of the major applications of CAs is the computer simulation of various dynamical processes. Even if CAs were not invented for this purpose, they possess peculiarities that make them particularly suitable for this task. The main advantage of using a CA for a dynamical simulation is due to their completely discrete nature that allows exact simulations on a computer. Thus, any spurious effect due to rounding errors is ruled out. Another advantage is that the EL of a CA can be seen as a function between finite sets. For this reason, one can specify the EL through a “lookup table” (see [2]): then when running the simulations, the computer has only to access the table instead of computing the function every time, shortening considerably the computation time. Another great advantage of CAs in computer simulations is that, for their very nature (at least for local EL), they can be implemented on parallel machines. These two concepts are at the basis of dedicated computers for CAs simulations developed by Toffoli, Margolus, and co-workers (CAM series). The possibility to use efficiently parallel computers for CA simulation could prove

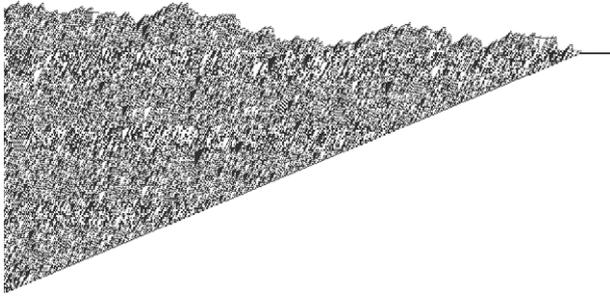


Figure 19 A CA that “computes” the $3n + 1$ Collatz–Ulam map. The ID for the CA is the initial number for the iterated map (binary notation, order 2^{300} , randomly chosen, displayed on the left vertical axis). The CA, according to the Collatz conjecture, ends up to the final stable configuration (horizontal line on the right for the CA, $1 \rightarrow 4 \rightarrow 2 \rightarrow 1$ for the map).

to be fundamental when computer speeds approach saturation. Moreover, CAs themselves can mimic parallel computations, see, for example, **Figure 19**, where a nonlocal CA “computes” very efficiently the celebrated Collatz–Ulam $3n + 1$ map.

CAs in Physics

Since Newton, physics has been described through differential equations and continuous functions. However, such a mathematical description is not fit for simulation on a computer, and some discretizations must be considered. First, one has to discretize space and time passing from differential equations to (finite systems of) finite difference equations; second, one has to round off the values of the functions to store them in the memory of the computer. The main drawback of this procedure is that in chaotic systems such approximations can rapidly lead to great differences between the real and the simulated behavior. As already noticed, this problem does not appear in CA. Thus, one would like to use this good characteristic of CAs in physical modeling taking due account of the continuous nature of the physics involved. This requires attention and ingenuity in constructing reliable CA models for physical processes. For example, this goal has been achieved in the so-called lattice gas automata (LGAs).

LGAs are CA models for the microscopic dynamics of fluids and gases. The thermodynamic limit of these CAs yields the correct continuous functions for the macroscopic quantities (density, pressure, viscosity, etc.).

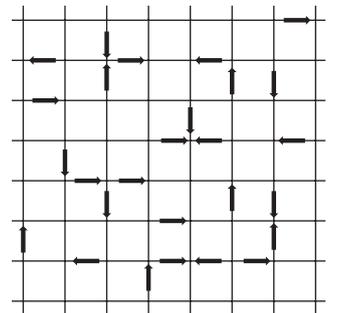
The first step toward LGAs was the discovery that the HPP model developed in the 1970s by Hardy, Pomeau and De Pazzis was in fact a CA. The HPP model describes the behavior of a fluid (or a gas) in a plane. The configuration space is given by a

bidimensional square lattice and the particles are described by arrows lying on the edges of the lattices and pointing to some vertex (see **Figure 20a**).

The particles are assumed to be all identical and with the same velocity, and particles on the same edge with the same direction are not allowed (exclusion principle). The EL prescribes that particles move with unitary velocity along the edges in the direction pointed by the arrow (free flight) unless there are exactly two particles on the edges connected to a given vertex and they point in opposite directions (collision); in this case they are replaced by two arrows pointing outward on the previously empty edges (see **Figure 20b**). Clearly, the EL conserves the number and the momentum of the particles.

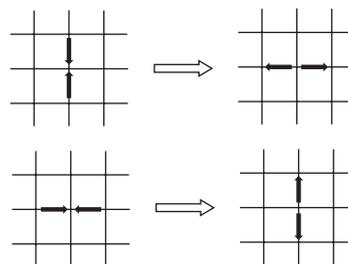
The HPP model can be described algebraically. The admissible particle velocities are just

$$c_1 = +\hat{x}, \quad c_2 = +\hat{y}, \quad c_3 = -\hat{x}, \quad c_4 = -\hat{y} \quad [26]$$

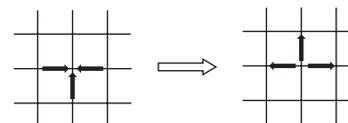


(a)

Collisions



Free flight



(b)

Figure 20 (a) An example of configuration for the HPP model. (b) Head on collisions and three particle collisions in the HPP model.

Accordingly, only four bits $n_j(\mathbf{x}, t)$, $j = 1, 2, 3, 4$, are required to denote the presence (1) or the absence (0) of a particle with velocity c_j pointing vertex \mathbf{x} at time t . The dynamical rule for HPP can be written in the form

$$n_j(\mathbf{x} + c_j, t + 1) = n_j(\mathbf{x}, t) + \omega_j(\mathbf{x}, t) \quad [27]$$

where term $n_j(\mathbf{x}, t)$ on the right-hand side accounts for the free flight of particles, while $\omega_j(\mathbf{x}, t)$ modifies the trajectories in the case of collisions. The ω_j are determined by the state of the system according to the following rules:

$$\begin{aligned} \omega_1 = & -n_1(1 - n_2)n_3(1 - n_4) \\ & + (1 - n_1)n_2(1 - n_3)n_4 \end{aligned} \quad [28a]$$

$$\begin{aligned} \omega_2 = & -n_2(1 - n_3)n_4(1 - n_1) \\ & + (1 - n_2)n_3(1 - n_4)n_1 \end{aligned} \quad [28b]$$

$$\begin{aligned} \omega_3 = & -n_3(1 - n_4)n_1(1 - n_2) \\ & + (1 - n_3)n_4(1 - n_1)n_2 \end{aligned} \quad [28c]$$

$$\begin{aligned} \omega_4 = & -n_4(1 - n_1)n_2(1 - n_3) \\ & + (1 - n_4)n_1(1 - n_2)n_3 \end{aligned} \quad [28d]$$

It is plain that eqns [27] and [28] can be interpreted as the EL for a CA.

In the thermodynamic limit, the equations governing the dynamics of the macroscopic quantities of the fluid are given by the continuity equation and by anisotropic Navier–Stokes equations. The anisotropy in the Navier–Stokes equations is due to the fact that the invariance group of the square lattice is too small. This problem was solved by Frisch, Hasslacher, and Pomeau in 1986, with the introduction of the FPP model. It turns out that a hexagonal lattice has enough symmetries to recover the isotropic Navier–Stokes equations in the thermodynamic limit. So, the FPP model is an example of a model where even if the microscopic dynamics is almost a caricature of the real dynamics, the thermodynamic limit gives rise to the correct physical equations.

CAs have been used to simulate many other physical processes (unfortunately, there is no space here for a sufficiently elaborate description). The principal fields of application are: percolation theory, magnetism, diffusion phenomena, sandpiles, models of earthquakes, crystal growth, etc.

The more intriguing aspect of some even simple CAs (e.g., CA9, CA10: see **Figures 16 and 18**) is their very rich particle-like dynamics. For instance, the existence of solitonic collisions suggested that the techniques recently developed to find and treat “integrable”

nonlinear dynamical systems (nonlinear continuous and discrete evolution equations, many-body problems) could profitably be extended to find “integrable” CAs. Indeed, many such CAs have been found that exhibit “solitons” and are endowed with non-trivial conservation laws (of course, this is very important in physical modeling). Moreover, the above-cited similarity between certain CA behaviors and elementary particle physics phenomena suggests that the fundamental structure of reality (at the Planck level) could indeed be that of a CA (cells of Planck length, discrete time flow): attempts to construct this underlying CA physics have been pursued.

Other Applications

CAs exhibit a great plasticity, which makes them well suited to model systems in a wide range of fields. This is mainly due to the fact that CAs with very simple rules can also simulate universal Turing machines, so that they can exhibit a very rich and complicated overall dynamics (in principle, one could simulate any dynamical system using a simple CA). There is another reason for the wide applicability of CA modeling even outside of physics: namely, it is well known that algorithms, not differential equations, are better instruments to schematize dynamical processes for complex and organized systems. Since simple algorithms can be naturally implemented on CAs, the latter are very useful for realizing simple models and simulations in many fields: biology, economics, ecology, neural networks, traffic models, etc.

Moreover, applications of CAs in informatics and specifically in cryptography and data compression have been investigated.

See also: Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Generic Properties of Dynamical Systems; Integrable Systems: Overview.

Further Reading

- Berlekamp ER, Conway JH, and Guy R (1982) *Winning Ways for Your Mathematical Plays*. London: Academic Press.
- Boghosian BM (1999) Lattice gases and cellular automata. *Future Generation Computer Systems* 16: 171–185.
- Boon JP, Dab D, Kapral R, and Lawniczak A (1996) Lattice gas automata for reactive systems. *Physics Reports* 273: 55–147.
- Burks AW (ed.) (1970) *Essay on Cellular Automata*. Urbana: University of Illinois Press.
- Chopard B and Droz M (1998) *Cellular Automata Modeling of Physical Systems*. Cambridge: Cambridge University Press.
- Doolen G (ed.) (1990) *Lattice Gas Methods for Partial Differential Equations*. New York: Addison-Wesley.
- Gardner M (1983) *Wheels, Life, and Other Mathematical Amusements*. New York: W H Freeman.

Jackson EA (1990) *Perspectives of Nonlinear Dynamics*. Cambridge: Cambridge University Press.
 Toffoli T and Margolus N (1987) *Cellular Automata Machines – A New Environment for Modeling*. Cambridge: The MIT Press.

von Neumann J (1966) In: Burks AW (ed.) *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press.
 Wolfram S (2002) *A New Kind of Science*. Champaign: Wolfram Media.

Central Manifolds, Normal Forms

P Bonckaert, Universiteit Hasselt, Diepenbeek, Belgium

© 2006 Elsevier Ltd. All rights reserved.

Introduction

We consider differentiable dynamical systems generated by a diffeomorphism or a vector field on a manifold. We restrict to the finite-dimensional case, although some of the ideas can also be developed in the general case (Vanderbauwhede and Iooss 1992). We also restrict to the behavior near a stationary point or a periodic orbit of a flow.

Let the origin 0 of \mathbf{R}^n be a stationary point of a C^1 vector field X , that is, $X(0) = 0$. We consider the linear approximation $A = dX(0)$ of X at 0 and its spectrum $\sigma(A)$, which we decompose as $\sigma(A) = \sigma_s \cup \sigma_c \cup \sigma_u$, where σ_s resp. σ_c resp. σ_u consists of those eigenvalues with real part < 0 resp. $= 0$ resp. > 0 . If $\sigma_c = \emptyset$ then there is no central manifold, and the stationary point 0 is called hyperbolic. Let E_s, E_c , and E_u be the linear A -invariant subspaces corresponding to σ_s resp. σ_c resp. σ_u . Then $\mathbf{R}^n = E_s \oplus E_c \oplus E_u$. We look for corresponding X -invariant manifolds in the neighborhood of 0, in the form of graphs of maps. More precisely:

Theorem 1 *Let the vector field X above be of class C^r ($1 \leq r < \infty$). There exist map germs $\phi_{ss}: (E_s, 0) \rightarrow E_c \oplus E_u$, $\phi_{sc}: (E_s \oplus E_c, 0) \rightarrow E_u$, $\phi_{uu}: (E_u, 0) \rightarrow E_s \oplus E_c$, $\phi_{cu}: (E_c \oplus E_u, 0) \rightarrow E_s$, and $\phi_c: (E_c, 0) \rightarrow E_s \oplus E_u$ of class C^r such that the graphs of these maps are invariant for the flow of X . Moreover, these maps are of class C^r , and their linear approximation at 0 is zero, that is, their graphs are tangent to, respectively, $E_s, E_s \oplus E_c, E_u, E_c \oplus E_u$, and E_c . If X is of class C^∞ then ϕ_{ss} and ϕ_{uu} are also of class C^∞ . If X is analytic then ϕ_{ss} and ϕ_{uu} are also analytic.*

The graph of ϕ_c is called the (local) central (or, center) manifold of X at 0 and it is often denoted by W^c . Thus, it is an invariant manifold of X tangent at the generalized eigenspace of $dX(0)$ corresponding to the eigenvalues having zero real part.

(Non) uniqueness, Smoothness

Most proofs in the literature (Vanderbauwhede 1989) use a cutoff in order to construct globally defined objects, and then obtain the invariant graph as the solution of some fixed-point problem of a contraction in an appropriate function space. Although this solution is unique for the globalized problem, this is not the case at the germ level: another cutoff may produce a different germ of a central manifold. In other words, locally a central manifold might not be unique, as is easily seen on the planar example $x^2\partial/\partial x - y\partial/\partial y$. On the other hand, the ∞ -jet of the map ϕ_c , in case of a C^∞ vector field, is unique, so if there would exist an analytic central manifold then this last one is unique; in the foregoing example, it is the x -axis. But for the (polynomial) example $(x - y^2)\partial/\partial x + y^2\partial/\partial y$ one can calculate that the ∞ -jet of $x = \phi_c(y)$ is given by $j_\infty\phi_c(y) = \sum_{n \geq 1} n!y^{n+1}$, which has a vanishing radius of convergence, so there is no analytic central manifold. On the other hand, by the Borel theorem we can choose a C^∞ -representative for ϕ_c . This can be generalized in the planar case:

Proposition 1 *If $n = 2$ and if X is C^∞ and if the ∞ -jet of X in the direction of the central manifold is nonzero, then this central manifold is C^∞ . In particular, if X is analytic then the central manifold is either an analytic curve of stationary points or is a C^∞ curve along which X has a nonzero jet.*

For proofs and additional reading, the reader is referred to Aulbach (1992). In general, a central manifold is not necessarily C^∞ (van Strien 1979, Arrowsmith and Place 1990): for the system in \mathbf{R}^3 given by

$$(x^2 - z^2) \frac{\partial}{\partial x} + (y + x^2 - z^2) \frac{\partial}{\partial y} + 0 \cdot \frac{\partial}{\partial z}$$

one can find a C^k central manifold for every k but there is no C^∞ central manifold. Indeed, in this case the domain of definition of ϕ_c shrinks to zero when k tends to infinity.

Central Manifold Reduction

The importance of a central manifold lies in the principle of central manifold reduction, which roughly says that for local bifurcation phenomena it is enough to study the behavior on the central manifold, that is, if two vector fields, restricted to their central manifolds, have homeomorphic integral curve portraits, and if the dimensions of E_s and E_u are equal, then the two vector fields have homeomorphic integral curve portraits in \mathbf{R}^n , at least locally near 0. Let us be more precise:

Theorem 2 *Let m be the dimension of E_c . There exists p , $0 \leq p \leq n - m$, such that X is locally C^0 -conjugate to*

$$X' = \sum_{i=1}^m \tilde{X}_i(z_1, \dots, z_m) \frac{\partial}{\partial z_i} + \sum_{i=m+1}^{m+p} z_i \frac{\partial}{\partial z_i} - \sum_{i=m+p+1}^n z_i \frac{\partial}{\partial z_i}$$

where (z_1, \dots, z_m) is a coordinate system on a central manifold, (z_1, \dots, z_n) is a coordinate system on \mathbf{R}^n extending (z_1, \dots, z_m) and $\sum_{i=1}^m \tilde{X}_i \partial/\partial z_i$ is the restriction of X to a central manifold. Moreover, if

$$Y = \sum_{i=1}^m \tilde{Y}_i(z_1, \dots, z_m) \frac{\partial}{\partial z_i} + \sum_{i=m+1}^{m+p} z_i \frac{\partial}{\partial z_i} - \sum_{i=m+p+1}^n z_i \frac{\partial}{\partial z_i}$$

and if $\sum_{i=1}^m \tilde{Y}_i \partial/\partial z_i$ is C^0 -equivalent (resp. C^0 -conjugate) to $\sum_{i=1}^m \tilde{X}_i \partial/\partial z_i$ then X is C^0 -equivalent (resp. -conjugate) to Y .

For a proof and further reading (a generalization) see [Palis and Takens \(1977\)](#).

In case that more smoothness than just C^0 is needed, we have the principle of normal linearization along the central manifold. More concretely, let x denote a coordinate in the central manifold and let y be a complementary variable, that is, let $X = X_c \partial/\partial x + X_b \partial/\partial y$. We define the normally linear part along the central manifold by

$$NX := X_c(x, 0) \frac{\partial}{\partial x} + \frac{\partial X_b}{\partial y}(x, 0) \cdot y \frac{\partial}{\partial y}$$

Under certain nonresonance conditions ([Takens 1971](#), [Bonckaert 1997](#)) on the real parts of the eigenvalues of $dX(0)$, there exists a C^r local conjugacy between X and NX for each $r \in \mathbf{N}$ (assuming X to be of class C^∞). If there are resonances, then one can conjugate with the

so-called seminormal or renormal form containing higher-order terms (see [Bonckaert \(1997, 2000\)](#) and references therein; here one can also find results for cases where extra constraints should be respected, like symmetry, reversibility, or invariance of some given foliation etc.).

Parameters

Having an eigenvalue with zero real part is ungeneric, so in bifurcation problems we consider p -parameter families X_λ near, say, $\lambda=0$. With respect to the results above, we remark that such a family can be considered as a vector field near $(0, 0) \in \mathbf{R}^n \times \mathbf{R}^p$ tangent to the leaves $\mathbf{R}^n \times \{\lambda\}$. In fact, the parameter direction \mathbf{R}^p is contained in E_c . In all the results mentioned, this structure “of being a family” is respected. For example, in [Theorem 2](#) we replace $\tilde{X}_i(z_1, \dots, z_m)$ by $\tilde{X}_i(z_1, \dots, z_m, \lambda)$. Hence, if \tilde{X}_λ is a versal unfolding of \tilde{X}_0 then X_λ is a versal unfolding of X_0 . By this, the search for versal unfoldings is reduced to the unfolding of singularities whose linear approximation at 0 has a purely imaginary spectrum.

Diffeomorphisms, Periodic Orbits

A completely analogous theory can be developed for fixed points of diffeomorphisms $f: (\mathbf{R}^n, 0) \rightarrow \mathbf{R}^n$. Here we split up the spectrum of the linear part $L = df(0)$ at 0 as $\sigma(L) = \sigma_s \cup \sigma_c \cup \sigma_u$, where σ_s resp. σ_c resp. σ_u consists of those eigenvalues with modulus <1 resp. $=1$ resp. >1 . This theory can be applied to the time- t map of a vector field (and will give the same invariant manifolds) and to the Poincaré map of a transversal section of a periodic orbit of a vector field ([Chow et al. 1994](#)).

Normal Forms

The general idea of a normal form is to put a (complicated) system into a form “as simple as possible” by means of a change of coordinates. This idea was already developed to a great extent by H Poincaré. Simple examples are: (1) putting a square matrix into Jordan form, (2) the flow box theorem ([Arrowsmith and Place 1990](#)) near a nonsingular point. Depending on the context and on the purpose of the simplification, this concept may vary greatly. It depends on the kind of changes of coordinates that are tolerated (linear, polynomial, formal series, smooth, analytic) and on the possible structures that must be preserved (e.g., symplectic, volume-preserving, symmetric, reversible etc.). Let us restrict to local normal forms, that is, in the vicinity of a stationary point of a vector field or a diffeomorphism (the latter can be

applied to the Poincaré map of a periodic orbit). We concentrate on the simplification of the Taylor series. The general idea is to apply consecutive polynomial changes of variables; at each step we simplify terms of a degree higher than in the step before. The ideal simplification would be to put all higher-order terms to zero, which would (at least at the level of formal series) linearize the system. But as soon as there are resonances (see below), this is impossible: the planar system $2x\partial/\partial x + (y + x^2)\partial/\partial y$ cannot be formally linearized.

Setting

Let X be a C^{r+1} vector field defined on a neighborhood of $0 \in \mathbb{R}^n$, and denote $A = dX(0)$ (its linear approximation at 0). The Taylor expansion of X at 0 takes the form

$$X(x) = A \cdot x + \sum_{k=2}^r X_k(x) + O(|x|^{r+1})$$

where $X_k \in H^k$, the space of vector fields whose components are homogeneous polynomials of degree k . The classical formal normal-form theorem is as follows. We define the operator L_A on H^k by putting $L_A b(x) = db(x) \cdot A \cdot x - A \cdot b(x)$; one calls L_A the homological operator. One checks that $L_A(H^k) \subset H^k$. One also denotes this by $\text{ad } A(b)(x)$; see further in the Lie algebra setting. Let R^k be the range of L_A , that is, $R^k = L_A(H^k)$. Let G^k denote any complementary subspace to R^k in H^k . The formal normal-form theorem states, under the above settings:

Theorem 3 (Chow *et al.* 1994, Dumortier 1991) *There exists a composition of near identity changes of variables of the form*

$$x = y + \xi^k(y) \tag{1}$$

where the components of ξ^k are homogeneous polynomials of degree k , such that the vector field X is transformed into

$$Y(y) = A \cdot y + \sum_{k=2}^r g_k(y) + O(|y|^{r+1})$$

where $g_k \in G^k, k = 2, \dots, r$.

Sometimes this theorem is applied to the restriction of a vector field to its central manifold, for reasons explained in the last section. This is the reason why we did not assume X to be C^∞ ; in the latter case one can let $r \rightarrow \infty$ and obtain a normal form on the level of formal Taylor series (also called ∞ -jets). Using a theorem of Borel, we infer the existence of a C^∞ change of variables ϕ such that

the Taylor series of $\phi_*(X)$ is $A \cdot y + \sum_{k=2}^\infty g_k(y)$. For practical computations, it is often appropriate to first simplify the linear part A and to diagonalize it whenever possible. Hence, it is convenient to use a complexified setting and to use complex polynomials or power series. One can show that all involved changes of variables preserve the property of “being a complex system coming from a real system,” that is, at the final stage we can return to a real system (see, e.g., Arrowsmith and Place (1990) for a more precise mathematical description).

Hence, we can assume that A is an upper triangular matrix. Let the eigenvalues be $\lambda_1, \dots, \lambda_n$. It can be calculated that the eigenvalues of L_A , as an operator $H^k \rightarrow H^k$, are then the numbers $\langle \lambda, \alpha \rangle - \lambda_j$ where $\alpha \in \mathbb{N}^n, \sum_{j=1}^n \alpha_j = k$ and $1 \leq j \leq n$. Hence, if these would all be nonzero then $B^k = H^k$, and then we have an ideal simplification, that is, all g_k equal to zero. However, if such a number is zero, that is,

$$\langle \lambda, \alpha \rangle - \lambda_j = 0 \tag{2}$$

it is called a resonance between the eigenvalues. In such a case, we have to choose a complementary space G^k . From linear algebra it follows that one can always choose

$$G^k = \ker(L_{A^*}) \tag{3}$$

where A^* is the adjoint operator. But this choice [3] is not unique and is, from the computational point of view, not always optimal, especially if there are nilpotent blocks. This fact has been exploited by many authors. A typical example is the case where $A = y\partial/\partial x$. On the other hand, if A is semisimple we can choose the complementary space to be $\ker(L_A)$, so $L_A g_k = 0$; we can assume it to be the (complex) diagonal $[\lambda_1, \dots, \lambda_n]$. In that case we can be more explicit as follows. Let $e_j = \partial/\partial x_j$ denote the standard basis on C^n . For a monomial one can calculate that

$$L_A(x^\alpha e_j) = (\langle \lambda, \alpha \rangle - \lambda_j)x^\alpha e_j \tag{4}$$

If the latter is zero, then the monomial is called resonant. This implies that the normal form can be chosen so that it only contains resonant monomials.

Putting a system into normal form not only simplifies the original system, it also gives more geometric insight on the Taylor series. To be more precise, suppose (for simplicity, this can be generalized (Dumortier 1997)) that A is semisimple. One can calculate that the condition $L_A g_k = 0$ implies: $\exp(-At)g_k(\exp(At)x) = g_k(x)$ for all $t \in \mathbb{R}$. This means that g_k is invariant for the one-parameter group $\exp(At)$. A typical example in the plane is: A has eigenvalues $i\lambda, -i\lambda$. Note that the (only) resonances are $\langle (i\lambda, -i\lambda), (p+1, p) \rangle - i\lambda = 0$ and

$\langle (i\lambda, -i\lambda), (p, p+1) \rangle + i\lambda = 0$ for all $p \in \mathbb{N}$. We suppose that the original system was real, that is, on \mathbb{R}^2 ; we can choose linear coordinates such that for $z = x + iy$, $\bar{z} = x - iy$ the linear part is $A = \text{diagonal}[i\lambda, -i\lambda]$. Applying the remarks above, we conclude that the normal form only contains the monomials $(z\bar{z})^p z \partial / \partial z$ and $(z\bar{z})^p \bar{z} \partial / \partial \bar{z}$. The geometric interpretation here is that these monomials are invariant for rotations around $(0, 0)$. This can also be seen on the real variant of this: the Taylor series of the (real) normalized system has the form $(\lambda + f(x^2 + y^2))(x\partial/\partial y - y\partial/\partial x) + g(x^2 + y^2)(x\partial/\partial x + y\partial/\partial y)$ and is invariant for rotations. Warning: the dynamic behavior of a formal normal form in the central manifold can be very different from that of the original vector field, since we are only looking at the formal level. A trivial example is (take $f = g = 0$ in the foregoing example) $X(x, y) = \lambda(x\partial/\partial y - y\partial/\partial x) - \exp(-1/(x^2))\partial/\partial x$, where orbits near $(0, 0)$ spiral to $(0, 0)$, whereas the normal form is just a linear rotation. This difference is due to the so-called flat terms, that is, the difference between the transformed vector field and a C^∞ -realization of its normalized Taylor series (or polynomial). In case of analyticity of X , one can ask for analyticity of the normalizing transformation ϕ . Generically, this is not the case in many situations. The precise meaning of this “genericity condition” is too elaborate to explain in this brief review article. We provide some suggestions for further reading in the next section. One could roughly say that, in the central manifold, the normal form has too much symmetry and is too poor to model more complicated dynamics of the system, which can be “hidden in the flat terms.” To quote Il'yashenko (1981): “In the theory of normal forms of analytic differential equations, divergence is the rule and convergence the exception . . .”

In many applications, we want to preserve some extra structure, such as a symplectic structure, a volume form, some symmetry, reversibility, some projection etc.; the case of a projection is important since it includes vector fields depending on a parameter. Sometimes a superposition of these structures appears (e.g., a family of volume-preserving systems). We would like that the normal-form procedure respects this structure at each step. One can often formulate this in terms of vector fields belonging to some Lie subalgebra \mathcal{L}_0 . The idea is then to use changes of variables like [1], where ξ_k is then generated by a vector field in \mathcal{L}_0 . This will guarantee that all changes of variables are “compatible” with the extra structure. Unlike the general case where we could work with monomials as in [4], we will have to consider vector fields h_k in \mathcal{L}_0 whose components are homogeneous polynomials of degree k . If this can be

done, one says that \mathcal{L}_0 respects the grading by the homogeneous polynomials. In order to fix ideas, suppose that \mathcal{L}_0 are the divergence-free planar vector fields. Note that a monomial $x^i y^j \partial / \partial x$ is not divergence free. We can instead use time mappings of homogeneous vector fields of the form $a(q+1)x^{p+1}y^q \partial / \partial x - a(p+1)x^p y^{q+1} \partial / \partial y$. Up to terms of higher order we can use the time-one map of h_k instead of $x + h_k(x)$. In case that one asks for a C^∞ -realization of the normalizing transformation, we need an extra assumption on the extra structure, that is, on \mathcal{L}_0 , called the Borel property: denote by $J_{\infty, 0}$ the set of formal series such that each truncation is the Taylor polynomial of an element of \mathcal{L}_0 . The extra assumption is: each element of $J_{\infty, 0}$ must be the Taylor series of a C^∞ vector field in \mathcal{L}_0 . It can be proved (Broer 1981) that the following structures respect the grading and satisfy the Borel property: being an r -parameter family, respecting a volume form on \mathbb{R}^n , being a Hamiltonian vector field (n even), and being reversible for a linear involution.

One could consider other types of grading of the Lie-algebras involved.

This method, using the framework of the so-called filtered Lie algebras, is explained and developed systematically in a more general and abstract context in Broer (1981).

In nonlocal bifurcations, such as near a homoclinic loop, for example, it is not enough to perform central manifold reduction near the singularity: a simplified smooth model in a full neighborhood of the singularity is often needed, for example, in order to compute Poincaré maps.

Let us start with the “purely” hyperbolic case (i.e., $\dim E_c = 0$). First we compute the formal normal form such as the above. If there are no resonances [2] then we can formally linearize the vector field X . If X is C^∞ then a classical theorem of Sternberg (1958) states that this linearization can be realized by a C^∞ change of variables (i.e., no more flat terms remaining). In case there are resonances, we must allow nonlinear terms: the resonant monomials. In this case we can also reduce C^∞ to this normal form. Using the same methods, it is also possible to reduce to a polynomial normal form, but this time using $C^k (k < \infty)$ changes of variables. More precisely, if k is a given number and if we write the vector field as $X = X_N + R_N$, where X_N is the Taylor polynomial up to order N (which can be assumed to be in normal form) and where $R_N(x) = O(|x|^{N+1})$, then for N sufficiently large there is a C^k change of variables conjugating X to X_N near 0. The number N depends on the spectrum of $A = dX(0)$. An elegant proof of these facts can be found in Il'yashenko and Yakovenko (1991). For the case when extra structure must be

preserved, see [Bonckaert \(1997\)](#), which also deals with the partially hyperbolic case ($\dim E_c \geq 1$). As already remarked above, the case of a parameter-dependent family can be regarded as a partially hyperbolic stationary point preserving this extra structure.

The question of an analytic normal form, also in the hyperbolic case, leads to convergence questions and calls upon the so-called small-divisor problems. The classical results are due to Poincaré and Siegel. Let us summarize them; they are formulated in the complex analytic setting:

Theorem 4

- (i) *If the convex hull of the spectrum of A does not contain $0 \in \mathbb{C}$ then X can locally be put into normal form by an analytic change of variables. Moreover, this normal form is polynomial.*
- (ii) *If the spectrum $\{\lambda_1, \dots, \lambda_n\}$ of A satisfies the condition that there exists $C > 0$ and $\mu > 0$ such that for any $m \in \mathbb{N}^n$ with $\sum_j m_j \geq 2$:*

$$|((\lambda_1, \dots, \lambda_n), m) - \lambda_j| \geq \frac{C}{|m|^\mu} \quad [5]$$

for $1 \leq j \leq n$ then X can be locally linearized by an analytic change of variables.

Note that case (i) contains the case where 0 is a hyperbolic source or sink. This case (i) in [Theorem 4](#) can be extended if there are parameters: if X depends analytically on a parameter $\varepsilon \in \mathbb{C}^p$ near $\varepsilon = 0$ then the change of variables is also analytic in ε ; moreover, the normal form is then a polynomial in the space variables whose coefficients are analytically dependent on the parameter ε .

For case (ii) this is surely not the case, since the condition [5] is fragile: a small distortion of the parameter generically causes resonances, be it of a high order. To fix ideas, consider $n = 2$ and suppose $\lambda_1 < 0 < \lambda_2$. By a generic but arbitrary small perturbation, we can have that the ratio of these eigenvalues becomes a negative rational number $-p/q$, which gives a resonance of the form [2] with $j = 1$ and $\alpha = (q + 1, p)$, so [5] is violated.

So analytic linearization, or even a polynomial analytic normal form, is ungeneric for families of such hyperbolic stationary points. The search for analytic normal forms, that is, simplified models, for families is still under investigation. A first simplification is obtained via the stable and unstable manifold from [Theorem 1](#), that is, the graphs of ϕ_{ss} and ϕ_{uu} . When X is analytic near 0 then these manifolds are also analytic. So, up to an analytic change of variables, we can assume that E_s and E_u are invariant, which gives a simplification of the expression of X . Moreover, there is analytic dependence on parameters.

For local diffeomorphisms there are completely similar theorems pertaining to all the cases considered above.

Concluding Remarks

The concept of central manifold can be extended to more general invariant sets (see [Chow et al. \(2000\)](#) and references therein). It can also be extended to the infinite-dimensional case and can be applied to partial differential equations ([Vanderbauwhede and Iooss 1992](#)).

Concerning the generic divergence of normalizing transformations, the reader is referred to [Broer and Takens \(1989\)](#), [Bruno \(1989\)](#), [Il'yashenko \(1981\)](#), and [Il'yashenko and Pyartli \(1991\)](#). Although the power series giving the normalizing transformation generally diverges, the study of the dynamics is often performed by truncating the normal form at a certain order. Recently, [Iooss and Lombardi \(2005\)](#) considered the question as to what an optimal truncation is. It is shown, in case $dX(0)$ is semisimple, that the order of the normal form can be optimized so that the remainder satisfies some estimate shrinking exponentially fast to zero as a function of the radius of the domain.

Concerning normal forms preserving the Hamiltonian structure, see [Birkhoff \(1966\)](#) and [Siegel and Moser \(1995\)](#) for a starting point; this is an extended subject on its own, sometimes called Birkhoff normal form, and it would require another review article.

Further simplifications of the normal form can sometimes be obtained by taking into account nonlinear terms (instead of just A) in order to obtain reductions of higher-order terms (see [Gaeta \(2002\)](#) and especially the references therein).

Applications of normal forms and central manifolds to bifurcation theory have been explained in [Dumortier \(1991\)](#).

See also: Averaging Methods; Bifurcation Theory; Dynamical Systems and Thermodynamics; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Finite Group Symmetry Breaking; Korteweg–de Vries Equation and Other Modulation Equations; Multiscale Approaches; Normal Forms and Semiclassical Approximation; Symmetry and Symmetry Breaking in Dynamical Systems.

Further Reading

Arrowsmith D and Place C (1990) *Dynamical Systems*. Cambridge: Cambridge University Press.
 Aubach B (1992) One-dimensional center manifolds are C^∞ . *Results in Mathematics* 21: 3–11.

- Birkhoff GD (1966) *Dynamical Systems*. With an addendum by Jurgen Moser. American Mathematical Society Colloquium Publications, vol. IX. Providence, RI: American Mathematical Society.
- Bonckaert P (1997) Conjugacy of vector fields respecting additional properties. *Journal of Dynamical and Control Systems* 3: 419–432.
- Bonckaert P (2000) Symmetric and reversible families of vector fields near a partially hyperbolic singularity. *Ergodic Theory and Dynamical Systems* 20: 1627–1638.
- Broer H (1981) Formal normal forms for vector fields and some consequences for bifurcations in the volume preserving case. In: *Dynamical Systems and Turbulence, Warwick 1980*, vol. 898, Lecture Notes in Mathematics. New York: Springer.
- Broer H and Takens F (1989) Formally symmetric normal forms and genericity. *Dynamics Reported. A Series in Dynamical Systems and their Applications* 2: 11–18.
- Bruno AD (1989) *Local Methods in Nonlinear Differential Equations*. New York: Springer.
- Chow S-N, Li C, and Wang D (1994) *Normal Forms and Bifurcations of Planar Vector Fields*. Cambridge: Cambridge University Press.
- Chow S-N, Liu W, and Yi Y (2000) Center manifolds for invariant sets. *Journal of Differential Equations* 168: 355–385.
- Dumortier F (1991) Local study of planar vector fields: singularities and their unfoldings. In: Van Groesen E and De Jager EM (eds.) *Structures in Dynamics, Studies in Mathematical Physics*, vol. 2, pp. 161–241. Amsterdam: Elsevier.
- Gaeta G (2002) Poincaré normal and renormalized forms. *Acta Applicandae Mathematicae* 70(1–3): 113–131 (symmetry and perturbation theory).
- Il'yashenko YS (1981) In the theory of normal forms of analytic differential equations violating the conditions of Bryuno divergence is the rule and convergence the exception. *Moscow University Mathematical Bulletin* 36(2): 11–18.
- Il'yashenko YS and Pyartli AS (1986) Materialization of resonances and divergence of normalizing series for polynomial differential equations. *Journal of Mathematical Sciences* 32(3): 300–313.
- Il'yashenko YS and Yakovenko SY (1991) Finitely smooth normal forms of local families of diffeomorphisms and vector fields. *Russian Mathematical Surveys* 46: 1–43.
- Iooss G and Lombardi E (2005) Polynomial normal forms with exponentially small remainder for analytic vector fields. *Journal of Differential Equations* 212: 1–61.
- Palis J and Takens F (1977) Topological equivalence of normally hyperbolic dynamical systems. *Topology* 16(4): 335–345.
- Siegel CL and Moser JK (1971) *Lectures on Celestial Mechanics*, (reprint 1995). Berlin: Springer.
- Sternberg S (1958) On the structure of local homeomorphisms of Euclidean n -space. II. *American Journal of Mathematics* 80: 623–631.
- Takens F (1971) Partially hyperbolic fixed points. *Topology* 10: 133–147.
- Vanderbauwhede A (1989) Center manifolds, normal forms and elementary bifurcations. In: Kirchgraber U and Walther O (eds.) *Dynamics Reported*, vol. 2, pp. 89–169. New York: Wiley.
- Vanderbauwhede A and Iooss G (1992) Center manifold theory in infinite dimensions. In: Jones CKRT *et al.* (eds.) *Dynamics Reported*, vol. 1, New Series, pp. 125–163. Berlin: Springer.

Channels in Quantum Information Theory

M Keyl, Università di Pavia, Pavia, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Consider a typical quantum system such as a string of ions in a trap. To “process” the quantum information the ions carry, we have to perform in general many steps of a quite different nature. Typical examples are: free time evolution (including unwanted but unavoidable interactions with the environment), controlled time evolution (e.g., the application of a “quantum gate” in a quantum computer), preparations and measurements. Each processing step can be described by a channel which transforms input systems into output system of a possibly different type (e.g., a measurement transforms quantum systems into classical information).

Systems, States, and Algebras

To get a unified mathematical description of systems of different physical nature, it is useful to consider

C^* -algebras (which are, in our case, always finite dimensional): quantum systems can be represented in terms of the algebra $\mathcal{B}(\mathcal{H})$ of (bounded) operators on the Hilbert space $\mathcal{H} = \mathbb{C}^d$; for classical information we have to choose the set $\mathcal{C}(X)$ of (continuous), complex-valued functions on the finite alphabet X ; and the tensor product of both $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ describes hybrid systems which are half-classical and half-quantum. Assume now that \mathcal{A} is one of these algebras. Effects (i.e., yes/no measurements on the system in question) are then described by $A \in \mathcal{A}$ satisfying $0 \leq A \leq \mathbb{1}$, states are positive, normalized linear functionals $\omega: \mathcal{A} \rightarrow \mathbb{C}$, and the probability to get the result “yes” during an A measurement on a system in the state ω is given by $\omega(A)$. Since \mathcal{A} is assumed to be finite dimensional, each state ω on $\mathcal{B}(\mathcal{H})$ is represented by a density operator ρ , that is, $\omega(A) = \text{tr}(\rho A)$. Likewise, a state ω on $\mathcal{C}(X)$ has the form $\omega(A) = \sum_x A(x)p_x$, where $(p_x)_{x \in X}$ denotes a probability distribution on X , and a state ω on $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ is described by a sequence $(\rho_x)_{x \in X}$ of positive (trace-class) operators on $\mathcal{B}(\mathcal{H})$ with $\sum_x \text{tr}(\rho_x) = 1$ such that $\omega(A) = \sum_x \text{tr}(\rho_x A_x)$. Here

we have used the fact that an element $A \in \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ can be represented in a canonical way by a sequence $(A_x)_{x \in X}$ of operators on \mathcal{H} . The set of states will be denoted in the following by $\mathcal{S}(\mathcal{A})$ and the set of effects by $\mathcal{E}(\mathcal{A})$.

Completely Positive Maps

Our aim is now to get a mathematical object which can be used to describe a channel. To this end, consider two C^* -algebras, \mathcal{A}, \mathcal{B} , describing the input and output system, respectively, and an effect $A \in \mathcal{B}$ of the output system. If we invoke first a channel which transforms \mathcal{A} systems into \mathcal{B} systems, and measure A afterwards on the output systems, we end up with a measurement of an effect $T(A)$ on the input systems. Hence, we get a map $T: \mathcal{E}(\mathcal{B}) \rightarrow \mathcal{E}(\mathcal{A})$ which completely describes the channel (note that the direction of the mapping arrow is reversed compared to the natural ordering of processing). Alternatively, we can look at the states and interpret a channel as a map $T^*: \mathcal{S}(\mathcal{A}) \rightarrow \mathcal{S}(\mathcal{B})$ which transforms \mathcal{A} systems in the state $\rho \in \mathcal{S}(\mathcal{A})$ into \mathcal{B} systems in the state $T^*(\rho)$. To distinguish between both maps, we can say that T describes the channel in the Heisenberg picture and T^* in the Schrödinger picture. On the level of the statistical interpretation, both points of view should coincide of course, that is, the probabilities $(T^*\rho)(A)$ and $\rho(TA)$ to get the result “yes” during an A measurement on \mathcal{B} systems in the state $T^*\rho$, respectively a TA measurement on \mathcal{A} systems in the state ρ , should be the same. Since $(T^*\rho)(A)$ is linear in A , we see immediately that T must be an affine map, that is, $T(\lambda_1 A_1 + \lambda_2 A_2) = \lambda_1 T(A_1) + \lambda_2 T(A_2)$ for each convex linear combination $\lambda_1 A_1 + \lambda_2 A_2$ of effects in \mathcal{B} , and this in turn implies that T can be extended naturally to a linear map, which we will identify in the following with the channel itself, that is, we say that T is the channel.

Let us now change slightly our point of view and start with a linear operator $T: \mathcal{A} \rightarrow \mathcal{B}$. To be a channel, T must map effects to effects, that is, T has to be positive: $T(A) \geq 0 \forall A \geq 0$ and bounded from above by $\mathbb{1}$, that is, $T(\mathbb{1}) \leq \mathbb{1}$. In addition, it is natural to require that two channels in parallel are again a channel. More precisely, if two channels $T: \mathcal{A}_1 \rightarrow \mathcal{B}_1$ and $S: \mathcal{A}_2 \rightarrow \mathcal{B}_2$ are given, we can consider the map $T \otimes S$ which associates to each $A \otimes B \in \mathcal{A}_1 \otimes \mathcal{A}_2$ the tensor product $T(A) \otimes S(B) \in \mathcal{B}_1 \otimes \mathcal{B}_2$. It is natural to assume that $T \otimes S$ is a channel which converts composite systems of type $\mathcal{A}_1 \otimes \mathcal{A}_2$ into $\mathcal{B}_1 \otimes \mathcal{B}_2$ systems. Hence, $S \otimes T$ should be positive as well.

Definition 1 Consider two observable algebras \mathcal{A}, \mathcal{B} and a linear map $T: \mathcal{A} \rightarrow \mathcal{B} \subset \mathcal{B}(\mathcal{H})$.

- (i) T is called positive if $T(A) \geq 0$ holds for all positive $A \in \mathcal{A}$.
- (ii) T is called completely positive (CP) if $T \otimes \text{Id}: \mathcal{A} \otimes \mathcal{B}(C^n) \rightarrow \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(C^n)$ is positive for all $n \in \mathbb{N}$. Here Id denotes the identity map on $\mathcal{B}(C^n)$.
- (iii) T is called unital if $T(\mathbb{1}) = \mathbb{1}$ holds.

Consider now the map $T^*: \mathcal{B}^* \rightarrow \mathcal{A}^*$ which is dual to T , that is, $T^*\rho(A) = \rho(TA)$ for all $\rho \in \mathcal{B}^*$ and $A \in \mathcal{A}$. It is called the Schrödinger-picture representation of the channel T , since it maps states to states provided T is unital. (Complete) positivity can be defined in the Schrödinger picture as in the Heisenberg picture, and we immediately see that T is (completely) positive iff T^* is.

It is natural to ask whether the distinction between positivity and complete positivity is really necessary, that is, whether there are positive maps which are not CP. If at least one of the algebras \mathcal{A} or \mathcal{B} is classical, the answer is no: each positive map is CP in this case. If both algebras are quantum however, complete positivity is not implied by positivity alone. The most prominent example for this fact is the transposition map.

If item (ii) holds only for a fixed $n \in \mathbb{N}$, the map T is called n -positive. This is obviously a weaker condition than complete positivity. However, n -positivity implies m -positivity for all $m \leq n$, and for $\mathcal{A} = \mathcal{B}(C^d)$ complete positivity is implied by n -positivity, provided $n \geq d$ holds.

Let us consider now the question whether a channel should be unital or not. We have already mentioned that $T(\mathbb{1}) \leq \mathbb{1}$ must hold since effects should be mapped to effects. If $T(\mathbb{1})$ is not equal to $\mathbb{1}$, we get $\rho(T\mathbb{1}) = T^*\rho(\mathbb{1}) < 1$ for the probability to measure the effect $\mathbb{1}$ on systems in the state $T^*\rho$, but this is impossible for channels which produce an output with certainty, because $\mathbb{1}$ is the effect which is always true. In other words, if a CP map is not unital, it describes a channel which sometimes produces no output at all and $T(\mathbb{1})$ is the effect which measures whether we have got an output. We will assume henceforth that channels are unital if nothing else is explicitly stated.

Quantum Channels

In this section we will discuss some basic properties of CP maps which transform quantum systems into quantum systems, in particular the Stinespring theorem, which constitutes the most important structural result. For a more detailed presentation, including generalizations to more general input/

output algebras the reader should consult the textbook by Paulsen (2002).

The Stinespring Theorem

Hence consider channels between quantum systems, i.e., $\mathcal{A}=\mathcal{B}(\mathcal{H}_1)$ and $\mathcal{B}=\mathcal{B}(\mathcal{H}_2)$. A fairly simple example (not necessarily unital) is given in terms of an operator $V:\mathcal{H}_1\rightarrow\mathcal{H}_2$ by $\mathcal{B}(\mathcal{H}_1)\ni A\mapsto VAV^*\in\mathcal{B}(\mathcal{H}_2)$. A second example is the restriction to a subsystem, which is given in the Heisenberg picture by $\mathcal{B}(\mathcal{H})\ni A\mapsto A\otimes 1_{\mathcal{K}}\in\mathcal{B}(\mathcal{H}\otimes\mathcal{K})$. Finally the composition $S\circ T=ST$ of two channels is again a channel. The following theorem says that each channel can be represented as a composition of these two examples [7].

Theorem 2 (Stinespring dilation theorem). *Every completely positive map $T:\mathcal{B}(\mathcal{H}_1)\rightarrow\mathcal{B}(\mathcal{H}_2)$ has the form*

$$T(A) = V^*(A \otimes 1_{\mathcal{K}})V \tag{1}$$

with an additional Hilbert space \mathcal{K} and an operator $V:\mathcal{H}_2\rightarrow\mathcal{H}_1\otimes\mathcal{K}$. Both (i.e., \mathcal{K} and V) can be chosen such that the span of all $(A\otimes 1)V\phi$ with $A\in\mathcal{B}(\mathcal{H}_1)$ and $\phi\in\mathcal{H}_2$ is dense in $\mathcal{H}_1\otimes\mathcal{K}$. This particular decomposition is unique (up to unitary equivalence) and is called the minimal decomposition.

By introducing a family $|\chi_j\rangle\langle\chi_j|$ of one-dimensional projectors with $\sum_j|\chi_j\rangle\langle\chi_j|=1$, we can define the ‘‘Kraus operators’’ $\langle\psi, V_j\phi\rangle=\langle\psi\otimes\chi_j, V\phi\rangle$. In terms of these, we can rewrite eqn [1] in the following form (Kraus 1983):

Corollary 3 (Kraus form). *Every CP map $T:\mathcal{B}(\mathcal{H}_1)\rightarrow\mathcal{B}(\mathcal{H}_2)$ can be written in the form*

$$T(A) = \sum_{j=1}^N V_j^* A V_j \tag{2}$$

with operators $V_j:\mathcal{H}_2\rightarrow\mathcal{H}_1$.

To get a third representation of channels, consider the Stinespring form [1] of T and a vector $\psi\in\mathcal{K}$ such that $U(\phi\otimes\psi)=V(\phi)$ can be extended to a unitary map $U:\mathcal{H}\otimes\mathcal{K}\rightarrow\mathcal{H}\otimes\mathcal{K}$. It is then easy to see that the dual T^* of T can be written as:

Corollary 4 (Ancilla form). *Assume that $T:\mathcal{B}(\mathcal{H})\rightarrow\mathcal{B}(\mathcal{H})$ is a channel. Then there is a Hilbert space \mathcal{K} , a pure state ρ_0 , and a unitary map $U:\mathcal{H}\otimes\mathcal{K}\rightarrow\mathcal{H}\otimes\mathcal{K}$ such that*

$$T^*(\rho) = \text{tr}_{\mathcal{K}}(U(\rho\otimes\rho_0)U^*) \tag{3}$$

holds.

This representation of a channel has a (seemingly) very nice physical interpretation, because we can look at eqn [3] as the unitary interaction of the system with an unobservable environment, which is initially in the state ρ_0 . The problem, however, is that there is a great arbitrariness in the choice of U and ρ_0 . This is the weakness of the ancilla form compared to the Stinespring representation.

Finally, let us state a related result. It characterizes all decompositions of a given completely positive map into completely positive summands. By analogy with results for states on abelian algebras (i.e., probability measures), we will call it a Radon–Nikodym theorem (see Arveson (1969) for a proof).

Theorem 5 (Radon–Nikodym theorem). *Let $T_x:\mathcal{B}(\mathcal{H}_1)\rightarrow\mathcal{B}(\mathcal{H}_2), x\in X$ be a family of CP maps and let $V:\mathcal{H}_2\rightarrow\mathcal{H}_1\otimes\mathcal{K}$ be the Stinespring operator of $\bar{T}=\sum_x T_x$; then there are uniquely determined positive operators F_x in $\mathcal{B}(\mathcal{K})$ with $\sum_x F_x=1$ and*

$$T_x(A) = V^*(A \otimes F_x)V \tag{4}$$

The Jamiołkowski Isomorphism

The subject of this section is a relation between CP maps and states of bipartite systems, first discovered by Jamiołkowski (1972), and which is very useful in translating properties of bipartite systems into properties of positive maps and vice versa.

The idea is based on the following setup. Alice and Bob share a bipartite system in a maximally entangled state

$$\chi = \frac{1}{\sqrt{d}} \sum_{\alpha=1}^d e_{\alpha} \otimes e_{\alpha} \in \mathcal{H} \otimes \mathcal{H} \tag{5}$$

(where e_1, \dots, e_d denote an orthonormal basis of \mathcal{H}). Alice applies to her subsystem a channel $T:\mathcal{B}(\mathcal{H})\rightarrow\mathcal{B}(\mathcal{H}')$ while Bob does nothing. At the end of the processing, the overall system ends up in a state

$$R_T = (T \otimes \text{Id})|\chi\rangle\langle\chi| \tag{6}$$

Mathematically, eqn [6] makes sense if T is only linear but not necessarily positive or CP (but then R_T is not positive either). If we denote the space of all linear maps from $\mathcal{B}(\mathcal{H})$ into $\mathcal{B}(\mathcal{H}')$ by \mathcal{L} , we get a map

$$\mathcal{L} \ni T \mapsto R_T \in \mathcal{B}(\mathcal{K} \otimes \mathcal{H}) \tag{7}$$

which is easily shown to be linear (i.e., $R_{\mu T+\lambda S}=\mu R_T+\lambda R_S$ for all $\lambda, \mu\in\mathbb{C}$ and all $T, S\in\mathcal{L}$). Furthermore, this map is bijective, hence a linear isomorphism.

Theorem 6 *The map defined in eqns [7] and [6] is a linear isomorphism. The inverse map is given by*

$$\mathcal{B}(\mathcal{H} \otimes \mathcal{H}') \ni \rho \mapsto T_\rho \in \mathcal{L} \quad [8]$$

with

$$\langle e'_\mu, T_\rho(\sigma)e'_\nu \rangle = d \operatorname{tr} \left(\rho(|e'_\nu\rangle\langle e'_\mu| \otimes \sigma^T) \right) \quad [9]$$

where $e'_1, \dots, e'_d \in \mathcal{H}'$ denote an (arbitrary) orthonormal basis of \mathcal{H}' and the transposition of σ is defined with respect to the basis $e_\alpha, \alpha = 1, \dots, d$ used to define χ in [5].

From the definition of R_T in eqn [6], it is obvious that R_T is positive, if T is CP. To see that the converse is also true is not as trivial (because a transposition is involved), but it requires only a short calculation, which is omitted here. Hence, we get:

Corollary 7 *The operator R_T is positive, iff the map T is CP.*

Examples

Let us return now to the general case (i.e., arbitrary input and output algebras) and discuss several examples.

Channels Under Symmetry

It is often useful to consider channels with special symmetry properties. To be more precise, consider a group G and two unitary representations π_1, π_2 on the Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 , respectively. A channel $T: \mathcal{B}(\mathcal{H}_1) \rightarrow \mathcal{B}(\mathcal{H}_2)$ is called covariant (with respect to π_1 and π_2) if

$$T[\pi_1(U)A\pi_1(U)^*] = \pi_2(U)T[A]\pi_2(U)^* \quad \forall A \in \mathcal{B}(\mathcal{H}_1) \forall U \in G \quad [10]$$

holds. The general structure of covariant channels is governed by a fairly powerful variant of Stinesprings theorem (Keyl and Werner 1999).

Theorem 8 *Let G be a group with finite-dimensional unitary representations $\pi_j: G \rightarrow \mathcal{U}(\mathcal{H}_j)$ and $T: \mathcal{B}(\mathcal{H}_1) \rightarrow \mathcal{B}(\mathcal{H}_2)$ a π_1, π_2 -covariant channel.*

- (i) *Then there is a finite-dimensional unitary representation $\tilde{\pi}: G \rightarrow \mathcal{U}(\mathcal{K})$ and an operator $V: \mathcal{H}_2 \rightarrow \mathcal{H}_1 \otimes \mathcal{K}$ with $V\pi_2(U) = \pi_1(U) \otimes \tilde{\pi}(U)V$ and $T(A) = V^*A \otimes \mathbb{1}_V$.*
- (ii) *If $T = \sum_\alpha T^\alpha$ is a decomposition of T in CP and covariant summands, there is a decomposition $\mathbb{1} = \sum_\alpha F^\alpha$ of the identity operator on \mathcal{K} into positive operators $F^\alpha \in \mathcal{B}(\mathcal{K})$ with $[F^\alpha, \tilde{\pi}(g)] = 0$ such that $T^\alpha(X) = V^*(X \otimes F^\alpha)V$.*

The most prominent examples of covariant channels arise with $\mathcal{H}_1 = \mathcal{H}_2 = \mathbb{C}^d, G = \mathcal{U}(d)$ and $\pi_1(U) = \pi_2(U) = U$. All channels of this type are of the form

$$T(A) = (1 - \vartheta)A + \vartheta d^{-1} \operatorname{tr}(A) \mathbb{1} \quad \text{with } \vartheta \in [0, d^2/(d^2 - 1)] \quad [11]$$

and are known as “depolarizing channels.” They often serve as a standard model for noise. Two particular cases are the ideal channel arising with $\vartheta = 0$, and the completely depolarizing channel ($\vartheta = 1$) which erases all information. If we choose $\pi_2(U) = \bar{U}$ (where the bar denotes complex conjugate) instead of $\pi_2(U) = U$, we get

$$T(A) = \frac{\vartheta}{d+1} [\operatorname{tr}(A)\mathbb{1} + A^T] + \frac{1-\vartheta}{d-1} [\operatorname{tr}(A)\mathbb{1} - A^T], \quad \vartheta \in [0, 1] \quad [12]$$

If we map these channels to states of bipartite systems (using the Jamiołkowski isomorphism from the last section), we get “Isotropic states” from eqn [11] and “Werner states” from [12].

Classical Channels

The classical analog to a quantum operation is a channel $T: \mathcal{C}(X) \rightarrow \mathcal{C}(Y)$ which describes the transmission or manipulation of classical information. As already mentioned in the subsection “Completely positive maps,” positivity and complete positivity are equivalent in this case. Hence, we have to assume only that T is positive and unital. Obviously, T is characterized by its matrix elements $T_{xy} = \delta_y(Te_x)$, where $\delta_y \in \mathcal{C}^*(Y)$ denotes the Dirac measure at $y \in Y$ and $e_x \in \mathcal{C}(X)$ is the canonical basis in $\mathcal{C}(X)$. More precisely, δ_y and e_x denote, respectively, the probability distribution and the function on X , given by

$$\delta_y = (\delta_{xy})_{x \in X} \quad \text{and} \quad e_x(y) = \delta_{xy} \quad [13]$$

We will keep this notation up to the end of this article. Positivity and normalization of T imply that $0 \leq T_{xy} \leq 1$ and

$$1 = \delta_y(\mathbb{1}) = \delta_y(T(\mathbb{1})) = \delta_y \left[T \left(\sum_x e_x \right) \right] = \sum_x T_{xy} \quad [14]$$

holds. Hence the family $(T_{xy})_{x \in X}$ is a probability distribution on X and T_{xy} is, therefore, the transition probability to get the information $x \in X$ at the output side of the channel if $y \in Y$ was sent.

Observables

Let us consider now a channel which transforms quantum information $\mathcal{B}(\mathcal{H})$ into classical information $\mathcal{C}(X)$. Since positivity and complete positivity are again equivalent, we just have to look at a positive and unital map $E: \mathcal{C}(X) \rightarrow \mathcal{B}(\mathcal{H})$. With the canonical basis $e_x, x \in X$, of $\mathcal{C}(X)$, we get a family $E_x = E(e_x), x \in X$, of positive operators $E_x \in \mathcal{B}(\mathcal{H})$ with $\sum_{x \in X} E_x = \mathbb{1}$. Hence, the E_x form a positive operator valued (POV) measure, i.e., an observable. If, on the other hand, a POV measure $E_x \in \mathcal{B}(\mathcal{H}), x \in X$, is given, we can define a quantum-to-classical channel $E: \mathcal{C}(X) \rightarrow \mathcal{B}(\mathcal{H})$ by

$$E(f) = \sum_{x \in X} f(x) E_x \quad [15]$$

This shows that the observable $E_x, x \in X$, and the channel E can be identified.

Preparations

Let us now exchange the role of $\mathcal{C}(X)$ and $\mathcal{B}(\mathcal{H})$; in other words, let us consider a channel $R: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{C}(X)$ with a classical input and a quantum output algebra. In the Schrödinger picture, we get a family of density matrices $\rho_x := R^*(\delta_x) \in \mathcal{B}^*(\mathcal{H}), x \in X$, where $\delta_x \in \mathcal{C}^*(X)$ denotes again the Dirac measure on X . Hence, we get a parameter-dependent preparation that can be used to encode the classical information $x \in X$ into the quantum information $\rho_x \in \mathcal{B}^*(\mathcal{H})$.

Instruments

An observable describes only the statistics of measuring results, but does not contain information about the state of the system after the measurement. To get a description which fills this gap, we have to consider channels which operate on quantum systems and produce hybrid systems as output, that is, $T: \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \rightarrow \mathcal{B}(\mathcal{K})$. Following Davies (1976), we will call such an object an instrument. From T we can derive the subchannel

$$\mathcal{C}(X) \ni f \mapsto T(\mathbb{1} \otimes f) \in \mathcal{B}(\mathcal{K}) \quad [16]$$

which is the observable measured by T , that is, $\text{tr}(T(\mathbb{1} \otimes e_x)\rho)$ is the probability to measure $x \in X$ on systems in the state ρ . On the other hand, we get for each $x \in X$ a quantum channel (which is not unital)

$$\mathcal{B}(\mathcal{H}) \ni A \mapsto T_x(A) = T(A \otimes e_x) \in \mathcal{B}(\mathcal{K}) \quad [17]$$

It describes the operation performed by the instrument T if $x \in X$ was measured. More precisely, if a measurement on systems in the state ρ gives the result $x \in X$, we get (up to normalization) the state $T_x^*(\rho)$ after the measurement, while

$$\text{tr}(T_x^*(\rho)) = \text{tr}(T_x^*(\rho)\mathbb{1}) = \text{tr}(\rho T(\mathbb{1} \otimes e_x)) \quad [18]$$

is (again) the probability to measure $x \in X$ on ρ . The instrument T can be expressed in terms of the operations T_x by

$$T(A \otimes f) = \sum_x f(x) T_x(A) \quad [19]$$

Hence, we can identify T with the family $T_x, x \in X$. Finally, we can consider the second marginal of T

$$\mathcal{B}(\mathcal{H}) \ni A \mapsto T(A \otimes \mathbb{1}) = \sum_{x \in X} T_x(A) \in \mathcal{B}(\mathcal{K}) \quad [20]$$

It describes the operation we get if the outcome of the measurement is ignored.

The best-known example of an instrument is a von Neumann–Lüders measurement associated with a PV measure given by family of projections $E_x, x = 1, \dots, d$; for example, the eigenprojections of a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})$. It is defined as the channel

$$T: \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \rightarrow \mathcal{B}(\mathcal{H})$$

with $X = \{1, \dots, d\}$ and $T_x(A) = E_x A E_x$ [21]

Hence, we get the final state $\text{tr}(E_x \rho)^{-1} E_x \rho E_x$ if we measure the value $x \in X$ on systems initially in the state ρ – this is well known from quantum mechanics.

Parameter-Dependent Operations

Let us change now the role of $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ and $\mathcal{B}(\mathcal{K})$; in other words, consider a channel $T: \mathcal{B}(\mathcal{K}) \rightarrow \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ with hybrid input and quantum output. It describes a device which changes the state of a system depending on the additional classical information. As for an instrument, T decomposes into a family of (unital!) channels $T_x: \mathcal{B}(\mathcal{K}) \rightarrow \mathcal{B}(\mathcal{H})$ such that we get $T^*(\rho \otimes p) = \sum_x p_x T_x^*(\rho)$ in the Schrödinger picture. Physically, T describes a parameter-dependent operation: depending on the classical information $x \in X$, the quantum information $\rho \in \mathcal{B}(\mathcal{K})$ is transformed by the operation T_x .

Finally, we can consider a channel $T: \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \rightarrow \mathcal{B}(\mathcal{K}) \otimes \mathcal{C}(Y)$ with hybrid input and output to get a parameter-dependent instrument: similarly to the above discussion, we can define a family of instruments $T_y: \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \rightarrow \mathcal{B}(\mathcal{K}), y \in Y$, by the equation $T^*(\rho \otimes p) = \sum_y p_y T_y^*(\rho)$. Physically, T describes the following device: it receives the classical information $y \in Y$ and a quantum system in the state $\rho \in \mathcal{B}(\mathcal{K})$ as input. Depending on y , a measurement with the instrument T_y is performed, which in turn produces the measuring value $x \in X$ and leaves the quantum system in the state (up to normalization) $T_{y,x}^*(\rho)$; with $T_{y,x}$ given as in eqn [17] by $T_{y,x}(A) = T_y(A \otimes e_x)$.

See also: Capacities Enhanced by Entanglement; Capacity for Quantum Information; Entanglement; Optimal Cloning of Quantum States; Positive Maps on C^* -Algebras; Quantum Channels: Classical Capacity; Quantum Dynamical Semigroups; Quantum Entropy; Quantum Spin Systems; Source Coding in Quantum Information Theory.

Further Reading

Arveson W (1969) Subalgebras of C^* -algebras. *Acta Mathematica* 123: 141–224.

Davies EB (1976) *Quantum Theory of Open Systems*. London: Academic Press.

Jamiolkowski A (1972) Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics* 3: 275–278.

Keyl M and Werner RF (1999) Optimal cloning of pure states, testing single clones. *Journal of Mathematical Physics* 40: 3283–3299.

Kraus K (1983) *States Effects and Operations*. Berlin: Springer.

Paulsen VI (2002) *Completely Bounded Maps and Dilations*. Cambridge: Cambridge University Press.

Stinespring WF (1955) Positive functions on C^* -algebras. *Proceedings of the American Mathematical Society* 6: 211–216.

Chaos and Attractors

R Gilmore, Drexel University, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Chaos is a type of behavior that can be exhibited by a large class of physical systems and their mathematical models. These systems are deterministic. They are modeled by sets of coupled nonlinear ordinary differential equations (ODEs):

$$\dot{x}_i = \frac{dx_i}{dt} = f_i(x; c) \quad [1]$$

called dynamical systems. The coordinates x designate points in a state space or phase space. Typically, $x \in R^n$ or some n -dimensional manifold for some $n \geq 3$, and $c \in R^k$ are called control parameters. They describe parameters that can be controlled in physical systems, such as pumping rates in lasers or flow rates in chemical mixing reactions. The most important mathematical property of dynamical systems is the uniqueness theorem, which states that there is a unique trajectory through every point at which $f(x; c)$ is continuous and Lipschitz and $f(x; c) \neq 0$. In particular, two distinct periodic orbits cannot have any points in common.

The properties of dynamical systems are governed, in lowest order, by the number, stability, and distribution of their fixed points, defined by $\dot{x}_i = f_i(x; c) = 0$. It can happen that a dynamical system has no stable fixed points and no stable limit cycles ($x(t) = x(t + T)$, some $T > 0$, all t). In such cases, if the solution is bounded and recurrent but not periodic, it represents an unfamiliar type of attractor. If the system exhibits “sensitivity to initial conditions” ($|x(t) - y(t)| \sim e^{\lambda t} |x(0) - y(0)|$ for $|x(0) - y(0)| = \epsilon$ and $\lambda > 0$ for most $x(0)$), the solution set is called a “chaotic attractor.” If the

attractor has fractal structure, it is called a “strange attractor.”

Tools to study strange attractors have been developed that depend on three types of mathematics: geometry, dynamics, and topology.

Geometric tools attempt to study the metric relations among points in a strange attractor. These include a spectrum of fractal dimensions. These real numbers are difficult to compute, require very long, very clean data sets, provide a number without error estimates for which there is no underlying statistical theory, and provide very little information about the attractor.

Dynamical tools include estimation of Lyapunov exponents and a Lyapunov dimension. They include globally averaged exponents and local Lyapunov exponents. These are eigenvalues related to the different stretching ($\lambda > 0$) and squeezing ($\lambda < 0$) eigendirections in the phase space. To each globally averaged Lyapunov exponent λ_i , $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, there corresponds a “partial dimension” ϵ_i , $0 \leq \epsilon_i \leq 1$, with $\epsilon_i = 1$ if $\lambda_i \geq 0$. The Lyapunov dimension is the sum of the partial dimensions $d_L = \sum_{i=1}^n \epsilon_i$. That the partial dimension $\epsilon_i = 1$ for $\lambda_i \geq 0$ indicates that the flow is smooth in the stretching ($\lambda_i > 0$) and flow directions and fractal in the squeezing ($\lambda_i < 0$) directions with $\epsilon_i < 1$. Dynamical indices provide some useful information about a strange attractor. In particular, they can be used to estimate some fractal properties of a strange attractor, but not vice versa.

Topological tools are very powerful for a restricted class of dynamical systems. These are dynamical systems in three dimensions ($n = 3$). For such systems there are three Lyapunov exponents $\lambda_1 > \lambda_2 > \lambda_3$, with $\lambda_1 > 0$ describing the stretching direction and responsible for “sensitivity to initial conditions,” $\lambda_2 = 0$ describing the direction of the flow, and $\lambda_3 < 0$ describing the squeezing direction

and responsible for “recurrence.” Strange attractors are generated by dissipative dynamical systems, which satisfy the additional condition $\lambda_1 + \lambda_2 + \lambda_3 < 0$. For such attractors, $\epsilon_1 = \epsilon_2 = 1$ and $\epsilon_3 = \lambda_1/|\lambda_3|$ by the Kaplan–Yorke conjecture, so that $d_L = 2 + \epsilon_3 = 2 + \lambda_1/|\lambda_3|$.

A number of tools from classical topology have been exploited to probe the structure of strange attractors in three dimensions. These include the Gauss linking number, the Euler characteristic, the Poincaré–Hopf index theorem, and braid theory. More recent topological contributions include several definitions for entropy, the development of a theory for knot holders or braid holders (also called branched manifolds), the Birman–Williams theorem for these objects, and relative rotation rates, a topological index for individual periodic orbits and orbit pairs.

Three-dimensional strange attractors are remarkably well understood; those in higher dimensions are not. As a result, the description that follows is largely restricted to strange attractors with $d_L < 3$ that exist in R^3 or other three-dimensional manifolds (e.g., $R^2 \times S^1$). The obstacle to progress in higher dimensions is the lack of a higher-dimensional analog of the Gauss linking number for orbit pairs in R^3 .

Overview

The program described below has two objectives:

1. classify the global topological structure of strange attractors in R^3 ; and
2. determine the “perestroikas” (changes) that such attractors can undergo as experimental conditions or control parameters change.

Four levels of structure are required to complete this program. Each is topological and discretely quantifiable. This provides a beautiful interaction between a rigidity of structure, demanded by topological constraints, and freedom within this rigidity. These four levels of structure are:

1. basis sets of orbits,
2. branched manifolds or knot holders,
3. bounding tori, and
4. embeddings of bounding tori.

Branched Manifolds: Stretching and Squeezing

A strange attractor is generated by the repetition of two mechanisms: stretching and squeezing. Stretching occurs in the directions identified by the positive

Lyapunov exponents and squeezing occurs in the directions identified by the negative Lyapunov exponents. In R^3 there is one stretching direction and one squeezing direction.

A simple stretch-and-squeeze mechanism that nature appears to be very fond of is illustrated in **Figure 1**. In this illustration, a cube of initial conditions at (a) is advected by the flow in a short time to (b). During this process, the cube is deformed by being stretched ($\lambda_1 > 0$). It also shrinks in a transverse direction ($\lambda_3 < 0$). During the initial phase of this deformation, two nearby points typically separate exponentially in time. If they were to continue to separate exponentially for all times, the invariant set would not be bounded. Therefore, this separation cannot continue indefinitely, and in fact it must somehow reverse itself after some time because the motion is recurrent. The mechanism shown in **Figure 1** involves folding, which begins between (b) and (c) and continues through to (d). Squeezing occurs where points from distant parts of the attractor approach each other exponentially, as at (d). Finally, the cube, shown deformed at (d), returns to the neighborhood of initial conditions (a). This process repeats itself and builds up the strange attractor. As can be inferred from this figure, the strange attractor constructed by the repetitive process is smooth in the expanding (λ_1) and flow ($\lambda_2 = 0$) directions but fractal in the squeezing (λ_3) direction. The attractor’s fractal dimension is $\epsilon_1 + \epsilon_2 + \epsilon_3 = 2 + \epsilon_3 = 2 + \lambda_1/|\lambda_3|$.

Figure 1 summarizes the boundedness and recurrence conditions that were introduced to define strange attractors, and illustrates one stretching and squeezing mechanism that occurs repetitively to build up the fractal structure of the strange attractor

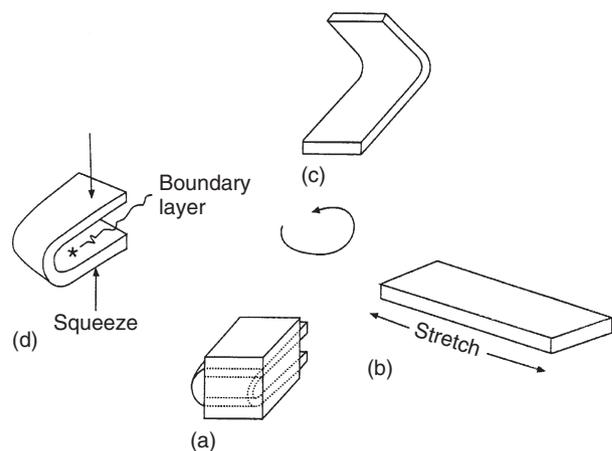


Figure 1 A common stretch-and-fold mechanism generates many experimentally observed strange attractors. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

and to organize all the (unstable) periodic orbits in it in a unique way. The particular mechanism shown in Figure 1 is called a stretch-and-fold mechanism. Other mechanisms involve stretch and roll, and tear and squeeze.

The stretch-and-squeeze mechanisms are well summarized by the cartoons shown in Figure 2. On the left, a cube of initial conditions (top) is deformed under the flow. The flow is downward. Stretching occurs in one direction (horizontal) and shrinking occurs in a transverse direction (perpendicular to the page). In the limit of extreme shrinking ($\lambda_3 \rightarrow -\infty$), the dynamics of the stretching part of the flow is represented by the two-dimensional surface shown on the bottom left. This surface fails to be a manifold because of the singularity, called a splitting point. This singularity represents an initial condition that flows to an unstable fixed point with at least one stable direction. On the right (squeezing), two distant cubes of initial conditions (top) are deformed and brought to each other's proximity under the flow (middle). In the limit of extreme dissipation, two two-dimensional surfaces representing inflows are joined at a branch line to a single surface representing an outflow. This surface fails to be a manifold because of the branch line, which is a singularity of a different kind. Points below the branch line in this representation of the flow (on the

outflow side of the branch line) have two preimages above the branch line, one in each inflow sheet. This structure generates positive entropy.

A beautiful theorem of Birman and Williams justifies the use of the two cartoons shown at the bottom of Figure 2 to characterize strange attractors in R^3 . As preparation for the theorem, Birman and Williams introduced an important identification for the nongeneric or atypical points that “are not sensitive to initial conditions”

$$x \sim y \text{ if } |x(t) - y(t)| \xrightarrow{t \rightarrow \infty} 0 \quad [2]$$

That is, two points in a strange attractor are identified if they have asymptotically the same future. In practice, this amounts to projecting the flow down along the stable ($\lambda_3 < 0$) direction onto a two-dimensional surface described by the stretching ($\lambda_1 > 0$) and the flow ($\lambda_2 = 0$) directions. This surface is not a manifold because of lower-dimensional singularities: splitting points and branch lines. The two-dimensional surface has many names, for example, knot holder (because it holds the periodic orbits that exist in abundance in strange attractors), braid holders, templates, branched manifolds. The flow, restricted to this surface, is called a semiflow. Under the semiflow, points in the branched manifold have a unique future but do not have a unique past. The degree of nonuniqueness is measured by the topological entropy of the dynamical system. The Birman–Williams theorem is:

Theorem Assume that a flow Φ_t

- (i) on R^3 is dissipative ($\lambda_1 > 0, \lambda_2 = 0, \lambda_3 < 0$ and $\lambda_1 + \lambda_2 + \lambda_3 < 0$);
- (ii) generates a hyperbolic strange attractor (the eigenvectors of the local Lyapunov exponents $\lambda_1, \lambda_2, \lambda_3$ span everywhere on the attractor).

Then the projection [2] maps the strange attractor SA to a branched manifold BM and the flow Φ_t on SA to a semiflow $\hat{\Phi}_t$ on BM in R^3 . The periodic orbits in SA under Φ_t correspond 1:1 with the periodic orbits in BM under $\hat{\Phi}_t$ with perhaps one or two specified exceptions. On any finite subset of orbits the correspondence can be taken via isotopy.

The beauty of this theorem is that it guarantees that a flow Φ_t that generates a (fractal) strange attractor SA can be continuously deformed to a new flow $\hat{\Phi}_t$ on a simple two-dimensional structure BM . During this deformation, periodic orbits are neither created nor destroyed. The uniqueness theorem for ODEs is satisfied during the deformation, so orbit segments do not pass through each other. As a result, the topological organization of all the

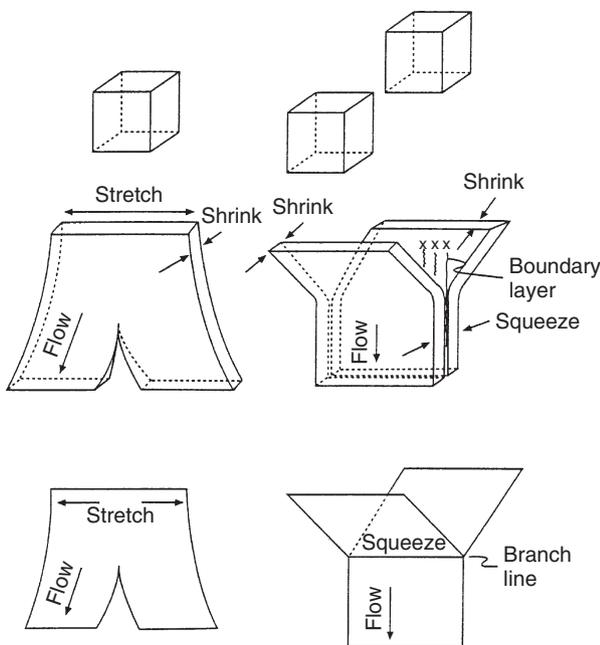


Figure 2 Left: The stretch mechanism is modeled by a two-dimensional surface with a splitting point singularity. Right: The squeeze mechanism is modeled by a two-dimensional surface with a branch line singularity. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

unstable periodic orbits in the strange attractor is the same as the topological organization of all the unstable periodic orbits in the branched manifold. In fact, the branched manifold (knot holder) defines the topological organization of all the unstable periodic orbits that it supports. Topological organization is defined by the Gauss linking number and the relative rotation rates, another braid index.

The significance of this theorem is that strange attractors can be characterized – in fact classified – by their branched manifolds. Figure 3 shows a branched manifold “for a figure-8 knot” as well as the figure-8 knot itself (dark curve). If a constant current is sent through a conducting wire tied into the shape of a figure-8 knot, a discrete countable set of magnetic field lines will be closed. These closed field lines can be deformed onto the two-dimensional surface shown in Figure 3. Each of the eight branches of this branched manifold can be named. One way to do this specifies the two branch lines that are joined by the branch in the sense of the flow (e.g., $(a\alpha)$ and $(\beta\alpha)$ (but not $(a\beta)$). Every closed field line can be labeled by a symbol sequence that is

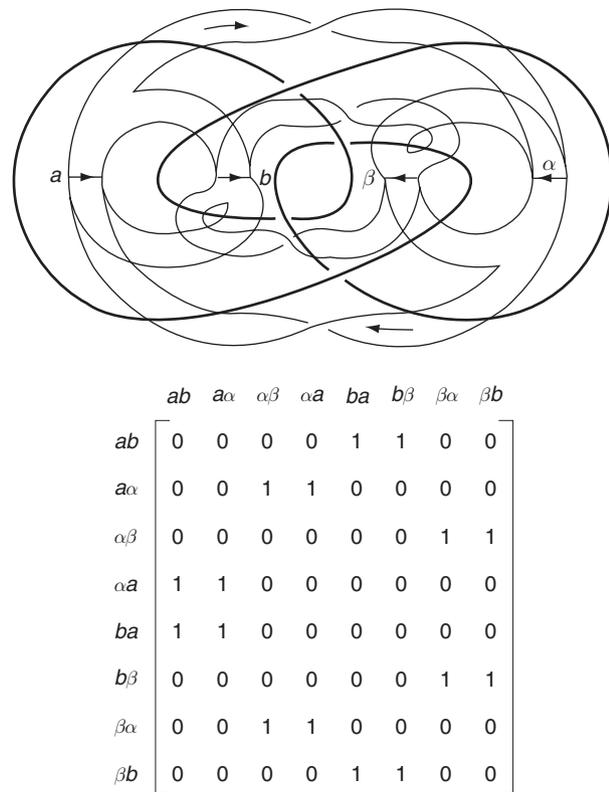


Figure 3 Figure-8 knot (dark curve) and the figure-8 branched manifold. Transition matrix for the eight branches of the figure-8 branched manifold is also shown. Flow direction is shown by arrows. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

unique up to cyclic permutation. This symbol sequence provides a symbolic name for the orbit. For example, $(a\alpha)(\alpha\beta)(\beta b)(ba)$ is a period-4 orbit. The structure of a branched manifold is determined in part by a transition matrix T . The matrix element T_{ij} is 1 if the transition from branch i to branch j is allowed, 0 otherwise. The transition matrix for the figure-8 branched manifold is shown in Figure 3.

The Birman–Williams theorem is stronger than its statement suggests. More systems satisfy the statement of the theorem than do the assumptions of the theorem. The figure-8 knot, and its attendant magnetic field, is not dissipative – in fact, it is not even a dynamical system, yet the closed loops can be isotoped to the figure-8 knot holder. There are other ways in which the Birman–Williams theorem is stronger than its statement suggests.

It is apparent from Figure 3 that the figure-8 branched manifold can be built up Lego[®] fashion from the two basic building blocks shown in Figure 2. This is more generally true. Every branched manifold can be built up, Lego[®] fashion, from the stretch (with a splitting point singularity) and the squeeze (with a branch line singularity) building blocks, subject to the following two conditions:

1. outputs flow to inputs and
2. there are no free ends.

The figure-8 branched manifold is built up from four stretch and four squeeze building blocks. As a result, there are eight branches and four branch lines.

Two often-studied strange attractors are shown in Figures 4 and 5. Figure 4 shows the details of the Rössler dynamical system. A similar spectrum of features is shown in Figure 5 for the Lorenz equations. The knot holder in Figure 5e is obtained from the caricature in Figure 5d by twisting the right-hand lobe by π radians.

Branched manifolds can be used to characterize all three-dimensional strange attractors. Branched manifolds that classify the strange attractors generated by four familiar sets of equations (for some control parameter values) are shown in Figure 6. The sets of equations, and one set of parameter values that generate strange attractors, are presented in Table 1.

The beauty of this topological classification of strange attractors is that it is apparent, just by inspection, that there is no smooth change of variables that will map any of these systems to any of the others for the parameter values shown.

Branched manifolds can be described algebraically. In Figure 7 we provide the algebraic

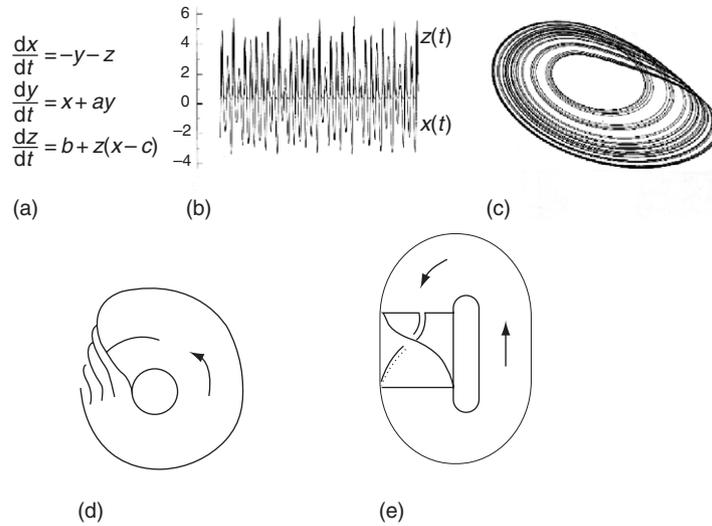


Figure 4 The Rössler dynamical system. (a) Rössler equations. (b) Time series $z(t)$ and $x(t)$ generated by these equations, and (c) projection of the strange attractor onto the x - y plane. (d) Caricature of the flow and (e) knot holder derived directly from the caricature. Control parameter values $(a, b, c) = (2.0, 4.0, 0.398)$. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

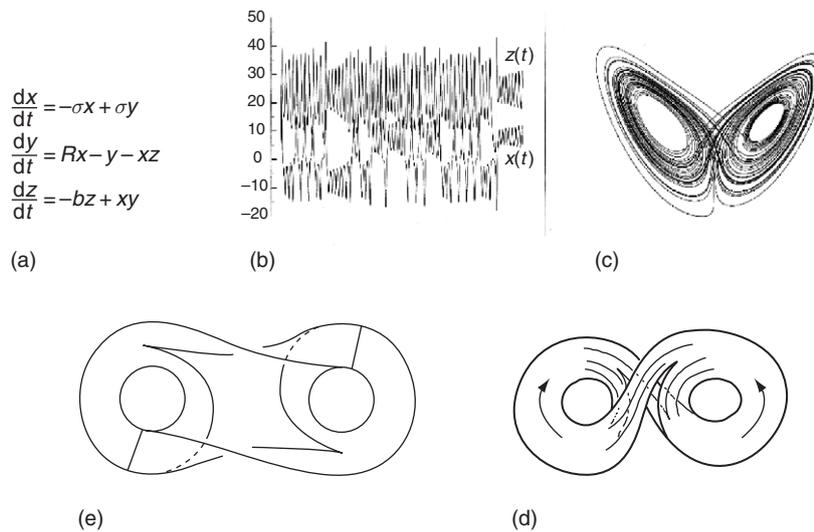


Figure 5 (a) Lorenz equations. (b) Time series $x(t)$ and $z(t)$ generated by these equations, and (c) projection of the strange attractor onto the x - y plane. (d) Caricature of the flow and (e) knot holder derived directly from the caricature by rotating the right-hand lobe by π radians. Control parameter values $(R, \sigma, b) = (26.0, 10.0, 8/3)$. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

description of two branched manifolds. **Figure 7a** shows the branched manifold that describes experimental data generated by many physical systems. The mechanism is a simple stretch-and-fold deformation with zero global torsion that generates a typical Smale horseshoe. There are two branches. The diagonal elements of the matrix identify the local torsion of the flow through the corresponding branch, measured in units of π . Branch 0 has no local torsion, and branch 1 shows a half-twist and has local torsion $+1$. The off-diagonal matrix

elements are twice the linking number of the period-1 orbits in the corresponding pair of branches. Since the period-1 orbits in these two branches do not link, the off-diagonal matrix elements are 0. The period-1 orbits in the branches labeled 1 and 2 in **Figure 7b** have linking number $+1$, so the off-diagonal matrix elements are $T(1, 2) = T(2, 1) = 2 \times +1$. The array identifies the order (above, below) that the two branches are joined at the branch line, the smaller the value, the closer to the viewer. These two pieces of information, four integers in **Figure 7a** and eight in

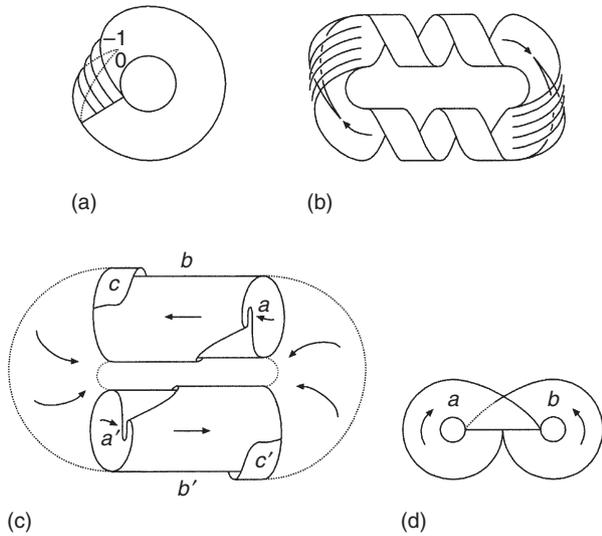


Figure 6 Branched manifolds for four standard sets of equations: (a) Rössler equations, (b) periodically driven Duffing equations, (c) periodically driven van der Pol equations, and (d) Lorenz equations. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

Table 1 Four sets of equations that generate strange attractors

Dynamical system	ODEs	Parameter values
Rössler	$\dot{x} = -y - z$ $\dot{y} = x + ay$ $\dot{z} = b + z(x - c)$	$(a, b, c) = (2.0, 4.0, 0.398)$
Duffing	$\dot{x} = y$ $\dot{y} = -\delta y - x^3 + x + A \sin(\omega t)$	$(\delta, A, \omega) = (0.4, 0.4, 1.0)$
van der Pol	$\dot{x} = by + (c - dy^2)x$ $\dot{y} = -x + A \sin(\omega t)$	$(b, c, d, A, \omega) = (0.7, 1.0, 10.0, 0.25, \pi/2)$
Lorenz	$\dot{x} = -\sigma x + \sigma y$ $\dot{y} = Rx - y - xz$ $\dot{z} = -bz + xy$	$(R, \sigma, b) = (26.0, 10.0, 8/3)$

Figure 7b, serve to determine the topological organization of all the unstable periodic orbits in any strange attractor with either branched manifold.

The periodic orbits are identified by a repeating symbol sequence of least period p , which is unique up to cyclic permutation. The symbol sequence consists of a string of integers, sequentially identifying the branches through which the orbit passes. For a branched manifold with two branches, there are two symbols. The number of orbits of period p , $N(p)$, obeys the recursion relation

$$pN(p) = 2^p - \sum_{1=k|p}^{k \leq p/2} kN(k) \quad [3]$$

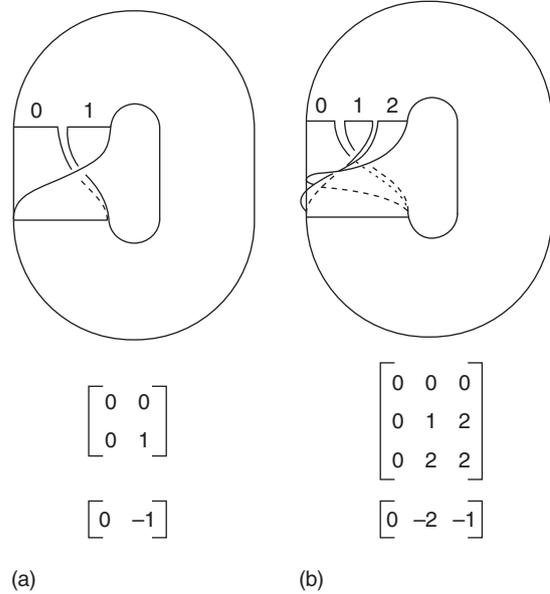


Figure 7 Branched manifolds are described algebraically. The diagonal matrix elements describe the twist of each branch. The off-diagonal matrix elements are twice the linking number of the period-1 orbits in each of the two branches. The array describes the order in which the branches are connected at the branch line. (a) Smale horseshoe branched manifold. (b) Beginning of a “gateau roulé” (jelly roll) branched manifold.

Table 2 shows the number of orbits of period $p \leq 20$ for the branched manifolds with two and three branches shown in Figure 7. The number of orbits of period p grows exponentially with p , and the limit $h_T = \lim_{p \rightarrow \infty} \log(N(p))/p$ defines the topological entropy h_T for the branched manifold. The limits are $\ln 2$ and $\ln 3$ for the branched manifolds with two and three branches, respectively. The linking numbers of orbits up to period 5 in the Smale horseshoe branched manifold are shown in Table 3, which identifies each of the orbits by its symbol sequence (e.g., 00111).

Table 2 Number of orbits of period p on the branched manifolds with two and three branches, shown in Figure 7. The integers $N_3(p)$ are constructed by replacing 2^p by 3^p in eqn [3]

Period p	Two branches	Three branches	Period p	Two branches	Three branches
	$N_2(p)$	$N_3(p)$		$N_2(p)$	$N_3(p)$
1	2	3	11	186	16 104
2	1	3	12	335	44 220
3	2	8	13	630	122 640
4	3	18	14	1 161	341 484
5	6	48	15	2 182	956 576
6	9	116	16	4 080	2 690 010
7	18	312	17	7 710	7 596 480
8	30	810	18	14 532	21 522 228
9	56	2184	19	27 954	61 171 656
10	99	5880	20	52 377	174 336 264

Table 3 Linking numbers of orbits to period 5 in the Smale horseshoe branched manifold with zero global torsion

		0	1	2 ₁	3 ₁	3 ₁	4 ₁	4 ₂	4 ₂	5 ₁	5 ₁	5 ₂	5 ₂	5 ₃	5 ₃
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	1	1	1	2	1	1	2	2	2	2	1	1
2 ₁	01	0	1	1	2	2	3	2	2	4	4	3	3	2	2
3 ₁	011	0	1	2	2	3	4	3	3	5	5	5	5	3	3
3 ₁	001	0	1	2	3	2	4	3	3	5	5	4	4	3	3
4 ₁	0111	0	2	3	4	4	5	4	4	8	8	7	7	4	4
4 ₂	0011	0	1	2	3	3	4	3	4	5	5	5	5	4	4
4 ₂	0001	0	1	2	3	3	4	4	3	5	5	5	5	4	4
5 ₁	01111	0	2	4	5	5	8	5	5	8	10	9	9	5	5
5 ₁	01101	0	2	4	5	5	8	5	5	10	8	8	8	5	5
5 ₂	00111	0	2	3	5	4	7	5	5	9	8	6	7	5	5
5 ₂	00101	0	2	3	5	4	7	5	5	9	8	7	6	5	5
5 ₃	00011	0	1	2	3	3	4	4	4	5	5	5	5	4	5
5 ₃	00001	0	1	2	3	3	4	4	4	5	5	5	5	5	4

Tables of linking numbers have been used successfully to identify mechanisms that nature uses to generate chaotic data. This analysis procedure is called topological analysis. Segments of data are identified that closely approximate unstable periodic orbits existing in the strange attractor. These data segments are then embedded in R^3 . Each orbit is given a trial identification (symbol sequence). Their pairwise linking numbers are computed either by counting signed crossings or using the time-parametrized data segments and estimating the integers numerically using the Gauss linking integral

$$\text{Link}(A, B) = \frac{1}{4\pi} \oint \oint \frac{\mathbf{r}_A(t_1) - \mathbf{r}_B(t_2)}{|\mathbf{r}_A(t_1) - \mathbf{r}_B(t_2)|^3} d\mathbf{r}_A(t_1) \times d\mathbf{r}_B(t_2)$$

This table of experimental integers is compared with the table of linking numbers for orbits with the same symbolic name on a trial branched manifold. This procedure serves to identify the branched manifold and refine the symbolic identifications of the experimental orbits, if necessary. The procedure is vastly overdetermined. For example, the linking numbers of only three low-period orbits serve to identify the four pieces of information required to specify a branched manifold with two branches. Since six or more surrogate periodic orbits can typically be extracted from experimental data, providing $\binom{6}{2} = 15$ or more linking numbers, this topological analysis procedure has built-in self-consistency checks, unlike analysis procedures based on geometric and dynamical tools.

Basis Sets of Orbits

A branched manifold determines the topological organization of all the periodic orbits that it

supports. Whenever a low-dimensional strange attractor is subjected to topological analysis, it is always the case that fewer periodic orbits are present and identified than are allowed by the branched manifold that classifies it. This is the case for strange attractors generated by experimental data as well as strange attractors generated by ODEs. The full spectrum occurs only in the hyperbolic limit, which has never been seen.

The orbits that are present are organized exactly as in the hyperbolic limit – that is, as determined by the underlying branched manifold. As control parameters change, the strange attractor undergoes perestroikas. New orbits are created and/or old orbits are annihilated in direct or inverse period-doubling and saddle-node bifurcations. The orbits that are present are always organized as determined by the branched manifold. Orbits are not created or annihilated independently of each other. Rather, there is a partial order (“forcing order”) involved in orbit creation and annihilation. This partial order is poorly understood for general branched manifolds. It is much better understood for the two-branch Smale horseshoe branched manifold.

The forcing diagram for this branched manifold is shown in **Figure 8** for orbits up to period 8. It is typically the case that the existence of one orbit in a strange attractor forces the presence of a spectrum of additional orbits. Forcing is transitive, so if orbit A forces orbit $B (A \Rightarrow B)$ and B forces C , then A forces C : if $A \Rightarrow B$ and $B \Rightarrow C$ then $A \Rightarrow C$. For this reason, it is sufficient to show only the first-order forcing in this figure. The orbits shown are labeled by their period and the order in which they are created in a particular highly dissipative limit of the dynamics: the logistic map (U-sequence order in **Figure 8**). For example, 5_2 describes the second (pair) of period-5 orbits created in the

Bounding Tori

As experimental conditions or control parameters change, strange attractors can undergo “grosser” perestroikas than those that can be described by a change in the basis set of orbits. This occurs when new orbits are created that cannot be contained on the initial branched manifold – for example, when orbits are created that must be described by a new symbol. This is seen experimentally in the transition from horseshoe type dynamics to gateau roulé type dynamics. This involves the addition of a third branch to the branched manifold with two branches, as shown in **Figures 7a and 7b**. Strange attractors can undergo perestroikas described by the addition of new branches to, or deletion of old branches from, a branched manifold. These perestroikas are in a very real sense “grosser” than the perestroikas that can be described by changes in the basis sets of orbits on a fixed branched manifold.

There is a structure that provides constraints on the allowed bifurcations of branched manifolds (creation/annihilation of branches), which is analogous to the constraints that a branched manifold provides on the bifurcations and topological organization of the periodic orbits that can exist on it. This structure is called a bounding torus.

Bounding tori are constructed as follows. The semiflow on a branched manifold is “inflated” or “blown up” to a flow on a thin open set in R^3 containing this branched manifold. The boundary of this open set is a two-dimensional surface. Such surfaces have been classified. They are uniquely tori of genus g ; $g=0$ (sphere), $g=1$ (tire tube), $g=2, 3, \dots$. The torus of genus g has Euler characteristic $\chi = 2 - 2g$. The flow is into this surface. The flow, restricted to the surface, exhibits a singularity wherever it is normal to the surface. At such singularities the stability is determined by the local Lyapunov exponents: $\lambda_1 > 0$ and $\lambda_3 < 0$, since the flow direction ($\lambda_2 = 0$) is normal to the

surface. As a result, all singularities are saddles; so, by the Poincaré–Hopf theorem, the number of singularities is strongly related to the genus. The number is $2(g - 1)$.

The flow, restricted to the genus- g surface, can be put into canonical form and these canonical forms can be classified. The classification involves projection of the genus- g torus onto a two-dimensional surface. The planar projection consists of a disk with outer boundary and g interior holes. All singularities can be placed on the interior holes. The flow on the interior holes without singularities is in the same direction as the flow on the exterior boundary. Interior holes with singularities have an even number, $4, 6, \dots$. Some canonical forms are shown in **Figure 9**.

Poincaré sections have been used to simplify the study of flows in low-dimensional spaces by effectively reducing the dimension of the dynamics. In three dimensions, a Poincaré surface of section for a strange attractor is a minimal two-dimensional surface with the property that all points in the attractor intersect this surface transversally an infinite number of times under the flow. The Poincaré surface need not be connected and in fact is often not connected.

The Poincaré section for the flow in a genus- g torus consists of the union of $g - 1$ disjoint disks ($g \geq 3$) or is a single disk ($g = 1$). The locations of the disks are determined algorithmically, as shown in **Figure 9**. The interior circles without singularities are labeled by capital letters A, B, C, \dots and those with singularities are labeled with lowercase letters a, b, c, \dots . The components of the global Poincaré surface of section are numbered sequentially $1, 2, \dots, g - 1$, in the order they are encountered when traversing the outer boundary in the direction of the flow, starting from any point on that boundary. Each component of the global Poincaré surface of section connects (in the projection) an interior circle without singularities to the exterior boundary. There is one component between each successive encounter of the flow with

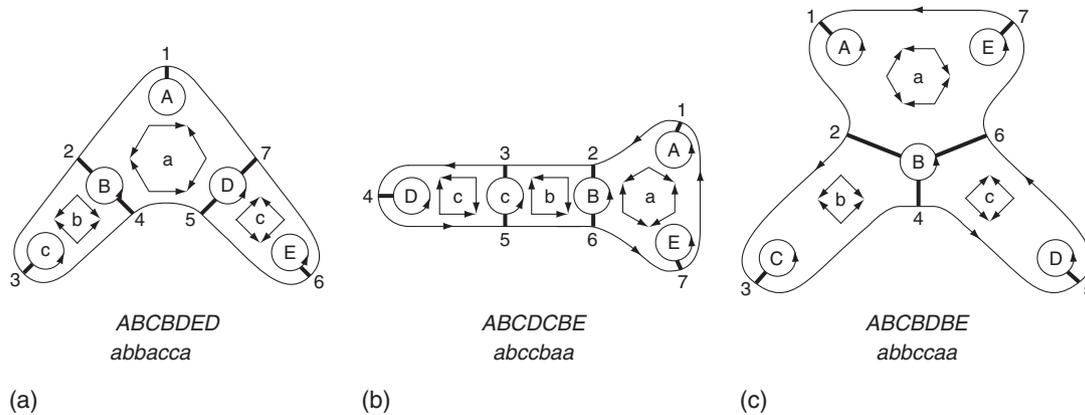


Figure 9 Three inequivalent canonical forms of genus 8 are shown. Each is identified by a “period-7 orbit” and its dual. Reprinted figure with permission from *Physical Review E*, 69, 056206, 2004. Copyright (2004) by the American Physical Society.

holes that have singularities. Heavy lines are used to show the location of the seven components of the global Poincaré surface of section for each of the three inequivalent genus-8 canonical forms shown in Figure 9. The structure of the flow is summarized by a transition matrix. For the canonical form shown in Figure 9c the transition matrix is

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where $T_{i,j} = 1$ if the flow can proceed directly from component i to component j , 0 otherwise.

Bounding tori, dressed with flows, can be labeled. In fact, two dual labeling schemes are possible. Following the outer boundary in the direction of the flow, one encounters the $g - 1$ components of the global Poincaré surface of section sequentially, the interior holes without singularities at least once each, and the interior holes with singularities at least twice each. The canonical form (genus- g torus dressed with a flow) on the genus-8 bounding torus shown in Figure 9a can be labeled by the sequence in which the holes without singularities are encountered (*ABCDED*) or the order in which the holes with singularities are encountered (*abbacca*). Both sequences contain $g - 1$ symbols. These labels are unique up to cyclic permutation.

Symbol sequences for canonical forms for bounding tori act in many ways like symbol sequences for periodic orbits on branched manifolds. Although there is a 1:1 correspondence between bounded closed two-dimensional surfaces in R^3 and genus g , the number of

Table 4 Number of canonical bounding tori as a function of genus g

g	$N(g)$	g	$N(g)$	g	$N(g)$
3	1	9	15	15	2211
4	1	10	28	16	5549
5	2	11	67	17	14290
6	2	12	145	18	36824
7	5	13	368	19	96347
8	6	14	870	20	252927

canonical forms grows rapidly with g , as shown in Table 4. In fact, the number, $N(g)$, grows exponentially and can even be assigned an entropy:

$$\lim_{g \rightarrow \infty} \frac{\ln(N(g))}{g - 1} = \ln 3 \quad [5]$$

In some sense, canonical forms that constrain branched manifolds within them behave like branched manifolds that constrain periodic orbits on them.

Every strange attractor that has been studied in R^3 has been described by a canonical bounding torus that contains it. This classification is shown in Table 5.

Branched manifold perestroikas are constrained by bounding tori as follows. Each branch line of any branched manifold can be moved into one of the $g - 1$ components of the global Poincaré surface of section. Any branched manifold contained in a genus- g bounding torus ($g \geq 3$) must have at least one branch between each pair of components of the global Poincaré surface of section between which the flow is allowed, as summarized by the canonical form's transition matrix. New branches can only be added in a way that is consistent with the canonical form's transition matrix, continuity requirements, and the no intersection condition.

Table 5 All known strange attractors of dimension $d_L < 3$ are bounded by one of the standard dressed tori. Dual labels for the bounding tori depend on $g - 1$ symbols describing holes with or without singularities

Strange attractor	Holes w/o singularities	Holes with singularities	Genus
Rosler, Duffing, Burke, and Shaw	A		1
Various lasers, gateau roulé	A		1
Neuron with subthreshold oscillations	A		1
Shaw-van der Pol	A		1
Lorenz, Shimizu-Morioka, Rikitake	AB	aa	3
C_2 covers of Rosler	AB	a^2	3
C_2 cover of Lorenz ^a	ABCD	a^4	5
C_2 cover of Lorenz ^b	ABCB	abba	5
2 → 1 Image of figure-8 branched manifold	ABCB	$ab(ab)^{-1}$	5
Figure-8 branched manifold	AEBECEDE	$a^2 b^2 c^2 d^2$	9
C_n covers of Rosler	$AB \dots N$	a^n	$n + 1$
C_n cover of Lorenz ^a	$AB \dots (2N)$	a^{2n}	$2n + 1$
C_n cover of Lorenz ^b	$(AZ)(BZ) \dots (NZ)$	$a^2 b^2 \dots n^2$	$2n + 1$
Multispiral attractors	$A(B \dots M)N(B \dots M)^{-1}$	$(ab \dots m)(ab \dots m)^{-1}$	$2m + 1$

^aRotation axis through origin.

^bRotation axis through one focus.

In the simplest case, $g=1$, a third branch can be added to a branched manifold with two branches only if its local torsion differs by ± 1 from the adjacent branch. In addition, the ordering of the new branch must be consistent with the continuity and no intersection (ODE uniqueness theorem) requirements.

Embeddings of Bounding Tori

The last level of topological structure needed for the classification of strange attractors in R^3 describes their embeddings in R^3 . The classification using genus- g bounding tori is intrinsic – that is, the canonical form shows how the flow looks from inside the torus. Strange attractors, and the tori that bound them, are actually embedded in R^3 . For a complete classification, we must specify not only the canonical form but also how this form sits in R^3 .

This program has not yet been completed, but we illustrate it with the genus-1 bounding torus in **Figure 10**. **Figure 10a** shows the canonical form, and two different embeddings of it in R^3 . The embedding on the left is unknotted. The embedding on the right is knotted like a figure-8 knot. Extrinsic embeddings of genus-1 tori are described by tame knots in R^3 , and tame knots can be used as “centerlines” for extrinsically embedded genus-1 tori. Higher-genus ($g \geq 3$) canonical forms – intrinsic genus- g tori dressed with a

canonical flow – have a larger (but discrete) variety of extrinsic embeddings in R^3 .

The Embedding Question

The mechanism that nature uses to generate chaotic behavior in physical systems is not directly observable, and must be deduced by examining the data that are generated. Typically, the data consist of a single scalar time series that is discretely recorded: $x_i, i=1, 2, \dots$. In order to exhibit a strange attractor, a mapping of the data into R^N must also be constructed. If the attractor is low dimensional ($d_L < 3$), one can hope that a mapping into R^3 can be constructed that exhibits no self-intersections or other degeneracies. Such a map is called an embedding. Once an embedding in R^3 is available, a topological analysis can be carried out. The analysis reveals the mechanism that underlies the creation of the embedded strange attractor.

But how do you know that the mechanism that generates the observed, embedded strange attractor has anything to do with the mechanism nature used to generate the experimental data?

If the embedding is contained in a genus-1 bounding torus, then the topological mechanism that generates the data, as defined by some unknown branched manifold $\mathcal{B}\mathcal{M}_{\text{EXP}}$, and the topological mechanism that is identified from the embedded strange attractor $\mathcal{B}\mathcal{M}_{\text{EMB}}$, are identical up to three degrees of freedom: parity, global torsion, and the knot type. As a result, in this case (genus-1) a topological analysis of embedded data does reveal nature’s hidden secrets.

See also: Ergodic theory; Fractal dimensions in dynamics; Generic Properties of Dynamical Systems; Gravitational N -body Problem (Classical); Homeomorphisms and Diffeomorphisms of the Circle; Homoclinic phenomena; Inviscid Flows; Lyapunov Exponents and Strange Attractors; Nonequilibrium Statistical Mechanics (Stationary): Overview; Random Algebraic Geometry, Attractors and Flux Vacua; Random Matrix Theory in Physics; Regularization for Dynamical Zeta Functions; Singularity and Bifurcation Theory; Symmetry and Symmetry Breaking in Dynamical Systems; Synchronization of Chaos.

Further Reading

- Abraham R and Shaw CD (1992) *Dynamics: The Geometry of Behavior*, Studies in Nonlinearity, 2nd edn. Reading, MA: Addison-Wesley.
- Eckmann J-P and Ruelle D (1985) Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* 57(3): 617–656.
- Gilmore R (1998) Topological analysis of chaotic dynamical systems. *Reviews of Modern Physics* 70(4): 1455–1529.
- Gilmore R and Lefranc M (2002) *The Topology of Chaos, Alice in Stretch and Squeezeland*. New York: Wiley.

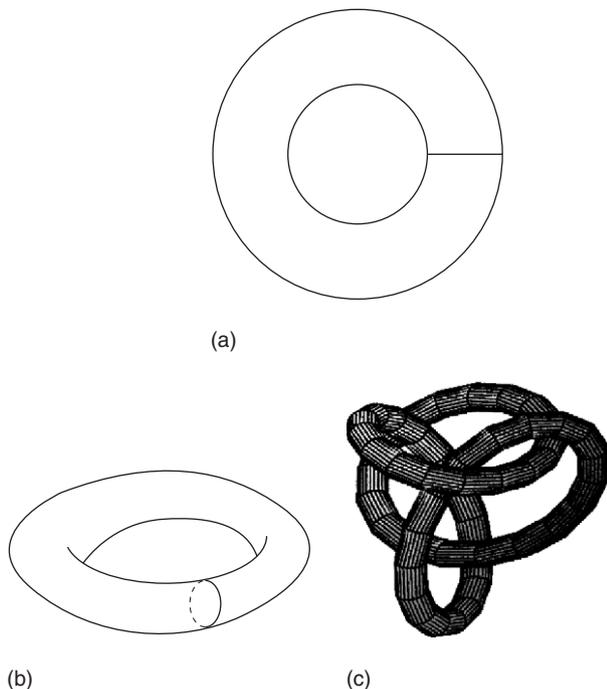


Figure 10 (a) Canonical form for genus-1 bounding torus. Extrinsic embeddings of the torus into R^3 that are (b) unknotted and (c) knotted like the figure-8 knot.

- Gilmore R and Letellier C (2006) *The Symmetry of Chaos Alice in the Land of Mirrors*. Oxford: Oxford University Press.
- Gilmore R and Pei X (2001) The topology and organization of unstable periodic orbits in Hodgkin–Huxley models of receptors with subthreshold oscillations. In: Moss F and Gielen S (eds.) *Handbook of Biological Physics, Neuro-informatics, Neural Modeling*, vol. 4, pp. 155–203. Amsterdam: North-Holland.

- Ott E (1993) *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.
- Solari HG, Natiello MA, and Mindlin GB (1996) *Nonlinear Physics and Its Mathematical Tools*. Bristol: IoP Publishing.
- Tufillaro NB, Abbott T, and Reilly J (1992) *An Experimental Approach to Nonlinear Dynamics and Chaos*. Reading, MA: Addison-Wesley.

Characteristic Classes

P B Gilkey, University of Oregon, Eugene, OR, USA
R Ivanova, University of Hawaii Hilo, Hilo, HI, USA
S Nikčević, SANU, Belgrade, Serbia and Montenegro

© 2006 Elsevier Ltd. All rights reserved.

Vector Bundles

Let $\text{Vect}_k(M, \mathbb{F})$ be the set of isomorphism classes of real ($\mathbb{F} = \mathbb{R}$) or complex ($\mathbb{F} = \mathbb{C}$) vector bundles of rank k over a smooth connected m -dimensional manifold M . Let

$$\text{Vect}(M, \mathbb{F}) = \bigcup_k \text{Vect}_k(M, \mathbb{F})$$

Principal Bundles – Examples

Let H be a Lie group. A fiber bundle

$$\rho: \mathcal{P} \rightarrow M$$

with fiber H is said to be a principal bundle if there is a right action of H on \mathcal{P} which acts transitively on the fibers, that is, if $\mathcal{P}/H = M$. If H is a closed subgroup of a Lie group G , then the natural projection $G \rightarrow G/H$ is a principal H bundle over the homogeneous space G/H . Let $O(k)$ and $U(k)$ denote the orthogonal and unitary groups, respectively. Let S^k denote the unit sphere in \mathbb{R}^{k+1} . Then we have natural principal bundles:

$$\begin{aligned} O(k) &\subset O(k+1) \rightarrow S^k \\ U(k) &\subset U(k+1) \rightarrow S^{2k+1} \end{aligned}$$

Let $\mathbb{R}P^k$ and $\mathbb{C}P^k$ denote the real and complex projective spaces of lines through the origin in \mathbb{R}^{k+1} and \mathbb{C}^{k+1} , respectively. Let

$$\begin{aligned} \mathbb{Z}_2 &= \{\pm \text{Id}\} \subset O(k) \\ \mathbb{S}^1 &= \{\lambda \cdot \text{Id} : |\lambda| = 1\} \subset U(k) \end{aligned}$$

One has \mathbb{Z}_2 and \mathbb{S}^1 principal bundles:

$$\begin{aligned} \mathbb{Z}_2 &\rightarrow S^{k-1} \rightarrow \mathbb{R}P^{k-1} \\ \mathbb{S}^1 &\rightarrow S^{2k-1} \rightarrow \mathbb{C}P^{k-1} \end{aligned}$$

Frames

A frame $\mathbf{s} := (s_1, \dots, s_k)$ for $V \in \text{Vect}_k(M, \mathbb{F})$ over an open set $\mathcal{O} \subset M$ is a collection of k smooth sections to $V|_{\mathcal{O}}$ so that $\{s_1(P), \dots, s_k(P)\}$ is a basis for the fiber V_P of V over any point $P \in \mathcal{O}$. Given such a frame \mathbf{s} , we can construct a local trivialization which identifies $\mathcal{O} \times \mathbb{F}^k$ with $V|_{\mathcal{O}}$ by the mapping

$$(P; \lambda_1, \dots, \lambda_k) \rightarrow \lambda_1 s_1(P) + \dots + \lambda_k s_k(P)$$

Conversely, given a local trivialization of V , we can take the coordinate frame

$$s_i(P) = P \times (0, \dots, 0, 1, 0, \dots, 0)$$

Thus, frames and local trivializations of V are equivalent notions.

Simple Covers

An open cover $\{\mathcal{O}_\alpha\}$ of M , where α ranges over some indexing set A , is said to be a simple cover if any finite intersection $\mathcal{O}_{\alpha_1} \cap \dots \cap \mathcal{O}_{\alpha_k}$ is either empty or contractible.

Simple covers always exist. Put a Riemannian metric on M . If M is compact, then there exists a uniform $\delta > 0$ so that any geodesic ball of radius δ is geodesically convex. The intersection of geodesically convex sets is either geodesically convex (and hence contractible) or empty. Thus, covering M by a finite number of balls of radius δ yields a simple cover. The argument is similar even if M is not compact where an infinite number of geodesic balls is used and the radii are allowed to shrink near ∞ .

Transition Cocycles

Let $\text{Hom}(\mathbb{F}, k)$ be the set of linear transformations of \mathbb{F}^k and let $\text{GL}(\mathbb{F}, k) \subset \text{Hom}(\mathbb{F}, k)$ be the group of all invertible linear transformations.

Let $\{s_\alpha\}$ be frames for a vector bundle V over some open cover $\{\mathcal{O}_\alpha\}$ of M . On the intersection $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$, one may express $s_\alpha = \psi_{\alpha\beta} s_\beta$, that is

$$s_{\alpha,i}(P) = \sum_{1 \leq j \leq k} \psi_{\alpha\beta,i}^j(P) s_{\beta,j}(P)$$

Gilmore R and Letellier C (2006) *The Symmetry of Chaos Alice in the Land of Mirrors*. Oxford: Oxford University Press.
 Gilmore R and Pei X (2001) The topology and organization of unstable periodic orbits in Hodgkin–Huxley models of receptors with subthreshold oscillations. In: Moss F and Gielen S (eds.) *Handbook of Biological Physics, Neuro-informatics, Neural Modeling*, vol. 4, pp. 155–203. Amsterdam: North-Holland.

Ott E (1993) *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.
 Solari HG, Natiello MA, and Mindlin GB (1996) *Nonlinear Physics and Its Mathematical Tools*. Bristol: IoP Publishing.
 Tufillaro NB, Abbott T, and Reilly J (1992) *An Experimental Approach to Nonlinear Dynamics and Chaos*. Reading, MA: Addison-Wesley.

Characteristic Classes

P B Gilkey, University of Oregon, Eugene, OR, USA
R Ivanova, University of Hawaii Hilo, Hilo, HI, USA
S Nikčević, SANU, Belgrade, Serbia and Montenegro

© 2006 Elsevier Ltd. All rights reserved.

Vector Bundles

Let $\text{Vect}_k(M, \mathbb{F})$ be the set of isomorphism classes of real ($\mathbb{F} = \mathbb{R}$) or complex ($\mathbb{F} = \mathbb{C}$) vector bundles of rank k over a smooth connected m -dimensional manifold M . Let

$$\text{Vect}(M, \mathbb{F}) = \bigcup_k \text{Vect}_k(M, \mathbb{F})$$

Principal Bundles – Examples

Let H be a Lie group. A fiber bundle

$$\rho: \mathcal{P} \rightarrow M$$

with fiber H is said to be a principal bundle if there is a right action of H on \mathcal{P} which acts transitively on the fibers, that is, if $\mathcal{P}/H = M$. If H is a closed subgroup of a Lie group G , then the natural projection $G \rightarrow G/H$ is a principal H bundle over the homogeneous space G/H . Let $O(k)$ and $U(k)$ denote the orthogonal and unitary groups, respectively. Let S^k denote the unit sphere in \mathbb{R}^{k+1} . Then we have natural principal bundles:

$$\begin{aligned} O(k) &\subset O(k+1) \rightarrow S^k \\ U(k) &\subset U(k+1) \rightarrow S^{2k+1} \end{aligned}$$

Let $\mathbb{R}P^k$ and $\mathbb{C}P^k$ denote the real and complex projective spaces of lines through the origin in \mathbb{R}^{k+1} and \mathbb{C}^{k+1} , respectively. Let

$$\begin{aligned} \mathbb{Z}_2 &= \{\pm \text{Id}\} \subset O(k) \\ \mathbb{S}^1 &= \{\lambda \cdot \text{Id} : |\lambda| = 1\} \subset U(k) \end{aligned}$$

One has \mathbb{Z}_2 and \mathbb{S}^1 principal bundles:

$$\begin{aligned} \mathbb{Z}_2 &\rightarrow S^{k-1} \rightarrow \mathbb{R}P^{k-1} \\ \mathbb{S}^1 &\rightarrow S^{2k-1} \rightarrow \mathbb{C}P^{k-1} \end{aligned}$$

Frames

A frame $\mathbf{s} := (s_1, \dots, s_k)$ for $V \in \text{Vect}_k(M, \mathbb{F})$ over an open set $\mathcal{O} \subset M$ is a collection of k smooth sections to $V|_{\mathcal{O}}$ so that $\{s_1(P), \dots, s_k(P)\}$ is a basis for the fiber V_P of V over any point $P \in \mathcal{O}$. Given such a frame \mathbf{s} , we can construct a local trivialization which identifies $\mathcal{O} \times \mathbb{F}^k$ with $V|_{\mathcal{O}}$ by the mapping

$$(P; \lambda_1, \dots, \lambda_k) \rightarrow \lambda_1 s_1(P) + \dots + \lambda_k s_k(P)$$

Conversely, given a local trivialization of V , we can take the coordinate frame

$$s_i(P) = P \times (0, \dots, 0, 1, 0, \dots, 0)$$

Thus, frames and local trivializations of V are equivalent notions.

Simple Covers

An open cover $\{\mathcal{O}_\alpha\}$ of M , where α ranges over some indexing set A , is said to be a simple cover if any finite intersection $\mathcal{O}_{\alpha_1} \cap \dots \cap \mathcal{O}_{\alpha_k}$ is either empty or contractible.

Simple covers always exist. Put a Riemannian metric on M . If M is compact, then there exists a uniform $\delta > 0$ so that any geodesic ball of radius δ is geodesically convex. The intersection of geodesically convex sets is either geodesically convex (and hence contractible) or empty. Thus, covering M by a finite number of balls of radius δ yields a simple cover. The argument is similar even if M is not compact where an infinite number of geodesic balls is used and the radii are allowed to shrink near ∞ .

Transition Cocycles

Let $\text{Hom}(\mathbb{F}, k)$ be the set of linear transformations of \mathbb{F}^k and let $\text{GL}(\mathbb{F}, k) \subset \text{Hom}(\mathbb{F}, k)$ be the group of all invertible linear transformations.

Let $\{s_\alpha\}$ be frames for a vector bundle V over some open cover $\{\mathcal{O}_\alpha\}$ of M . On the intersection $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$, one may express $s_\alpha = \psi_{\alpha\beta} s_\beta$, that is

$$s_{\alpha,i}(P) = \sum_{1 \leq j \leq k} \psi_{\alpha\beta,i}^j(P) s_{\beta,j}(P)$$

The maps $\psi_{\alpha\beta} : \mathcal{O}_\alpha \cap \mathcal{O}_\beta \rightarrow \text{GL}(\mathbb{F}, k)$ satisfy

$$\begin{aligned} \psi_{\alpha\alpha} &= \text{Id} \quad \text{on } \mathcal{O}_\alpha \\ \psi_{\alpha\beta} &= \psi_{\alpha\gamma} \psi_{\gamma\beta} \quad \text{on } \mathcal{O}_\alpha \cap \mathcal{O}_\beta \cap \mathcal{O}_\gamma \end{aligned} \tag{1}$$

Let G be a Lie group. Maps belonging to a collection $\{\psi_{\alpha\beta}\}$ of smooth maps from $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$ to G which satisfy eqn [1] are said to be transition cocycles with values in G ; if $G \subset \text{GL}(\mathbb{F}, k)$, they can be used to define a vector bundle by making appropriate identifications.

Reducing the Structure Group

If G is a subgroup of $\text{GL}(\mathbb{F}, k)$, then V is said to have a G -structure if we can choose frames so the transition cocycles belong to G ; that is, we can reduce the structure group to G .

Denote the subgroup of orientation-preserving linear maps by

$$\text{GL}^+(\mathbb{R}, k) := \{\psi \in \text{GL}(\mathbb{R}, k) : \det(\psi) > 0\}$$

If $V \in \text{Vect}_k(M, \mathbb{R})$, then V is said to be orientable if we can choose the frames so that

$$\psi_{\alpha\beta} \in \text{GL}^+(\mathbb{R}, k)$$

Not every real vector bundle is orientable; the first Stiefel–Whitney class $sw_1(V) \in H^1(M; \mathbb{Z}_2)$, which is defined later, vanishes if and only if V is orientable. In particular, the Möbius line bundle over the circle is not orientable.

Similarly, a real (resp. complex) bundle V is said to be Riemannian (resp. Hermitian) if we can reduce the structure group to the orthogonal group $\text{O}(k) \subset \text{GL}(\mathbb{R}, k)$ (resp. to the unitary group $\text{U}(k) \subset \text{GL}(\mathbb{C}, k)$).

We can use a partition of unity to put a positive-definite symmetric (resp. Hermitian symmetric) fiber metric on V . Applying the Gram–Schmidt process then constructs orthonormal frames and shows that the structure group can always be reduced to $\text{O}(k)$ (resp. to $\text{U}(k)$); if V is a real vector bundle, then the structure group can be reduced to the special orthogonal group $\text{SO}(k)$ if and only if V is orientable.

Lifting the Structure Group

Let τ be a representation of a Lie group H to $\text{GL}(\mathbb{F}, k)$. One says that the structure group of V can be lifted to H if there exist frames $\{s_\alpha\}$ for V and smooth maps $\phi_{\alpha\beta} : \mathcal{O}_\alpha \cap \mathcal{O}_\beta \rightarrow H$, so $\tau\phi_{\alpha\beta} = \psi_{\alpha\beta}$ where eqn [1] holds for ϕ .

Spin Structures

For $k \geq 3$, the fundamental group of $\text{SO}(k)$ is \mathbb{Z}_2 . Let $\text{Spin}(k)$ be the universal cover of $\text{SO}(k)$ and let

$$\tau : \text{Spin}(k) \rightarrow \text{SO}(k)$$

be the associated double cover; set $\text{Spin}(2) = S^1$ and let $\tau(\lambda) = \lambda^2$. An oriented bundle V is said to be spin if the transition functions can be lifted from $\text{SO}(k)$ to $\text{Spin}(k)$; this is possible if and only if the second Stiefel–Whitney class of V , which is defined later, vanishes. There can be inequivalent spin structures, which are parametrized by the cohomology group $H^1(M; \mathbb{Z}_2)$.

The Tangent Bundle of Projective Space

The tangent bundle TRP^m of real projective space is orientable if and only if m is odd; TRP^m is spin if and only if $m \equiv 3 \pmod{4}$. If $m \equiv 3 \pmod{4}$, there are two inequivalent spin structures on this bundle as $H^1(\mathbb{RP}^m; \mathbb{Z}_2) = \mathbb{Z}_2$.

The tangent bundle TCP^m of complex projective space is always orientable; TCP^m is spin if and only if m is odd.

Principal and Associated Bundles

Let H be a Lie group and let

$$\phi_{\alpha\beta} : \mathcal{O}_\alpha \cap \mathcal{O}_\beta \rightarrow H$$

be a collection of smooth functions satisfying the compatibility conditions given in eqn [1]. We define a principal bundle \mathcal{P} by gluing $\mathcal{O}_\alpha \times H$ to $\mathcal{O}_\beta \times H$ using ϕ :

$$(P, h)_\alpha \sim (P, \phi_{\alpha\beta}(P)h)_\beta \quad \text{for } P \in \mathcal{O}_\alpha \cap \mathcal{O}_\beta$$

Because right multiplication and left multiplication commute, right multiplication gives a natural action of H on \mathcal{P} :

$$(P, h)_\alpha \cdot \tilde{h} := (P, h \cdot \tilde{h})_\alpha$$

The natural projection $\mathcal{P} \rightarrow \mathcal{P}/H = M$ is an H fiber bundle.

Let τ be a representation of H to $\text{GL}(\mathbb{F}, k)$. For $\xi \in \mathcal{P}$, $\lambda \in \mathbb{F}^k$, and $h \in H$, define a gluing

$$(\xi, \lambda) \sim (\xi \cdot h^{-1}, \tau(h)\lambda)$$

The associated vector bundle is then given by

$$\mathcal{P} \times_\tau \mathbb{F}^k := \mathcal{P} \times \mathbb{F}^k / \sim$$

Clearly, $\{\tau\phi_{\alpha\beta}\}$ are the transition cocycles of the vector bundle $\mathcal{P} \times_\tau \mathbb{F}^k$.

Frame Bundles

If V is a vector bundle, the associated principal $GL(\mathbb{F}, k)$ bundle is the bundle of all frames; if V is given an inner product on each fiber, then the associated principal $O(k)$ or $U(k)$ bundle is the bundle of orthonormal frames. If V is an oriented Riemannian vector bundle, the associated principal $SO(k)$ bundle is the bundle of oriented orthonormal frames.

Direct Sum and Tensor Product

Fiber-wise direct sum (resp. tensor product) defines the direct sum (resp. tensor product) of vector bundles:

$$\begin{aligned} \oplus : \text{Vect}_k(M, \mathbb{F}) \times \text{Vect}_n(M, \mathbb{F}) &\rightarrow \text{Vect}_{k+n}(M, \mathbb{F}) \\ \otimes : \text{Vect}_k(M, \mathbb{F}) \times \text{Vect}_n(M, \mathbb{F}) &\rightarrow \text{Vect}_{kn}(M, \mathbb{F}) \end{aligned}$$

The transition cocycles of the direct sum (resp. tensor product) of two vector bundles are the direct sum (resp. tensor product) of the transition cocycles of the respective bundles.

The set of line bundles $\text{Vect}_1(M, \mathbb{F})$ is a group under \otimes . The unit in the group is the trivial line bundle $1 := M \times \mathbb{F}$; the inverse of a line bundle L is the dual line bundle $L^* := \text{Hom}(L, \mathbb{F})$ since

$$L \otimes L^* = 1$$

Pullback Bundle

Let $\rho: V \rightarrow M$ be the projection associated with $V \in \text{Vect}_k(M, \mathbb{F})$. If f is a smooth map from N to M , then the pullback bundle f^*V is the vector bundle over N which is defined by setting

$$f^*V := \{(P, v) \in N \times V : f(P) = \rho(v)\}$$

The fiber of f^*V over P is the fiber of V over $f(P)$.

Let $\{s_\alpha\}$ be local frames for V over an open cover $\{\mathcal{O}_\alpha\}$ of M . For $P \in f^{-1}(\mathcal{O}_\alpha)$, define

$$\{f^*s_\alpha\}(P) := (P, s_\alpha(f(P)))$$

This gives a collection of frames for f^*V over the open cover $\{f^{-1}(\mathcal{O}_\alpha)\}$ of N . Let

$$f^*\psi_{\alpha\beta} := \psi_{\alpha\beta} \circ f$$

be the pullback of the transition functions. Then

$$\begin{aligned} \{f^*s_\alpha\}(P) &= (P, \psi_{\alpha\beta}(f(P))s_\beta(f(P))) \\ &= \{(f^*\psi_{\alpha\beta})(f^*s_\beta)\}(P) \end{aligned}$$

This shows that the pullback of the transition functions for V are the transition functions of the pullback $f^*(V)$.

Homotopy

Two smooth maps f_0 and f_1 from N to M are said to be homotopic if there exists a smooth map $F: N \times I \rightarrow M$ so that $f_0(P) = F(P, 0)$ and so that $f_1(P) = F(P, 1)$. If f_0 and f_1 are homotopic maps from N to M , then f_1^*V is isomorphic to f_0^*V .

Let $[N, M]$ be the set of all homotopy classes of smooth maps from N to M . The association $V \rightarrow f^*V$ induces a natural map

$$[N, M] \times \text{Vect}_k(M, \mathbb{F}) \rightarrow \text{Vect}_k(N, \mathbb{F})$$

If M is contractible, then the identity map is homotopic to the constant map c . Consequently, $V = \text{Id}^*V$ is isomorphic to $c^*V = M \times \mathbb{F}^k$. Thus, any vector bundle over a contractible manifold is trivial. In particular, if $\{\mathcal{O}_\alpha\}$ is a simple cover of M and if $V \in \text{Vect}(M, \mathbb{F})$, then $V|_{\mathcal{O}_\alpha}$ is trivial for each α . This shows that a simple cover is a trivializing cover for every $V \in \text{Vect}(M, \mathbb{F})$.

Stabilization

Let $1 \in \text{Vect}_1(M, \mathbb{F})$ denote the isomorphism class of the trivial line bundle $M \times \mathbb{F}$ over an m -dimensional manifold M . The map $V \rightarrow V \oplus 1$ induces a stabilization map

$$s : \text{Vect}_k(M, \mathbb{F}) \rightarrow \text{Vect}_{k+1}(M, \mathbb{F})$$

which induces an isomorphism

$$\begin{aligned} \text{Vect}_k(M, \mathbb{R}) &= \text{Vect}_{k+1}(M, \mathbb{R}) \quad \text{for } k > m \\ \text{Vect}_k(M, \mathbb{C}) &= \text{Vect}_{k+1}(M, \mathbb{C}) \quad \text{for } 2k > m \end{aligned} \quad [2]$$

These values of k comprise the stable range.

The K-Theory

The direct sum \oplus and tensor product \otimes make $\text{Vect}(M, \mathbb{F})$ into a semiring; we denote the associated ring defined by the Grothendieck construction by $\text{KF}(M)$. If $V \in \text{Vect}(M, \mathbb{F})$, let $[V] \in \text{KF}(M)$ be the corresponding element of K -theory; $\text{KF}(M)$ is generated by formal differences $[V_1] - [V_2]$; such formal differences are called virtual bundles.

The Grothendieck construction (see K -theory) introduces nontrivial relations. Let S^m denote the standard sphere in \mathbb{R}^{m+1} . Since

$$T(S^m) \oplus 1 = (m + 1)1$$

we can easily see that $[TS^m] = m[1]$ in $\text{KR}(S^m)$, despite the fact that $T(S^m)$ is not isomorphic to $m1$ for $m \neq 1, 3, 7$.

Let L denote the nontrivial real line bundle over $\mathbb{R}P^k$. Then $\text{TRP}^k \oplus 1 = (k + 1)L$, so

$$[\text{TRP}^k] = (k + 1)[L] - [1]$$

The map $V \rightarrow \text{Rank}(V)$ extends to a surjective map from $\text{KF}(M)$ to \mathbb{Z} . We denote the associated ideal of virtual bundles of virtual rank 0 by

$$\widetilde{\text{KF}}(M) := \ker(\text{Rank})$$

In the stable range, $V \rightarrow [V] - k[1]$ identifies

$$\begin{aligned} \text{Vect}_k(M, \mathbb{R}) &= \widetilde{\text{KR}}(M) & \text{if } k > m \\ \text{Vect}_k(M, \mathbb{C}) &= \widetilde{\text{KC}}(M) & \text{if } 2k > m \end{aligned} \quad [3]$$

These groups contain nontrivial torsion. Let L be the nontrivial real line bundle over $\mathbb{R}P^k$. Then

$$\widetilde{\text{KR}}(\mathbb{R}P^k) = \mathbb{Z} \cdot \{[L] - [1]\} / 2^{\nu(k)} \mathbb{Z} \{[L] - [1]\}$$

where $\nu(k)$ is the Adams number.

Classifying Spaces

Let $\text{Gr}_k(\mathbb{F}, n)$ be the Grassmannian of k -dimensional subspaces of \mathbb{F}^n . By mapping a k -plane π in \mathbb{F}^n to the corresponding orthogonal projection on π , we can identify $\text{Gr}_k(\mathbb{F}, n)$ with the set of orthogonal projections of rank k :

$$\{\xi \in \text{Hom}(\mathbb{F}^n) : \xi^2 = \xi, \xi^* = \xi, \text{tr}(\xi) = k\}$$

There is a natural associated tautological k -plane bundle

$$V_k(\mathbb{F}, n) \in \text{Vect}_k(\text{Gr}_k(\mathbb{F}, n), \mathbb{F})$$

whose fiber over a k -plane π is the k -plane itself:

$$V_k(\mathbb{F}, n) := \{(\xi, x) \in \text{Hom}(\mathbb{F}^n) \times \mathbb{F}^n : \xi x = x\}$$

Let $[M, \text{Gr}_k(\mathbb{F}, n)]$ denote the set of homotopy equivalence classes of smooth maps f from M to $\text{Gr}_k(\mathbb{F}, n)$. Since $[f_1] = [f_2]$ implies that $f_1^* V$ is isomorphic to $f_2^* V$, the association

$$f \rightarrow f^* V_k(\mathbb{F}, n) \in \text{Vect}_k(M, \mathbb{F})$$

induces a map

$$[M, \text{Gr}_k(\mathbb{F}, n)] \rightarrow \text{Vect}_k(M, \mathbb{F})$$

This map defines a natural equivalence of functors in the stable range:

$$\begin{aligned} [M, \text{Gr}_k(\mathbb{R}, \nu + k)] &= \text{Vect}_k(M, \mathbb{R}) & \text{for } \nu > m \\ [M, \text{Gr}_k(\mathbb{C}, \nu + k)] &= \text{Vect}_k(M, \mathbb{C}) & \text{for } 2\nu > m \end{aligned} \quad [4]$$

The natural inclusion of \mathbb{F}^n in \mathbb{F}^{n+1} induces natural inclusions

$$\begin{aligned} \text{Gr}_k(\mathbb{F}, n) &\subset \text{Gr}_k(\mathbb{F}, n + 1) \\ V_k(\mathbb{F}, n) &\subset V_k(\mathbb{F}, n + 1) \end{aligned} \quad [5]$$

Let $\text{Gr}_k(\mathbb{F}, \infty)$ and $V_k(\mathbb{F}, \infty)$ be the direct limit spaces under these inclusions; these are the infinite-dimensional Grassmannians and classifying bundles,

respectively. The topology on these spaces is the weak or inductive topology. The Grassmannians are called classifying spaces. The isomorphisms of eqn [4] are compatible with the inclusions of eqn [5] and we have

$$[M, \text{Gr}_k(\mathbb{F}, \infty)] = \text{Vect}_k(M, \mathbb{F}) \quad [6]$$

Spaces with Finite Covering Dimension

A metric space X is said to have a covering dimension at most m if, given any open cover $\{\mathcal{U}_\alpha\}$ of X , there exists a refinement $\{\mathcal{O}_\beta\}$ of the cover so that any intersection of more than $m + 1$ of the $\{\mathcal{O}_\beta\}$ is empty. For example, any manifold of dimension m has covering dimension at most m . More generally, any m -dimensional cell complex has covering dimension at most m .

The isomorphisms of [2]–[4], and [6] continue to hold under the weaker assumption that M is a metric space with covering dimension at most m .

Characteristic Classes of Vector Bundles

The Cohomology of $\text{Gr}_k(\mathbb{F}, \infty)$

The cohomology algebras of the Grassmannians are polynomial algebras on suitably chosen generators:

$$\begin{aligned} H^*(\text{Gr}_k(\mathbb{R}, \infty); \mathbb{Z}_2) &= \mathbb{Z}_2[\text{sw}_1, \dots, \text{sw}_k] \\ H^*(\text{Gr}_k(\mathbb{C}, \infty); \mathbb{Z}) &= \mathbb{Z}[c_1, \dots, c_k] \end{aligned} \quad [7]$$

The Stiefel–Whitney Classes

Let $V \in \text{Vect}_k(M, \mathbb{R})$. We use eqn [6] to find $\Psi : M \rightarrow \text{Gr}_k(\mathbb{R}, \infty)$ which classifies V ; the map Ψ is uniquely determined up to homotopy and, using eqn [7], one sets

$$\text{sw}_i(V) := \Psi^* \text{sw}_i \in H^i(M; \mathbb{Z}_2)$$

The total Stiefel–Whitney class is then defined by

$$\text{sw}(V) = 1 + \text{sw}_1(V) + \dots + \text{sw}_k(V)$$

The Stiefel–Whitney class has the properties:

1. If $f : X_1 \rightarrow X_2$, then $f^*(\text{sw}(V)) = \text{sw}(f^*V)$.
2. $\text{sw}(V \oplus W) = \text{sw}(V)\text{sw}(W)$.
3. If L is the Möbius bundle over S^1 , then $\text{sw}_1(L)$ generates $H^1(S^1; \mathbb{Z}_2) = \mathbb{Z}_2$.

The cohomology algebra of real projective space is a truncated polynomial algebra:

$$H^*(\mathbb{R}P^k; \mathbb{Z}_2) = \mathbb{Z}_2[x]/x^{k+1} = 0$$

Since $\text{TRP}^k \oplus 1 = (k + 1)L$, one has

$$\begin{aligned} \text{sw}(\text{TRP}^k) &= (1 + x)^{k+1} \\ &= 1 + kx + \frac{(k + 1)k}{2}x^2 + \dots \end{aligned} \quad [8]$$

Orientability and Spin Structures

The Stiefel–Whitney classes have real geometric meaning. For example, $\text{sw}_1(V) = 0$ if and only if V is orientable; if $\text{sw}_1(V) = 0$, then $\text{sw}_2(V) = 0$ if and only if V admits a spin structure. With reference to the discussion on the tangent bundle or projective space, eqn [8] yields

$$\text{sw}_1(\text{TRP}^k) = \begin{cases} 0 & \text{if } k \equiv 0 \pmod{2} \\ x & \text{if } k \equiv 1 \pmod{2} \end{cases}$$

Thus, RP^k is orientable if and only if k is odd. Furthermore,

$$\text{sw}_2(\text{TRP}^k) = \begin{cases} 0 & \text{if } k \equiv 3 \pmod{4} \\ x & \text{if } k \equiv 1 \pmod{4} \end{cases}$$

Thus, TRP^k is spin if and only if $k \equiv 3 \pmod{4}$.

Chern Classes

Let $V \in \text{Vect}_k(M, \mathbb{C})$. We use eqn [6] to find $\Psi : M \rightarrow \text{Gr}_k(\mathbb{C}, \infty)$ which classifies V ; the map Ψ is uniquely determined up to homotopy and, using eqn [7], one sets

$$c_i(V) := \Psi^* c_i \in H^{2i}(M; \mathbb{Z})$$

The total Chern class is then defined by

$$c(V) := 1 + c_1(V) + \dots + c_k(V)$$

The Chern class has the properties:

1. If $f : X_1 \rightarrow X_2$, then $f^*(c(V)) = c(f^*V)$.
2. $c(V \oplus W) = c(V)c(W)$.
3. Let L be the classifying line bundle over $S^2 = \mathbb{C}\mathbb{P}^1$. Then $\int_{S^2} c_1(L) = -1$.

The cohomology algebra of complex projective space also is a truncated polynomial algebra

$$H^*(\mathbb{C}\mathbb{P}^k; \mathbb{Z}) = \mathbb{Z}[x]/x^{k+1}$$

where $x = c_1(L)$ and L is the complex classifying line bundle over $\mathbb{C}\mathbb{P}^k = \text{Gr}_1(\mathbb{C}, k + 1)$. If $T_c\mathbb{C}\mathbb{P}^k$ is the complex tangent bundle, then

$$c(T_c\mathbb{C}\mathbb{P}^k) = (1 + x)^{k+1}$$

The Pontrjagin Classes

Let V be a real vector bundle over a topological space X of rank $r = 2k$ or $r = 2k + 1$. The Pontrjagin

classes $p_i(V) \in H^{4i}(X; \mathbb{Z})$ are characterized by the properties:

1. $p(V) = 1 + p_1(V) + \dots + p_k(V)$.
2. If $f : X_1 \rightarrow X_2$, then $f^*(p(V)) = p(f^*V)$.
3. $p(V \oplus W) = p(V)p(W) \pmod{\text{elements of order 2}}$.
4. $\int_{\mathbb{C}\mathbb{P}^2} p_1(T\mathbb{C}\mathbb{P}^2) = 3$.

We can complexify a real vector bundle V to construct an associated complex vector bundle $V_{\mathbb{C}}$. We have

$$p_i(V) := (-1)^i c_{2i}(V_{\mathbb{C}})$$

Conversely, if V is a complex vector bundle, we can construct an underlying real vector bundle $V_{\mathbb{R}}$ by forgetting the underlying complex structure. Modulo elements of order 2, we have

$$p(V_{\mathbb{R}}) = c(V)c(V^*)$$

Let $T\mathbb{C}\mathbb{P}^k$ be the real tangent bundle of complex projective space. Then

$$p(T\mathbb{C}\mathbb{P}^k) = (1 - x^2)^{k+1}$$

Line Bundles

Tensor product makes $\text{Vect}_1(M, \mathbb{F})$ into an abelian group. One has natural equivalences of functors which are group homomorphisms:

$$\text{sw}_1 : \text{Vect}_1(M, \mathbb{R}) \rightarrow H^1(M; \mathbb{Z}_2)$$

$$c_1 : \text{Vect}_1(M, \mathbb{C}) \rightarrow H^2(M; \mathbb{Z})$$

A real line bundle L is trivial if and only if it is orientable or, equivalently, if $\text{sw}_1(L)$ vanishes. A complex line bundle L is trivial if and only if $c_1(L) = 0$. There are nontrivial vector bundles with vanishing Stiefel–Whitney classes of rank $k > 1$. For example, $\text{sw}_i(TS^k) = 0$ for $i > 0$ despite the fact that TS^k is trivial if and only if $k = 1, 3, 7$.

Curvature and Characteristic Classes

de Rham Cohomology

We can replace the coefficient group \mathbb{Z} by \mathbb{C} at the cost of losing information concerning torsion. Thus, we may regard $p_i(V) \in H^{4i}(M; \mathbb{C})$ if V is real or $c_i(V) \in H^{2i}(M; \mathbb{C})$ if V is complex. Let M be a smooth manifold. Let $C^\infty \Lambda^p M$ be the space of smooth p -forms and let

$$d : C^\infty \Lambda^p M \rightarrow C^\infty \Lambda^{p+1} M$$

be the exterior derivative. The de Rham cohomology groups are then defined by

$$H_{\text{deR}}^p(M) := \frac{\ker(d : C^\infty \Lambda^p M \rightarrow C^\infty \Lambda^{p+1} M)}{\text{im}(d : C^\infty \Lambda^{p-1} M \rightarrow C^\infty \Lambda^p M)}$$

The de Rham theorem identifies the topological cohomology groups $H^p(M; \mathbb{C})$ with the de Rham cohomology groups $H_{\text{deR}}^p(M)$ which are given differentially.

Given a connection on V , the Chern–Weyl theory enables us to compute Pontrjagin and Chern classes in de Rham cohomology in terms of curvature.

Connections

Let V be a vector bundle over M . A connection

$$\nabla : C^\infty(V) \rightarrow C^\infty(T^*M \otimes V)$$

on V is a first-order partial differential operator which satisfies the Leibnitz rule, that is, if s is a smooth section to V and if f is a smooth function on M ,

$$\nabla(fs) = df \otimes s + f \nabla s$$

If X is a tangent vector field, we define

$$\nabla_X s = \langle X, \nabla s \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes the natural pairing between the tangent and cotangent spaces. This generalizes to the bundle setting the notion of a directional derivative and has the properties:

1. $\nabla_{fX} s = f \nabla_X s$.
2. $\nabla_X(fs) = X(f)s + f \nabla_X s$.
3. $\nabla_{X_1+X_2} s = \nabla_{X_1} s + \nabla_{X_2} s$.
4. $\nabla_X(s_1 + s_2) = \nabla_X s_1 + \nabla_X s_2$.

The Curvature 2-Form

Let ω_p be a smooth p -form. Then

$$\nabla : C^\infty(\Lambda^p M \otimes V) \rightarrow C^\infty(\Lambda^{p+1} M \otimes V)$$

can be extended by defining

$$\nabla(\omega_p \otimes s) = d\omega_p \otimes s + (-1)^p \omega_p \wedge \nabla s$$

In contrast to ordinary exterior differentiation, ∇^2 need not vanish. We set

$$\Omega(s) := \nabla^2 s$$

This is not a second-order partial differential operator; it is a zeroth-order operator, that is,

$$\begin{aligned} \Omega(fs) &= ddf \otimes s - df \wedge \nabla s + df \wedge \nabla s + f \nabla^2 s \\ &= f \Omega(s) \end{aligned}$$

The curvature operator Ω can also be computed locally. Let (s_i) be a local frame. Expand

$$\nabla s_i = \sum_j \omega_i^j \otimes s_j$$

to define the connection 1-form ω . One then has

$$\nabla^2 s_i = \left(d\omega_i^j - \omega_i^k \wedge \omega_k^j \right) \otimes s_k$$

and so

$$\Omega_i^j = d\omega_i^j - \omega_i^k \wedge \omega_k^j$$

If $\tilde{s} = \psi_i^j s_j$ is another local frame, we compute

$$\tilde{\omega} = dg g^{-1} + g \omega g^{-1} \quad \text{and} \quad \tilde{\Omega} = g \Omega g^{-1}$$

Although the connection 1-form ω is not tensorial, the curvature is an invariantly defined 2-form-valued endomorphism of V .

Unitary Connections

Let (\cdot, \cdot) be a nondegenerate Hermitian inner product on V . We say that ∇ is a unitary connection if

$$(\nabla s_1, s_2) + (s_1, \nabla s_2) = d(s_1, s_2)$$

Such connections always exist and, relative to a local orthonormal frame, the curvature is skew-symmetric, that is,

$$\Omega + \Omega^* = 0$$

Thus, Ω can be regarded as a 2-form-valued element of the Lie algebra of the structure group, $O(V)$ in the real setting or $U(V)$ in the complex setting.

Projections

We can always embed V in a trivial bundle 1^ν of dimension ν ; let π_V be the orthogonal projection on V . We project the flat connection to V to define a natural connection on V . For example, if M is embedded isometrically in the Euclidean space \mathbb{R}^n , this construction gives the Levi-Civita connection on the tangent bundle TM . The curvature of this connection is then given by

$$\Omega = \pi_V d\pi_V d\pi_V$$

Let V_P be the fiber of V over a point $P \in M$. The inclusion $i : V \subset \mathbb{R}^n$ defines the classifying map $f : P \rightarrow Gr_k(\mathbb{R}, n)$ where we set

$$f(P) = i(V_P)$$

Chern–Weyl Theory

Let ∇ be a Riemannian connection on a real vector bundle V of rank k . We set

$$p(\Omega) := \det\left(I + \frac{1}{2\pi}\Omega\right)$$

Let Ω^T denote the transpose matrix of differential form. Since $\Omega + \Omega^T = 0$, the polynomials of odd degree in Ω vanish and we may expand

$$p(\Omega) = 1 + p_1(\Omega) + \dots + p_r(\Omega)$$

where $k = 2r$ or $k = 2r + 1$ and the differential forms $p_i(\Omega) \in C^\infty \Lambda^{4i}(M)$ are forms of degree $4i$.

Changing the gauge (i.e., the local frame) replaces Ω by $g\Omega g^{-1}$ and hence $p(\Omega)$ is independent of the local frame chosen. One can show that $dp_i(\Omega) = 0$; let $[p_i(\Omega)]$ denote the corresponding element of de Rham cohomology. This is independent of the particular connection chosen and $[p_i(\Omega)]$ represents $p_i(V)$ in $H^{4i}(M; \mathbb{C})$.

Similarly, let V be a complex vector bundle of rank k with a Hermitian connection ∇ . Set

$$\begin{aligned} c(\Omega) &:= \det\left(I + \frac{\sqrt{-1}}{2\pi}\Omega\right) \\ &= 1 + c_1(\Omega) + \dots + c_k(\Omega) \end{aligned}$$

Again $c_i(\Omega)$ is independent of the local gauge and $dc_i(\Omega) = 0$. The de Rham cohomology class $[c_i(\Omega)]$ represents $c_i(V)$ in $H^{2i}(M; \mathbb{C})$.

The Chern Character

The total Chern character is defined by the formal sum

$$\begin{aligned} \text{ch}(\Omega) &:= \text{tr}(e^{\sqrt{-1}\Omega/2\pi}) \\ &= \sum_{\nu} \frac{(\sqrt{-1})^\nu}{(2\pi)^\nu \nu!} \text{tr}(\Omega^\nu) \\ &= \text{ch}_0(\Omega) + \text{ch}_1(\Omega) + \dots \end{aligned}$$

Let $\text{ch}(V) = [\text{ch}(\Omega)]$ denote the associated de Rham cohomology class; it is independent of the particular connection chosen. We then have the relations

$$\begin{aligned} \text{ch}(V \oplus W) &= \text{ch}(V) + \text{ch}(W) \\ \text{ch}(V \otimes W) &= \text{ch}(V)\text{ch}(W) \end{aligned}$$

The Chern character extends to a ring isomorphism from $KU(M) \otimes \mathbb{Q}$ to $H^e(M; \mathbb{Q})$, which is a natural equivalence of functors; modulo torsion, K theory and cohomology are the same functors.

Other Characteristic Classes

The Chern character is defined by the exponential function. There are other characteristic classes which appear in the index theorem that are defined using other generating functions that appear in index theory. Let $\mathbf{x} := (x_1, \dots)$ be a collection of indeterminates. Let $s_\nu(\mathbf{x})$ be the ν th elementary symmetric function;

$$\prod_{\nu} (1 + x_\nu) = 1 + s_1(\mathbf{x}) + s_2(\mathbf{x}) + \dots$$

For a diagonal matrix $A := \text{diag}(\lambda_1, \dots)$, denote the normalized eigenvalues by $x_j := \sqrt{-1}\lambda_j/2\pi$. Then

$$c(A) = \det\left(1 + \frac{\sqrt{-1}}{2\pi}A\right) = 1 + s_1(\mathbf{x}) + \dots$$

Thus, the Chern class corresponds in a certain sense to the elementary symmetric functions.

Let $f(\mathbf{x})$ be a symmetric polynomial or more generally a formal power series which is symmetric. We can express $f(\mathbf{x}) = F(s_1(\mathbf{x}), \dots)$ in terms of the elementary symmetric functions and define $f(\Omega) = F(c_1(\Omega), \dots)$ by substitution. For example, the Chern character is defined by the generating function

$$f(\mathbf{x}) := \sum_{\nu=1}^k e^{x_\nu}$$

The Todd class is defined using a different generating function:

$$\begin{aligned} \text{td}(\mathbf{x}) &:= \prod_{\nu} x_\nu (1 - e^{-x_\nu})^{-1} \\ &= 1 + \text{td}_1(\mathbf{x}) + \dots \end{aligned}$$

If V is a real vector bundle, we can define some additional characteristic classes similarly. Let $\{\pm\sqrt{-1}\lambda_1, \dots\}$ be the nonzero eigenvalues of a skew-symmetric matrix A . We set $x_j = -\lambda_j/2\pi$ and define the Hirzebruch polynomial L and the \hat{A} genus by

$$\begin{aligned} L(\mathbf{x}) &:= \prod_{\nu} \frac{x_\nu}{\tanh(x_\nu)} \\ &= 1 + L_1(\mathbf{x}) + L_2(\mathbf{x}) + \dots \\ \hat{A}(\mathbf{x}) &:= \prod_{\nu} \frac{x_\nu}{2 \sinh((1/2)x_\nu)} \\ &= 1 + \hat{A}_1(\mathbf{x}) + \hat{A}_2(\mathbf{x}) + \dots \end{aligned}$$

The generating functions

$$\frac{x}{\tanh(x)} \quad \text{and} \quad \frac{x}{2 \sinh((1/2)x)}$$

are even functions of x , so the ambiguity in the choice of sign in the eigenvalues plays no role. This defines characteristic classes

$$L_i(V) \in H^{4i}(M; \mathbb{C}) \quad \text{and} \quad \hat{A}_i(V) \in H^{4i}(M; \mathbb{C})$$

Summary of Formulas

We summarize below some of the formulas in terms of characteristic classes:

1. $c_1(\Omega) = \frac{\sqrt{-1}\text{tr}(\Omega)}{2\pi},$
2. $c_2(\Omega) = \frac{1}{8\pi^2} \{\text{tr}(\Omega^2) - \text{tr}(\Omega)^2\},$
3. $p_1(\Omega) = -\frac{1}{8\pi^2} \text{tr}(\Omega^2),$
4. $\text{ch}(V) = k + \left\{ c_1 + \frac{c_1^2 - 2c_2}{2} + \dots \right\} (V),$
5. $\text{td}(V) = \left\{ 1 + \frac{c_1}{2} + \frac{(c_1^2 + c_2)}{12} + \frac{c_1 c_2}{24} + \dots \right\} (V),$
6. $\hat{A}(V) = \left\{ 1 - \frac{p_1}{24} + \frac{7p_1^2 - 4p_2}{5760} + \dots \right\} (V),$
7. $L(V) = \left\{ 1 + \frac{p_1}{3} + \frac{7p_2 - p_1^2}{45} + \dots \right\} (V),$
8. $\text{td}(V \oplus W) = \text{td}(V)\text{td}(W),$
9. $\hat{A}(V \oplus W) = \hat{A}(V)\hat{A}(W),$
10. $L(V \oplus W) = L(V)L(W).$

The Euler Form

So far, this article has dealt with the structure groups $O(k)$ in the real setting and $U(k)$ in the complex setting. There is one final characteristic class which arises from the structure group $SO(k)$. Suppose $k = 2n$ is even. While a real antisymmetric matrix A of shape $2n \times 2n$ cannot be diagonalized, it can be put in block off 2-diagonal form with blocks,

$$\begin{pmatrix} 0 & \lambda_\nu \\ -\lambda_\nu & 0 \end{pmatrix}$$

The top Pontrjagin class $p_n(A) = x_1^2 \cdots x_n^2$ is a perfect square. The Euler class

$$e_{2n}(A) := x_1 \cdots x_n$$

is the square root of p_n . If V is an oriented vector bundle of dimension $2n$, then

$$e_{2n}(V) \in H^{2n}(M; \mathbb{C})$$

is a well-defined characteristic class satisfying $e_{2n}(V)^2 = p_n(V)$.

If V is the underlying real oriented vector bundle of a complex vector bundle W ,

$$e_{2n}(V) = c_n(W)$$

If M is an even-dimensional manifold, let $e_m(M) := e_m(TM)$. If we reverse the local orientation of M , then $e_m(M)$ changes sign. Consequently, $e_m(M)$ is a measure rather than an m -form; we can use the Riemannian measure on M to regard $e_m(M)$ as a scalar. Let R_{ijkl} be the components of the curvature of the Levi-Civita connection with respect to some local orthonormal frame field; we adopt the convention that $R_{1221} = 1$ on the standard sphere S^2 in \mathbb{R}^3 . If $\varepsilon^{IJ} := (e^I, e^J)$ is the totally antisymmetric tensor, then

$$e_{2n} := \sum_{I,J} \frac{\varepsilon^{IJ} R_{i_1 i_2 j_1 \dots j_{n-1} i_{n-1} i_n j_n}}{(8\pi)^n n!}$$

Let $\mathcal{R} := R_{ijji}$ and $\rho_{ij} := R_{ikkj}$ be the scalar curvature and the Ricci tensor, respectively. Then

$$e_2 = \frac{1}{4\pi} \mathcal{R}$$

$$e_4 = \frac{1}{32\pi^2} (\mathcal{R}^2 - 4|\rho|^2 + |R|^2)$$

Characteristic Classes of Principal Bundles

Let \mathfrak{g} be the Lie algebra of a compact Lie group G . Let $\pi : \mathcal{P} \rightarrow M$ be a principal G bundle over M . For $\xi \in \mathcal{P}$, let

$$\mathcal{V}_\xi := \ker \pi_* : T_\xi \mathcal{P} \rightarrow T_{\pi\xi} M \quad \text{and} \quad \mathcal{H}_\xi := \mathcal{V}_\xi^\perp$$

be the vertical and horizontal distributions of the projection π , respectively. We assume that the metric on \mathcal{P} is chosen to be G -invariant and such that $\pi_* : \mathcal{H}_\xi \rightarrow T_{\pi\xi} M$ is an isometry; thus, π is a Riemannian submersion. If F is a tangent vector field on M , let $\mathcal{H}F$ be the corresponding vertical lift. Let $\rho_\mathcal{V}$ be orthogonal projection on the distribution \mathcal{V} . The curvature is defined by

$$\Omega(F_1, F_2) = \rho_\mathcal{V}[\mathcal{H}(F_1), \mathcal{H}(F_2)]$$

the horizontal distribution \mathcal{H} is integrable if and only if the curvature vanishes. Since the metric is G -invariant, $\Omega(F_1, F_2)$ is invariant under the group action. We may use a local section s to \mathcal{P} over a contractible coordinate chart \mathcal{O} to split $\pi^{-1}\mathcal{O} = \mathcal{O} \times G$. This permits us to identify \mathcal{V} with TG and to regard Ω as a \mathfrak{g} -valued 2-form. If we replace the section s by a section \tilde{s} , then $\tilde{\Omega} = g\Omega g^{-1}$ changes by the adjoint action of G on \mathfrak{g} .

If V is a real or complex vector bundle over M , we can put a fiber metric on V to reduce the structure group to the orthogonal group $O(r)$ in the real setting or the unitary group $U(r)$ in the complex setting. Let \mathcal{P}_V be the associated frame bundle. A Riemannian connection ∇ on V induces an invariant splitting of $T\mathcal{P}_V = \mathcal{V} \oplus \mathcal{H}$ and defines a natural

metric on \mathcal{P}_V ; the curvature Ω of the connection ∇ defined here agrees with the definition previously.

Let $\mathcal{Q}(G)$ be the algebra of all polynomials on \mathfrak{g} which are invariant under the adjoint action. If $Q \in \mathcal{Q}(G)$, then $Q(\Omega)$ is well defined. One has $dQ(\Omega) = 0$. Furthermore, the de Rham cohomology class $Q(P) := [Q(\Omega)]$ is independent of the particular connection chosen. We have

$$\begin{aligned} Q(U(k)) &= \mathbb{C}[c_1, \dots, c_k] \\ Q(SU(k)) &= \mathbb{C}[c_2, \dots, c_k] \\ Q(O(2k)) &= \mathbb{C}[p_1, \dots, p_k] \\ Q(O(2k+1)) &= \mathbb{C}[p_1, \dots, p_k] \\ Q(SO(2k)) &= \mathbb{C}[p_1, \dots, p_k, e_k]/e_k^2 = p_k \\ Q(SO(2k+1)) &= \mathbb{C}[p_1, \dots, p_k] \end{aligned}$$

Thus, for this category of groups, no new characteristic classes ensue. Since the invariants are Lie-algebra theoretic in nature,

$$Q(\text{Spin}(k)) = Q(\text{SO}(k))$$

Other groups, of course, give rise to different characteristic rings of invariants.

Acknowledgments

Research of P Gilkey was partially supported by the MPI (Leipzig, Germany), that of R Ivanova by the UHH Seed Money Grant, and of S Nikčević by MM 1646 (Serbia), DAAD (Germany), and Dierks von Zweck Stiftung (Esen, Germany).

See also: Cohomology Theories; Gerbes in Quantum Field theory; Instantons: Topological Aspects; K -Theory; Mathai-Quillen Formalism; Riemann Surfaces.

Further Reading

Besse AL (1987) Einstein manifolds. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)* [Results in Mathematics and Related Areas (3)], p. 10. Berlin: Springer-Verlag.

- Bott R and Tu LW (1982) Differential forms in algebraic topology. *Graduate Texts in Mathematics*, p. 82. New York–Berlin: Springer-Verlag.
- Chern S (1944) A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds. *Annals of Mathematics* 45: 747–752.
- Chern S (1945) On the curvatura integra in a Riemannian manifold. *Annals of Mathematics* 46: 674–684.
- Conner PE and Floyd EE (1964) Differentiable periodic maps. *Ergebnisse der Mathematik und ihrer Grenzgebiete, N.F., Band 33*. New York: Academic Press; Berlin–Göttingen–Heidelberg: Springer-Verlag.
- de Rham G (1950) Complexes à automorphismes et homéomorphie différentiable (French). *Ann. Inst. Fourier Grenoble* 2: 51–67.
- Eguchi T, Gilkey PB, and Hanson AJ (1980) Gravitation, gauge theories and differential geometry. *Physics Reports* 66: 213–393.
- Eilenberg S and Steenrod N (1952) *Foundations of Algebraic Topology*. Princeton, NJ: Princeton University Press.
- Greub W, Halperin S, and Vanstone R (1972) *Connections, Curvature, and Cohomology. Vol. I: De Rham Cohomology of Manifolds and Vector Bundles*. Pure and Applied Mathematics, vol. 47. New York–London: Academic Press.
- Hirzebruch F (1956) Neue topologische Methoden in der algebraischen Geometrie (German). *Ergebnisse der Mathematik und ihrer Grenzgebiete (N.F.)*, Heft 9. Berlin–Göttingen–Heidelberg: Springer-Verlag.
- Husemoller D (1966) *Fibre Bundles*. New York–London–Sydney: McGraw-Hill.
- Karoubi M (1978) *K-theory. An introduction*. Grundlehren der Mathematischen Wissenschaften, Band 226. Berlin–New York: Springer-Verlag.
- Kobayashi S (1987) *Differential Geometry of Complex Vector Bundles*. Publications of the Mathematical Society of Japan, 15. Kanô Memorial Lectures, 5. Princeton, NJ: Princeton University Press; Tokyo: Iwanami Shoten.
- Milnor JW and Stasheff JD (1974) *Characteristic Classes*. Annals of Mathematics Studies, No. 76. Princeton, NJ: Princeton University Press; Tokyo: University of Tokyo Press.
- Steenrod NE (1962) *Cohomology Operations*. Lectures by NE Steenrod written and revised by DBA Epstein. Annals of Mathematics Studies, No. 50. Princeton, NJ: Princeton University Press.
- Steenrod NE (1951) *The Topology of Fibre Bundles*. Princeton Mathematical Series, vol. 14. Princeton, NJ: Princeton University Press.
- Stong RE (1968) *Notes on Cobordism Theory*. Mathematical Notes. Princeton, NJ: Princeton University Press; Tokyo: University of Tokyo Press.
- Weyl H (1939) *The Classical Groups. Their Invariants and Representations*. Princeton, NJ: Princeton University Press.

Chern–Simons Models: Rigorous Results

A N Sengupta, Louisiana State University,
Baton Rouge, LA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The relationship between topological invariants and functional integrals from quantum Chern–Simons theory discovered by Witten (1989) raised several

challenges for mathematicians. Most of the tremendous amount of mathematical activity generated by Witten’s discovery has been concerned primarily with issues that arise after one has accepted the functional integral as a formal object. This has left, as an important challenge, the task of giving rigorous meaning to the functional integrals themselves and to rigorously derive their relation to topological invariants. The present article will discuss efforts to put the functional integral itself on a rigorous basis.

Chern–Simons Functional Integrals

We shall describe here the typical Chern–Simons functional integral. For the purposes of this article, we will confine ourselves to a simpler setting rather than the most general possible one. In fact, we shall work with fields over three-dimensional Euclidean space \mathbb{R}^3 (instead of a general 3-manifold).

The typical Chern–Simons functional integral is of the form

$$\int_{\mathcal{A}} e^{i(k/4\pi)S_{CS}(A)} W_{C_1, R_1}(A) \dots W_{C_n, R_n}(A) DA \quad [1]$$

Our objective in this section will be to specify what the terms in this formal integral mean. Very briefly, the integration is with respect to a formal “Lebesgue measure” on \mathcal{A} , an infinite-dimensional space of geometric objects A called connections over \mathbb{R}^3 with values in the Lie algebra LG of a group G . In the first term in the integrand, in the exponent, k is a real number, and $S_{CS}(A)$ is the Chern–Simons action for the connection A . Each term $W_{C_i, R_i}(A)$ is a Wilson loop observable, the trace in some representation R_i of the holonomy of the connection A around the loop C_i . The entire integral, formal though it may be, provides an invariant associated with the system of loops C_1, \dots, C_n .

Let G be a compact Lie group; for ease of exposition, let us take G to be a closed, connected subgroup of $U(n)$. Thus, each element of G is an $n \times n$ complex matrix g with $g^*g = I$, the identity. The Lie algebra LG consists of all $n \times n$ matrices A which are skew-Hermitian, that is, satisfy $A^* = -A$, and for which $e^{tA} \in G$ for all real numbers t . On LG there is a convenient inner product given by

$$\langle A, B \rangle = \text{tr}(AB^*)$$

This inner product is invariant under the conjugation action of the group G on its Lie algebra LG .

By a connection over \mathbb{R}^3 we shall mean a C^∞ 1-form with values in LG . The set of all connections is an affine (in our case, actually a linear) space \mathcal{A} . If $A \in \mathcal{A}$, then define

$$S_{CS}(A) = \int_{\mathbb{R}^3} \text{tr}(A \wedge dA + \frac{2}{3}A \wedge A \wedge A) \quad [2]$$

This is, up to constant multiple, the Chern–Simons action functional.

Let A be a connection and consider a piecewise smooth path

$$C : [0, 1] \rightarrow \mathbb{R}^3$$

With this one can associate a G -valued path $[0, 1] \rightarrow G : t \mapsto g(t) \in G$ satisfying the differential equation

$$g'(t)g(t)^{-1} = -A(C'(t))$$

subject to the initial condition $g(0) = I$, the identity. The path $t \mapsto g(t)$ describes parallel transport along C by the connection A . If C is a loop then the final value $g(1)$ is the holonomy of A around C . If R is a representation of G on some finite-dimensional vector space then the trace of $R(g(1))$ is the Wilson loop observable:

$$W_{C,R}(A) = \text{tr}(R(g(1))) \quad [3]$$

Thus, we have specified the meaning of the terms appearing in the formal integral [1], where C_1, \dots, C_n of eqn [1] form a link (a family of nonintersecting, imbedded loops) in \mathbb{R}^3 and R_1, \dots, R_n are finite-dimensional representations of G . Witten showed that, at least for suitable values of k , integrals of this form ought to produce topological invariants, which he identified, for the link.

The integral [1] is problematic for several reasons. First, there is no reasonable and useful analog of Lebesgue measure on an infinite-dimensional space. Even if one were to regularize this measure in some simple way, one would run into the problem that the measure would not live on the space of smooth connections, and so the integrand would become meaningless.

There are several different approaches to a mathematical interpretation of [1]. The approach that is often taken in practice is to simply ignore the analytical problem and define the value of the integral [1] to be what Witten’s calculations have given. One approach, used, for instance, by [BarNatan \(1995\)](#) is to expand the integrand in a series and relate each individual integral in this expansion separately to topological invariants. Discrete approximation procedures to the continuum integral have also been explored. In the abelian case, infinite-dimensional oscillatory integral techniques have been used to understand the functional integral. [Fröhlich and King \(1999\)](#) showed the possibility of interpreting parallel transport using ideas from stochastic differential equations. Such an approach has been used successfully in the case of two-dimensional Yang–Mills theory, where the functional integral actually corresponds to integration with respect to a measure. In this article, we focus on a method of understanding the normalized Chern–Simons functional integral in terms of infinite-dimensional distribution theory and examining some ideas for understanding Wilson loop expectation values in this setting.

Infinite Dimensional Distributions

Let (x^0, x^1, x^2) denote the usual coordinates on \mathbb{R}^3 . Gauge symmetry, an issue which will not be examined here, may be used to simplify the problem of the Chern–Simons integral. In particular, one

need only focus on connections which vanish in the x^2 -direction, that is, connections of the form $A = A_0 dx^0 + A_1 dx^1$. For such A , the triple wedge-product term in the Chern–Simons action disappears, and we are left with the quadratic expression:

$$S_{CS}(A) = \int_{\mathbb{R}^3} \text{tr}(A \wedge dA) \quad [4]$$

This is good, since the functional integral now involves a quadratic exponent and so stands a good chance of rigorous realization, just as Gaussian measure can be given rigorous meaning in infinite dimensions. However, in the Chern–Simons situation, there is no hope of actually getting a measure, not even a complex measure.

The next best thing to a measure is a distribution or “generalized function.” A distribution over a space Y is a continuous linear functional on a topological vector space of functions on Y . Thus, the objective is to realize the Chern–Simons functional integral as a continuous linear functional on some space of test functions over \mathcal{A} (more precisely, on an extension of \mathcal{A}). Before turning to the specific case of the Chern–Simons integral, let us examine some elements of the theory of infinite-dimensional distributions, in as much as they are relevant to our needs.

Let us consider a Hilbert space \mathcal{E}_0 , and a positive Hilbert–Schmidt operator T on \mathcal{E}_0 . For each integer $p \geq 0$, let $\mathcal{E}_p = T^p(\mathcal{E}_0)$, which is a Hilbert space with the inner product $\langle x, y \rangle_p = \langle T^{-p}x, T^{-p}y \rangle$. Then we have the chain of inclusions

$$\mathcal{E} = \bigcap_{p \geq 1} \mathcal{E}_p \subset \cdots \subset \mathcal{E}_2 \subset \mathcal{E}_1 \subset \mathcal{E}_0 \quad [5]$$

with each inclusion $\mathcal{E}_{p+1} \rightarrow \mathcal{E}_p$ being Hilbert–Schmidt. Let $\mathcal{E}_{-p} = \mathcal{E}'_p$ be the topological dual of \mathcal{E}_p , the space of continuous linear functionals on \mathcal{E}_p , and let \mathcal{E}' be the topological dual of \mathcal{E} , where the latter is given the topology generated by all the norms $\|\cdot\|_p$. Then we have the inclusions

$$\mathcal{E}_0 \simeq \mathcal{E}'_0 \subset \mathcal{E}_{-1} \subset \mathcal{E}_{-2} \subset \cdots \subset \mathcal{E}' = \bigcup_{p \geq 0} \mathcal{E}_{-p} \quad [6]$$

For each $x \in \mathcal{E}$ there is the evaluation map $\hat{x}: \mathcal{E}' \rightarrow \mathbb{R}: \phi \mapsto \phi(x)$. A very special case of a general theorem of Minlos guarantees that on the dual \mathcal{E}' there is a measure μ on the sigma algebra generated by all the functions \hat{x} such that each \hat{x} is a Gaussian random variable of mean zero and variance $|x|_0^2$, that is,

$$\int_{\mathcal{E}'} e^{it\hat{x}} d\mu = e^{-t^2|x|_0^2/2}$$

for all $x \in \mathcal{E}$ and $t \in \mathbb{R}$. This measure μ is the standard Gaussian measure on \mathcal{E}' for the infinite-dimensional nuclear space \mathcal{E} .

The inner products $\langle \cdot, \cdot \rangle_p$ give rise to a nuclear space structure on function spaces over \mathcal{E} . Let \mathcal{U} be the algebra of functions on \mathcal{E}' generated by the exponentials $e^{\lambda \hat{x}}$, with x running over \mathcal{E} and λ over \mathbb{C} . For each $p \geq 0$, there is an inner product $\langle \langle \cdot, \cdot \rangle \rangle_p$ on \mathcal{U} such that

$$\left\langle \left\langle e^{\lambda \hat{x} - \lambda^2 |x|_p^2/2}, e^{\mu \hat{y} - \mu^2 |y|_p^2/2} \right\rangle \right\rangle_p = e^{\lambda \bar{\mu} \langle x, y \rangle_p} \quad [7]$$

For $p=0$ the left-hand side coincides with the $L^2(\mu)$ inner product. Let $[\mathcal{E}]_p$ be the Hilbert space completion of \mathcal{U} in the $\langle \langle \cdot, \cdot \rangle \rangle_p$ inner product. Then

$$\cdots [\mathcal{E}]_3 \subset [\mathcal{E}]_2 \subset [\mathcal{E}]_1 \subset [\mathcal{E}]_0 = L^2(\mathcal{E}', \mu) \quad [8]$$

Let $[\mathcal{E}] = \bigcap_{p \geq 0} [\mathcal{E}]_p$, equipped with topology from all the norms $\|\cdot\|_p$, and $[\mathcal{E}]'$ its topological dual. Elements of $[\mathcal{E}]'$, being continuous linear functionals on the “test function space” $[\mathcal{E}]$, are called distributions over \mathcal{E} , in the language of white-noise analysis.

A fundamental tool in the study of infinite-dimensional distributions is the S -transform. This generalizes the traditional Segal–Bargmann transform from the L^2 -setting to the context of distributions. Let \mathcal{E}_c be the complexification of \mathcal{E} . The inner product $\langle \cdot, \cdot \rangle_0$ on \mathcal{E} extends to a complex-bilinear pairing $\mathcal{E}_c \times \mathcal{E}_c \rightarrow \mathbb{C}: (z, w) \mapsto z \cdot w$. The evaluation pairing $\mathcal{E}' \times \mathcal{E} \rightarrow \mathbb{R}$ also extends naturally to the complexifications. For Φ a distribution belonging to $[\mathcal{E}]'$, define a function $S\Phi$ on \mathcal{E} by

$$S\Phi(z) = \Phi(c_z)$$

for all $z \in \mathcal{E}_c$. Here c_z is the coherent state function on \mathcal{E}' given by $c_z(\phi) = e^{\phi(z) - (1/2)z \cdot z}$. A fundamental and useful result in white-noise analysis, due originally to Pothoff and Streit, specifies the range of the transform S and allows reconstruction of a distribution Φ from the function $S\Phi$. Briefly, the range of S consists of functions which are holomorphic, in an appropriate sense, and have at most quadratic exponential growth. In particular, this theorem implies that a function of the form $z \mapsto e^{az \cdot z}$, for any constant a , is in the range of Φ .

Rigorous Realization of Chern–Simons Integrals

We return to the Chern–Simons context. As mentioned earlier, gauge symmetry may be invoked to reduce the space of connections to the smaller space:

$$\mathcal{E} = X \oplus X \quad [9]$$

where $X = \mathcal{S}(\mathbb{R}^3) \otimes LG$ is the space of rapidly decreasing functions with values in the Lie algebra LG . Let

$$T_1 = \left(-\frac{d^2}{dx^2} + \frac{x^2}{4} \right)^{-1}$$

as a linear operator on $L^2(\mathbb{R}^3)$, $T_2 = T_1^{\otimes 3} \otimes I$ the induced operator on $L^2(\mathbb{R}^3) \otimes LG$, and $T = T_2 \oplus T_2$. Then, as described in the preceding section, we have the space \mathcal{E} and its dual \mathcal{E}' . There is then the standard Gaussian measure μ on \mathcal{E}' , and the space $[\mathcal{E}]'$ of distributions over \mathcal{E}' .

The normalized Chern–Simons integral may be viewed as a linear functional

$$\Phi_{CS} : F \mapsto \frac{1}{N} \int_{\mathcal{E}} e^{i(k/4\pi)S_{CS}(A)} F(A) D A \quad [10]$$

where N is a “normalizing” factor. Rigorous meaning can be given to this by first formally working out what the S -transform of Φ_{CS} ought to be. Calculation shows that $S\Phi$ is indeed a holomorphic function on \mathcal{E}_c of quadratic growth. The Potthoff–Streit theorem then implies that Φ_{CS} does exist as a distribution in the space $[\mathcal{E}]'$. Let us examine this in some more detail.

As before, we take A to be of the form $A = A_0 dx^0 + A_1 dx^1$, with the component A_2 equal to 0. Integration by parts shows that

$$\frac{k}{4\pi} S_{CS}(A) = -\frac{k}{2\pi} \int_{\mathbb{R}^3} \text{tr}(A_0 \partial_2 A_1) \text{dvol} \quad [11]$$

A formal computation reveals that $S(\Phi_{CS})(j)$ should be given by

$$\exp\left(\frac{2\pi i}{k} \text{tr}(j_0 \partial_2^{-1} j_1)\right) \quad [12]$$

where $j = (j_0, j_1)$, and

$$\partial_2^{-1} f(x) = \frac{1}{2} \int ds [1_{(-\infty, x_2]}(s) - 1_{[x_2, \infty)}(s)] f(x^0, x^1, s)$$

The Potthoff–Streit criterion implies the existence of a distribution Φ_{CS} , whose S -transform is given by the above expression.

The distribution Φ_{CS} is, however, not a sufficiently powerful object to allow determination of the Wilson loop expectations that one would really like to have. For instance, Φ_{CS} does not live on the space of smooth connections and so the meaning of parallel transport needs to be defined. The state of knowledge, at the rigorous level, at this point is still evolving, with progress reported by A. Hahn. We describe some ideas for the Wilson loop expectations in the following.

The strategy for defining parallel transport along a path is to smear out the path by means of bump functions and essentially replace the path by a path of test functions in \mathcal{E} . The description given here is mainly for the case of abelian G . Choose first a C^∞ non-negative bump function ψ on \mathbb{R}^3 , vanishing outside the unit ball and having L^1 norm equal to 1. For $\epsilon > 0$, let ψ^ϵ be the scaled bump function given

by $\psi^\epsilon(x) = \epsilon^{-3} \psi(x/\epsilon)$. Next, for a smooth loop $[0, 1] \rightarrow l(t) = (l_0(t), l_1(t), l_2(t))$, let $l^\epsilon(t) = \psi^\epsilon(\cdot - l(t))$, the scaled bump function centered now at the path point $l(t)$. Now consider a generalized connection $A = (A_0, A_1) \in \mathcal{E}'$. Set

$$B_A^\epsilon(t) = A_0(l^\epsilon(t))l'(t)_0 + A_1(l^\epsilon(t))l'(t)_1 \quad [13]$$

The equation of parallel transport can be reformulated as a differential equation for a matrix-valued path $t \mapsto P_A^\epsilon(t)$ satisfying

$$\frac{d}{dt} P_A^\epsilon(t) + B_A^\epsilon(t) P_A^\epsilon(t) = 0 \quad [14]$$

and the initial condition $P_A^\epsilon(t) = I$. With this smearing, one can consider functions of the form

$$W_\epsilon(L; A) = \prod_{i=1}^n \text{tr}(P_A^\epsilon(A)) \quad [15]$$

for a link L consisting of loops l_1, \dots, l_n , instead of the classical Wilson loop variable.

At this stage, it would be natural to consider taking $\epsilon \downarrow 0$ in $\Phi(W_\epsilon(L))$. However, this is still problematic. A further regularization is needed, roughly corresponding to the geometric notion of framing. In the definition of Φ_{CS} , alteration is made to the quadratic form $Q(j, j)$ in the exponent which appears in the expression for $S(\Phi_{CS})$, replacing it with $Q(j, \phi_s^* j)$, where $\{\phi_s\}_{s>0}$ is a family of suitable diffeomorphisms of \mathbb{R}^3 , with ϕ_0 being the identity. In a sense, this splits a single loop l into l and a neighboring loop $\phi_s \circ l$. At the end, one has to take $s \downarrow 0$. The resulting limiting value is the expected link-invariant. We shall not go into the case of nonabelian G , which is more complex, for which work continues to be in progress.

Infinite-dimensional distributions can be used to formulate a rigorous theory for normalized Chern–Simons functional integrals. The more specific questions raised by the Wilson-loop integrals in this setting opens up new problems for further developments in the distribution theory, connecting geometry, topology, and infinite-dimensional analysis.

Acknowledgments

This research is supported by US NSF grant DMS-0201683.

See also: BF Theories; Feynman Path Integrals; Fractional Quantum Hall Effect; Knot Theory and Physics; Large- N and Topological Strings; Large- N Dualities; Quantum 3-Manifold Invariants; Quantum Hall Effect; Spin Foams; String Field Theory; Topological Quantum Field Theory: Overview; Twistor Theory: Some Applications.

Further Reading

- Albeverio S, Hahn A, and Sengupta AN (2003) Chern–Simons theory, Hida distributions, and state models. *Infinite Dimensional Analysis Quantum Probability and Related Topics* 6: 65–81.
- Albeverio S and Schäfer J (1994) Abelian Chern–Simons theory and linking numbers via oscillatory integrals. *Journal of Mathematical Physics (N.Y.)* 36 (suppl. 5): 2135–2169.
- Albeverio S and Sengupta A (1997) A mathematical construction of the non-Abelian Chern–Simons functional integral. *Communications in Mathematical Physics* 186: 563–579.
- Altschuler D and Freidel L (1997) Vassiliev Knot invariants and Chern–Simons perturbation theory to all orders. *Communications in Mathematical Physics* 187: 261–287.
- Atiyah M (1990) *The Geometry and Physics of Knot Polynomials*. Cambridge: Cambridge University Press.
- Bar-Natan D (1995) Perturbative Chern–Simons theory. *Journal of Knot Theory and its Ramifications* 4: 503.
- Chern S-S and Simons J (1974) Characteristic forms and geometric invariants. *Annals of Mathematics* 99: 48–69.
- Fröhlich J and King C (1989) The Chern–Simons theory and Knot polynomials. *Communications in Mathematical Physics* 126: 167–199.
- Kondratiev Yu, Leukert P, Potthoff J, Streit L, and Westerkamp W (1996) Generalized functionals in Gaussian spaces – the characterization theorem revisited. *Journal of Functional Analysis* 141 (suppl. 2): 301–318.
- Kuo H-H (1996) *White Noise Distribution Theory*. Boca Raton, FL: CRC Press.
- Landsman NP, Pflaum M, and Schlichenmaier M (2001) *Quantization of Singular Symplectic Quotients*. Basel–Boston–Berlin: Birkhäuser.
- Leukert P and Schäfer J (1996) A rigorous construction of Abelian Chern–Simons path integrals using White Noise analysis. *Rev. Math. Phys.* 8 (suppl. 3): 445–456.
- Sen Samik, Sen Siddhartha, Sexton JC, and Adams DH (2000) Geometric discretization scheme applied to the Abelian Chern–Simons theory. *Physical Review E* 61: 3174–5185.
- Simon B (1971) Distributions and their Hermite expansions. *Journal of Mathematical Physics (N.Y.)* 12: 140–148.
- Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.

Classical Groups and Homogeneous Spaces

S Gindikin, Rutgers University, Piscataway, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Classical groups are Lie groups corresponding to three classical geometries – linear, metric, and symplectic. Let us start with the complex field \mathbb{C} . We consider the linear space \mathbb{C}^n and the group $GL(n; \mathbb{C})$ of its automorphisms – nondegenerate (invertible) linear transformations. The complex linear metric space is the space \mathbb{C}^n endowed by a nondegenerate symmetric bilinear form; the orthogonal group $O(n; \mathbb{C})$ is the subgroup in $GL(n; \mathbb{C})$ of automorphisms of this structure. If, for $n = 2l$, we replace the symmetric form by a nondegenerate skew-symmetric form, we obtain the linear symplectic space and the group $Sp(l; \mathbb{C})$ of its automorphisms – the symplectic group.

A fundamental observation of nineteenth century geometry was that the transfer from the complex field to the real one, gives not only three corresponding groups for \mathbb{R} but a much richer collection of real forms of complex classical groups: unitary, pseudounitary, pseudoorthogonal, etc. (see below). Classical geometries correspond to homogeneous manifolds with classical groups of transformations. Geometers understood that this produces a very rich world of non-Euclidean geometries, including the first example of non-Euclidean geometry – hyperbolic geometry. Some classical algebraic theories through such an approach obtain a geometrical

interpretation (see below the consideration of the cone of symmetric positive forms). Between classical manifolds there are Minkowski space, Grassmannians, and multidimensional analogs of the disk and the half-plane. A substantial part of this theory is a matrix geometry, which serves as a background for matrix analysis. A rich geometry on classical manifolds with many symmetries is a background for a rich multidimensional analysis with many explicit formulas. Classical geometries, starting with Minkowski geometry, have appeared in some problems of mathematical physics.

A crucial technical fact is the embedding of the classical groups in the class of semisimple Lie groups; it gives a very strong unified method to work with semisimple groups and corresponding geometries – the method of roots. Nevertheless, some special realizations and constructions for classical groups can also be very useful. A very impressive example is the twistors of Penrose, where an initial construction is the realization of points of four-dimensional Minkowski space as lines in three-dimensional complex projective space. We mention below some general facts about semisimple groups and homogeneous manifolds, but the focus will be on special possibilities for the classical groups. The class of simple Lie groups contains, besides the classical groups, only a finite number of exceptional groups which are also very interesting and are connected, in particular, with noncommutative and nonassociative geometries; they have applications to mathematical physics.

Complex Groups and Homogeneous Manifolds

Complex Classical Groups

The complete linear group $GL(n; \mathbb{C})$ is the group of nongenerate matrices g of order n ($\det g \neq 0$) and the special linear group $SL(n; \mathbb{C})$ is its subgroup of matrices with the determinant equal 1 (unimodular condition). The unimodular condition kills the one-dimensional center, perhaps, leaving only a finite center. We realize the direct products of several copies of complete linear groups with different dimensions, for example, $GL(k; \mathbb{C}) \times GL(l; \mathbb{C})$, as the groups of the blockdiagonal nondegenerate matrices. The letter S always means that we take matrices with determinant 1. So the notation $S(L(k; \mathbb{C}) \times L(l; \mathbb{C}))$ means that we take blockdiagonal matrices with blocks of sizes k, l and with the determinant 1.

Let I be a nondegenerate symmetric matrix of order n ; then the orthogonal group $O(n; \mathbb{C})$ is the subgroup in $GL(n; \mathbb{C})$ of matrices preserving the corresponding symmetric form so that

$$g^T I g = I$$

These matrices can have the determinant ± 1 . The special orthogonal group $SO(n; \mathbb{C})$ is the subgroup of orthogonal matrices with determinant 1. Different I 's give isomorphic orthogonal groups since they are all linearly equivalent. If we take as I the unit matrix $E = E_n$, then we receive the group of orthogonal matrices in the classical sense: $g^T g = E$.

If $n = 2l$ and we replace in this definition the symmetric matrix I by a nondegenerate skew-symmetric matrix J , we obtain the symplectic group $Sp(l; \mathbb{C})$. Again, different J 's give isomorphic groups. The typical example of J is

$$J = \begin{pmatrix} 0 & E_l \\ -E_l & 0 \end{pmatrix}$$

It is convenient then to represent matrices g as

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where the blocks A, B, C, D are matrices of order l . Then the symplectic condition is that $A^T D - C^T A = E$ and matrices $A^T C$ and $D^T B$ are symmetric. If $C = 0$ then $D = (A^T)^{-1}$ and $A^{-1} B$ is a symmetric matrix. In this way, we have in $Sp(l; \mathbb{C})$ a subgroup P of blocktriangular matrices of a very simple structure; it is an example of subgroups which are called parabolic.

There are two principal classes of homogeneous spaces with complex semisimple Lie groups: flag manifolds and Stein manifolds.

Flag Manifolds

These homogeneous spaces $F = G/P$ with semi-simple (in our case with classical) groups G have parabolic subgroups P as the isotropy subgroups. The group $G = GL(n; \mathbb{C})$ transitively acts on the flag manifolds $F(n_1, \dots, n_r), 0 < n_1 < \dots < n_r < n$, whose elements are (n_1, \dots, n_r) -flags – sequences of embedded subspaces in \mathbb{C}^n of the dimensions (n_1, \dots, n_r) . The isotropy subgroup $P = P(n_1, \dots, n_r)$ is the subgroup of blocktriangle matrices with the diagonal blocks of sizes $k_1, \dots, k_{r+1}, k_j = (n_j - n_{j-1}), n_0 = 0, n_{r+1} = n$. The flag manifolds are compact complex manifolds. The matrices proportional to the unit matrix E_n act trivially and we can consider instead of the action of $G = GL(n; \mathbb{C})$ the transitive action of $G = SL(n; \mathbb{C})$.

Let us pay particular attention to two extremal cases. The first one is the case of the maximal flag manifold when we have the sequence of all integers $(1, 2, 3, \dots, n - 1)$ – complete flags; the subgroup P in this case is called Borelian. Another case is minimal flag manifolds with $r = 1$ (for them the unipotent radical of the parabolic subgroups is commutative). Then in the case of $SL(n; \mathbb{C})$ the sequence has only one element $n_1 = k < n$ and we have Grassmannian manifolds $Gr_{\mathbb{C}}(k; n) = F(k)$ of k -dimensional subspaces in \mathbb{C}^n . If $k = 1$ or $k = n - 1$, we obtain the dual realizations of the complex projective space CP^{n-1} . We can interpret points of $Gr_{\mathbb{C}}(k; n)$ also as $(k - 1)$ -dimensional planes in CP^{n-1} .

We can define points of the projective space CP^{n-1} by homogeneous coordinates – as the equivalency classes $(z \sim cz, z \in \mathbb{C}^n \setminus \{0\}, c \in \mathbb{C} \setminus \{0\})$. For the Grassmannians we can similarly use matrix homogeneous coordinates (Stiefel's coordinates): classes of $(k \times n)$ -matrices $Z \in Mat(k, n)$ of the maximal rank k relative to the equivalency

$$Z \sim uZ, \quad u \in GL(k; \mathbb{C})$$

The rows of a matrix Z correspond to a base in subspace with the homogeneous coordinate Z ; the left multiplication on a matrix u replaces this base, but does not change the subspace. The group $GL(n; \mathbb{C})$ acts by right multiplications:

$$Z \mapsto Zg$$

and this action preserves the equivalency classes. Suppose $k \leq n - k$ and the left k -minor of Z is not zero. Such matrices give the dense coordinate chart $\mathbb{C}^{k(n-k)}$: we can pick in the equivalency classes the representatives $(E_k, z), z \in Mat(k, n - k)$, and consider the matrices z as (inhomogeneous) local coordinates. In the inhomogeneous coordinates the

action of the group has a matrix fractional linear form: let

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

$$A \in \text{Mat}(k), \quad D \in \text{Mat}(n - k),$$

$$B \in \text{Mat}(k, n - k), \quad C \in \text{Mat}(n - k, k)$$

Then we have the transformation in inhomogeneous coordinates:

$$z \mapsto (A + zC)^{-1}(B + zD)$$

The condition $C=0$ defines the parabolic subgroup which has affine action in inhomogeneous coordinates which is transitive in the coordinate chart. In such a way the Grassmannian is a compactification of $\mathbb{C}^{k(n-k)}$ (realized as a space of $k \times (n - k)$ matrices). If $n = 2k$, we can consider it as the compactification of the space of square matrices z of order k with the flat generalized conformal structure defined by translations of the isotropy cone $\{\det z = 0\}$.

There are similar constructions of flag manifolds for other classical groups. We will consider only the minimal flag manifolds. For $O(2k; \mathbb{C})$ we consider the isotropic Grassmannian $\text{Gr}_{\mathbb{C}}^I(2k; \mathbb{C})$ of isotropic k -subspaces relative to the symmetric form I . We take the matrix realization of $\text{Gr}_{\mathbb{C}}(k; 2k)$, using Stiefel's homogeneous coordinates, and add the matrix equation

$$ZIZ^T = 0$$

which is well defined in the homogeneous coordinates (compatible with the equivalency classes) and defines isotropic subspaces relative to I . This matrix cone is preserved by the subgroup $O(2k; \mathbb{C}) \subset GL(2k; \mathbb{C})$ corresponding to the matrix I . If we take the symmetric matrix

$$I = \begin{pmatrix} 0 & E_k \\ E_k & 0 \end{pmatrix}$$

then in inhomogeneous coordinates (z is a square k -matrix) this equation is transformed into the condition that the matrix z is skew-symmetric. So, in a natural sense, the isotropic Grassmannian is the compactification of the linear space of skew-symmetric matrices $\text{Alt}(k) = \mathbb{C}^N, N = k(k - 1)/2$.

A similar construction makes sense for the symplectic group: if we replace the symmetric form I with the skew-symmetric form J , we obtain the equation of the matrix cone representing the Lagrangian Grassmannian $\text{Gr}_{\mathbb{C}}^L(k; 2k)$ of Lagrangian subspaces in $2k$ -dimensional linear symplectic space. If we were to choose J as above, then in the

(inhomogeneous) coordinate chart we obtain the condition that the matrix z is symmetric. Thus, we have the (dense) coordinate chart on the Lagrangian Grassmannian $\mathbb{C}^N = \text{Sym}(k), N = k(k + 1)/2$ – the linear space of symmetric matrices.

There is one more type of minimal flag manifolds for the orthogonal group $SO(n; \mathbb{C})$ – the quadric \mathcal{Q} in the projective space:

$$I(z) = zIz^T = 0$$

where rows $z \in \mathbb{C}^n \setminus \{0\}$ represent, in homogeneous coordinates, points in $\mathbb{C}P^{n-1}$. If $I = E_n$ we have the equation $(z_1)^2 + \dots + (z_n)^2 = 0$. This quadric is the complex compact conformal flat manifold $\mathbb{C}C^N, N = n - 2$; it is the compactification of \mathbb{C}^N endowed with the flat conformal structure corresponding to the quadratic isotropic cone. The parabolic group is generated by linear conformal transformations and translations. On the quadric \mathcal{Q} the conformal structure is defined by intersections of tangent spaces with \mathcal{Q} . Apparently, this structure is invariant relative to the natural action of $SO(n; \mathbb{C})$.

Classical Stein Manifolds

Such homogeneous complex manifolds $X = G/H$ have complex reductive isotropy subgroups H . Contrary to the flag manifolds which are compact, these manifolds are Stein ones and there are many holomorphic functions on them. The typical examples for $G = GL(n; \mathbb{C})$ are homogeneous spaces $S(k_1, \dots, k_{r+1}), n = k_1 + \dots + k_{r+1}$, for which the isotropy subgroups are blockdiagonal matrices with the blocks of sizes k_1, \dots, k_{r+1} . Then points of the manifold can be realized as generic sets of subspaces $L_j \subset \mathbb{C}^n, \dim L_j = k_j, 1 \leq j \leq r + 1$ or, what is equivalent, generic sets of $(k_j - 1)$ -dimensional planes in $\mathbb{C}P^{n-1}$. Since the isotropy subgroup of such a homogeneous space is a subgroup of the parabolic subgroup $P(n_1, \dots, n_r), k_j = n_j - n_{j-1}$, we have the natural fibering $S(k_1, \dots, k_{r+1}) \rightarrow F(n_1, \dots, n_r)$ (it is simple to see this geometrically: the i th subspace of a flag in the base is the direct sum of first i subspaces representing a point in the fiber). This is a convenient tool to apply complex analysis on S to the compact manifold F where there are no nontrivial holomorphic functions. Let us emphasize that such a connection exists only for special classes of classical Stein manifolds.

Let us pay special attention to the subclass of symmetric Stein manifolds. For such manifolds X , the isotropy subgroup H is fixed relative to a holomorphic involutive automorphism of G . Complex semisimple Lie groups G (including classical ones) are symmetric Stein manifolds relative to the action of their square $G \times G$ by left and right multiplications.

Classical Stein manifolds for $SL(n; \mathbb{C})$ considered above are symmetric if $r=1$ and we have the manifold of pairs of subspaces of complimentary dimensions intersecting only on $\{0\}$. The simplest example is the manifold of pairs of different points of the projective line CP^1 . Let us point out again that the transition to the generic pairs of points transforms the compact complex manifold without nonconstant holomorphic functions into a Stein manifold with a large collection of holomorphic functions.

Some other examples of symmetric Stein manifolds are connected with classical geometry and linear algebra. The affine hyperboloid in C^n ,

$$Q(z) = 1$$

is a symmetric space for $G = O(n; \mathbb{C}), H = O(n - 1; \mathbb{C})$. We can compare it with the projective quadric $Q(z) = 0$ which is a minimal flag manifold. Let us remark that there is a duality here: it is possible to interpret points of the hyperboloid of dimension n as generic hyperplane sections of the projective quadric of dimension $n - 1$.

The space X of complex symmetric matrices of order n with determinant 1 is symmetric for the group $SL(n; \mathbb{C})$ which acts by the changes of variables in the corresponding quadratic forms:

$$z \mapsto g^T z g, g \in SL(n; \mathbb{C})$$

The transitive action reflects the possibility of transforming such a form into a sum of squares. The isotropy subgroup is $SO(n; \mathbb{C})$.

The Stein symmetric manifold $X = SO(n; \mathbb{C}) / SO(k; \mathbb{C}) \times O(n - k; \mathbb{C})$ is realized as the manifold of k -dimensional subspaces in C^n on which the restriction of the principal symmetric form I is nondegenerate.

Isomorphisms in Small Dimensions

Isomorphisms of classical groups in small dimensions produce isomorphisms of some classical homogeneous manifolds. Such isomorphisms were very important in the history of geometry; below are a few examples. We will consider local isomorphisms (up to a finite center). We have $SL(2; \mathbb{C}) \cong SO(3; \mathbb{C})$. Let us realize C^3 as the space of symmetric matrices z of order 2. Then, as we remarked above, the two-dimensional submanifold X of matrices with determinant 1 is the symmetric Stein manifold for the group $SL(2; \mathbb{C})$. On the other hand, we can take $\det z$ as the quadratic symmetric form I in C^3 ; then X is the hyperboloid for this form and the action of $SL(2; \mathbb{C})$ on symmetric matrices gives the orthogonal transformations relative to this form I .

Similarly, we can interpret the local isomorphism $SO(4; \mathbb{C}) \cong SL(2; \mathbb{C}) \times SL(2; \mathbb{C})$. We realize C^4 as the space of square matrices z of order 2 with the symmetric quadratic form $I(z, z) = \det(z)$. Then left and right multiplications of z on unimodular matrices ($z \mapsto uzv, u, v \in SL(2; \mathbb{C})$) induce orthogonal transforms for the form I and any orthogonal transform can be represented in such a form (one can see it by the calculation of dimensions).

The local isomorphism $SL(4; \mathbb{C}) \cong SO(6; \mathbb{C})$ has a slightly more complicated nature. Let us consider the Grassmannian $Gr_C(2; 4)$ of lines in the projective space CP^3 with 2×4 matrices Z as matrix homogeneous coordinates. Let $p_{ij}, i < j$, be the minors of Z with i th and j th columns. They are called Plücker coordinates on $Gr_C(2; 4)$: the equivalency class of Z is defined by the sequence of six numbers $p = (p_{ij}, 1 \leq i < j \leq 4) \neq (0, \dots, 0)$ up to a constant factor. Thus, we have an imbedding of $Gr_C(2; 4)$ in the projective space CP^5 . The image will be the quadric

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0$$

Thus, we have the isomorphism of two flag manifolds and the action of $SL(4; \mathbb{C})$ on the Grassmannian transforms in orthogonal transformations of four-dimensional quadric in CP^5 . The Plücker coordinates can be defined for any Grassmannian, but they do not produce in other cases some isomorphisms with other flag manifolds; nevertheless, they realize them as intersections of quadrics in projective spaces.

Compact Classical Homogeneous Manifolds

Compact classical groups $U(n), SU(n), O(n), SO(n), Sp(l)$ are maximal compact subgroups in the corresponding classical complex groups $GL(n; \mathbb{C}), SL(n; \mathbb{C}), O(n; \mathbb{C}), SO(n; \mathbb{C}), Sp(l; \mathbb{C})$. This condition defines them up to an isomorphism. They are fixed subgroups of some antiholomorphic involutive automorphisms. The unitary groups $U(n)$ and $SU(n)$ are the groups of unitary matrices ($g^*g = E$), correspondingly, of unitary matrices with determinant 1. As the compact orthogonal group we can take the intersection $U(n) \cap O(n; \mathbb{C})$. For the standard form I , it will be the group of real orthogonal matrices: $g^T g = E$ (so the involution in $O(n; \mathbb{C})$ is the conjugation $g \mapsto \bar{g}$). Similarly, we can take $Sp(l) = SU(2l) \cap Sp(l; \mathbb{C})$ (then the involution is $g \mapsto -\bar{g}j$).

Compact classical groups act on compact homogeneous Riemann manifolds. There are two mechanisms connecting compact and complex homogeneous manifolds. We observe the first possibility in the case of flag manifolds which are

compact. We considered them so far relative to the action of complex (noncompact) groups. It turns out that on the flag manifold $F = G/P$ the maximal compact subgroup $U \subset G$ continues to be transitive: so we can consider flag manifolds also as being homogeneous with compact groups. Then $F = U/C$, where C is the centralizer of a torus in U . There is a Kähler metric on F , invariant relative to U . Thus, G is the group of all automorphisms of F as the complex manifold, but U is the group of its automorphisms as the Kähler manifold. It defines two sides of geometry of flag manifolds: complex and Kähler. Flag manifolds are the only compact homogeneous Kähler manifolds with semisimple Lie groups (the class of all compact Kähler manifolds also contains locally flat compact manifolds – toruses). In the example considered above we have $F(n_1, \dots, n_r) = \text{SU}(n)/\text{S}(\text{U}(k_0) \times \dots \times \text{U}(k_r))$. In the language of Stiefel (homogeneous) coordinates, we fix a positive Hermitian form in \mathbb{C}^n and characterize subspaces by orthonormal bases. For $r = 1$ we have Grassmannians $\text{Gr}_{\mathbb{C}}(k; n)$, in particular the projective space $\mathbb{C}\text{P}^{n-1}$ which we consider relative to the action of the unitary groups. Relative to this action they are Hermitian symmetric spaces. In the case of minimal flag manifolds for other groups the action of maximal compact subgroups also defines on them the structure of compact Hermitian symmetric spaces. Let us emphasize that relative to noncompact groups of biholomorphic automorphisms G , the minimal flag manifolds (including the Grassmannians) are not symmetric.

In the case of homogeneous Stein manifolds $X = G/H$, the picture is different: the maximal compact subgroups have no open orbits. There are totally real orbits which are the compact forms of X : $X_{\mathbb{R}} = G_{\mathbb{R}}/H_{\mathbb{R}}$, where $G_{\mathbb{R}}$ and $H_{\mathbb{R}}$ are compact forms of G and H , respectively. It is the canonical embedding of compact homogeneous manifolds in their complexifications. The important special case is the embedding of compact symmetric manifolds in the Stein symmetric manifolds – their complexifications.

For compact symmetric manifolds $X = U/K$ the groups U, K are compact Lie groups and elements of K are fixed for an involutive automorphism σ such that K contains the connected component of the subgroup of all fixed elements of σ . This possibility to connect several symmetric manifolds with one involution is illustrated by the next example. The sphere $S^{n-1} \subset \mathbb{R}^n$ is the symmetric space $\text{SO}(n)/\text{SO}(n-1)$; the real projective space $\mathbb{R}\text{P}^{n-1}$ is $\text{SO}(n)/\text{O}(n-1)$. Here $\text{SO}(n-1)$ is the connected component of $\text{O}(n-1)$ and S^{n-1} is a double covering of $\mathbb{R}\text{P}^{n-1}$. A few more examples, the

real Grassmannian $\text{Gr}_{\mathbb{R}}(k; n)$ of k -subspaces in \mathbb{R}^n can be defined as $\text{SO}(n)/\text{S}(\text{O}(k) \times \text{O}(n-k))$. This representation corresponds to the characterization of subspaces by orthonormal bases. The consideration of arbitrary bases defines the action of the larger group $\text{GL}(n; \mathbb{R})$ on $\text{Gr}_{\mathbb{R}}(k; n)$. Relative to this action, the real Grassmannian is not symmetric since the isotropy subgroup is parabolic and is not involutive. Such a possibility to extend the group is typical for a class of compact symmetric manifolds called symmetric R -spaces. They are real forms of Hermitian compact symmetric manifolds (minimal flag manifolds). Let us also mention compact symmetric spaces $\text{SU}(n)/\text{SO}(n)$, which is the compact form of the space of unimodular symmetric matrices and can be presented by the submanifold of unitary matrices in it. Also, all compact Lie groups G are symmetric spaces relative to the action of $G \times G$.

Noncompact Riemannian Symmetric Manifolds

This class of symmetric manifolds has the strongest connections with classical mathematics. Let us consider noncompact real semisimple Lie groups – real forms of complex semisimple Lie groups. They correspond to antiholomorphic involutions in complex groups.

Between real forms of $\text{SL}(\mathbb{C}, n)$ there are real and quaternionic unimodular groups $\text{SL}(\mathbb{R}, n)$, $\text{SL}(\mathbb{H}, n)$ and pseudounitary groups $\text{SU}(p, q)$ of complex matrices preserving a Hermitian form H of the signature (p, q) . The complex orthogonal group has as real forms, in particular, pseudoorthogonal groups $\text{SO}(p, q)$ of real matrices preserving a quadratic form of the signature (p, q) .

Let G be a real simple Lie group and K be its maximal compact subgroup. Then $X = G/K$ is a Riemann symmetric manifold of noncompact type; K is defined by an involutive automorphism of G . Therefore, in irreducible situation there is a correspondence between noncompact Riemann symmetric manifolds and real simple noncompact Lie groups. K -orbits on X are parametrized by points of the orbit on X of a maximal abelian subgroup A – the Cartan subgroup of the symmetric space X . Its dimension l is the important invariant of X – its rank. The algebraic base for geometry of X is the Iwasawa decomposition

$$G = KAN$$

where N is a maximal unipotent subgroup (in a natural sense compatible with A). Then the parabolic subgroup $P = AN$ is transitive on X .

Symmetric Cones

Let us start with $X = GL(n, \mathbb{R})/O(n)$. This manifold corresponds to the classical theory of quadratic forms: X can be realized as the manifold $Sym_+(n)$ of symmetric positive matrices $x \gg 0$ of order n (corresponding to positive quadratic forms). Then the transitivity of $GL(n; \mathbb{R})$ on X corresponds to the possibility to transform positive forms to a sum of squares. The sufficiency of triangle matrices for such transformations corresponds to the transitivity on $X = Sym_+(n)$ of the parabolic subgroup P of (upper) triangle matrices with positive diagonal elements. So A is the group of diagonal matrices with positive elements and the submanifold of diagonal matrices in X parametrizes K -orbits. The general fact about A -parametrization in this example is the classical fact about the reduction of quadratic forms to diagonal form by orthogonal transformations.

There are complex and quaternionic versions of this picture. The symmetric manifold $X = GL(n; \mathbb{C})/U(n)$ is realized as the manifold $Herm_+(n)$ of positive complex Hermitian matrices (forms) and similarly classical facts of linear algebra on Hermitian quadratic forms are transformed into geometrical statements on symmetric spaces. Let us emphasize that we consider here the group $GL(n; \mathbb{C})$ as the real group. The same situation exists with the manifold $Herm_+(\mathbb{H}, n)$ of positive quaternionic Hermitian matrices, which is the symmetric manifold for the real group $GL(n; \mathbb{H})$.

These three manifolds can be included in an impressive geometrical structure. They all are convex homogeneous cones V in linear spaces \mathbb{R}^N which are self-dual ($V = V^*$) relative to a bilinear form $\langle \cdot, \cdot \rangle$. Let us recall that

$$V^* = \{x; \langle x, y \rangle > 0, y \in \bar{V} \setminus 0\}$$

Here \bar{V} is the closure of V . So these three symmetric manifolds are linear homogeneous self-dual cones.

There is only one more type of classical homogeneous self-dual cones – quadratic (Lorentzian) cones

$$L_n = \{x \in \mathbb{R}^{n+1}; x_1^2 - x_2^2 - \dots - x_{n+1}^2 > 0, x_1 > 0\}$$

which is also called the future light cone (the condition $x_1 < 0$ defines the past light cone). The group of linear automorphisms of this cone is $SO(1, n) \times \mathbb{R}^+$; the first factor is the Lorentz group.

There is also one exceptional 27-dimensional cone; it is possible to interpret this cone as the cone of positive Hermitian matrices of third order over Cayley numbers. There is a very nice structural theory of homogeneous self-dual cones; it is convenient to develop this theory in the language of

Jordan algebras (Faraut and Koranyi 1994). Such cones participate as elements of explicit constructions of other classes of symmetric spaces (see below).

Following Siegel, it is possible to connect with homogeneous self-dual cones multidimensional versions of Euler integrals (Γ - and B -functions) (Faraut and Koranyi 1994). They have many applications, including those to integral formulas for complex symmetric domains.

Riemann Symmetric Manifolds of Rank 1

The first example of non-Euclidean geometry is connected with the Riemann symmetric manifolds of rank 1 – hyperbolic spaces; $X = SO(1, n)/O(n)$ is the hyperbolic space of dimension n . It can be realized as the upper sheet of the two-sheeted hyperboloid:

$$x_0^2 - x_1^2 - \dots - x_n^2 = 1, x_0 > 0$$

Pseudoorthogonal linear transformations from $SO(1, n)$ preserve this surface; they play the role of hyperbolic motions. The equivalent realization is in the real ball $x_1^2 - \dots - x_n^2 < 1$ relative to the projective transformations preserving this ball.

Another example of a Riemann symmetric manifold of rank 1 is the complex hyperbolic symmetric space $X = SU(1; n)/U(n)$. It can similarly be realized either as the hyperboloid

$$|z_0|^2 - |z_1|^2 - \dots - |z_n|^2 = 1$$

in \mathbb{C}^{n+1} relative to pseudounitary linear transformations or as the complex ball $|z_1|^2 + \dots + |z_n|^2 < 1$ relative to complex projective transformations preserving it. There are also quaternionic hyperbolic spaces which are realized as the quaternionic balls in the quaternionic projective spaces. These three series exhaust all classical Riemann symmetric manifolds of rank 1 of noncompact type. There is only one exceptional symmetric manifold of rank 1: it has the dimension 16 and can be interpreted as a two-dimensional ball for Cayley numbers.

Classical Symmetric Domains in \mathbb{C}^n (Cartan Domains)

Riemann symmetric manifolds of noncompact type which admit an invariant complex structure also have an invariant Hermitian form corresponding to the Riemann metrics. For this reason, we will call them noncompact Hermitian symmetric manifolds (we considered above the compact Hermitian symmetric manifolds). They are Stein manifolds, but as opposed to symmetric Stein manifolds, which we considered above, they are homogeneous relative to real groups. The condition for a Riemann symmetric

manifold $X = G/K$ to be Hermitian is that K has an one-dimensional center. All Hermitian symmetric manifolds of noncompact type can be realized as bounded domains in \mathbb{C}^n (but, of course, not all their holomorphic automorphisms extend in \mathbb{C}^n). In the case of classical manifolds, these domains are called Cartan's domains: Cartan gave their explicit matrix realizations.

The nature of groups of holomorphic automorphisms of symmetric domains $X = G/K \subset \mathbb{C}^N$ is explained by Cartan's duality. Each such domain (Hermitian symmetric manifold of noncompact type) admits an embedding in a Hermitian symmetric manifold of compact type $X_{\mathbb{C}}$ such that the complexification $G_{\mathbb{C}}$ of G is the group of holomorphic automorphisms of $X_{\mathbb{C}}$ (correspondingly, D is an open G -orbit on $X_{\mathbb{C}}$). Moreover, X lies inside a (Zariski open) coordinate chart \mathbb{C}^N , which is an orbit of a parabolic subgroup.

The simplest example is the complex ball CB^n (complex hyperbolic space) imbedded in the complex projective space CP^n . The affine chart \mathbb{C}^n is the orbit of the parabolic subgroup of affine transformations. Let us consider more complicated examples.

Let $X_{\mathbb{C}}$ be the Grassmannian $Gr_{\mathbb{C}}(k; n)$, $q = n - k \geq p$; we will use matrix homogeneous coordinates $Z - k \times n$ matrices - for the description of the symmetric domain. Then $G_{\mathbb{C}} = SL(n; \mathbb{C})$. Let us take its real form $G = SU(k; q)$, $k + q = n$. We fix a Hermitian form H of the signature (k, q) and realize G as the group of matrices preserving H :

$$gHg^* = H$$

Then $X = X_{k,q} = SU(k, q)/S(U(k) \times U(q))$ can be realized as the domain in the Grassmannian

$$ZHZ^* \gg 0$$

so that this Hermitian matrix of order k must be positive. It is essential that this condition is invariant relative to multiplications of Z on nondegenerate matrices u on the left and, therefore, it is a well-defined condition in homogeneous coordinates.

Let us specify the choice of H :

$$H_1 = \begin{pmatrix} E_k & 0 \\ 0 & -E_q \end{pmatrix}$$

Then the corresponding domain X_1 is defined in inhomogeneous coordinates $Z = (E_k, z)$, $z \in Mat(k, q)$, by the condition

$$E_k - zz^* \gg 0$$

This matrix ball lies completely in the coordinate chart \mathbb{C}^{kq} . Its rank is equal to $\min(k, q)$. Thus, we

have the realization of this Hermitian symmetric space as a bounded domain in \mathbb{C}^N , $N = kq$. In the case $k = 1$, we have the usual (scalar) complex ball. Let us remark that the edge of the boundary (Shilov's boundary) is the compact symmetric space

$$zz^* = E_k$$

with the group of automorphisms $S(U(k) \times U(q))$ (the isotropy subgroup of X). For $k = q$ the edge coincides with the set of unitary matrix $U(k)$.

Different forms H of the signature (k, q) are linearly equivalent and they correspond to different (biholomorphically equivalent) realizations of this Hermitian symmetric spaces. Let us, in the beginning, set $k = q$; the inhomogeneous matrix coordinates are square matrices of order k . Let us take the form

$$H_2 = \begin{pmatrix} 0 & iE_k \\ -iE_k & 0 \end{pmatrix}$$

Then, in inhomogeneous matrix coordinates, we have the domain X_2 :

$$\frac{1}{i}(z - z^*) \gg 0$$

(complex matrices with positive skew-Hermitian parts). This domain (but not its boundary) lies in the chart. It has the structure of the tube domain $T = \mathbb{R}^n + iV$, $n = k^2$, corresponding to the symmetric cone of positive Hermitian matrices (we take the space of such matrices as a real form of \mathbb{C}^n). The group of affine transformations of the tube domain:

$$z \mapsto uzu^* + a, \quad u \in GL(k; \mathbb{C}), a \in Herm(k)$$

is transitive on X_2 ; it is the parabolic subgroup in $SU(k, q)$.

The biholomorphic equivalency of the realizations of X corresponding to different H is induced by the equivalency of these forms. We have

$$H_2 = \lambda H_1 \lambda^*, \quad \lambda = \frac{\sqrt{2}}{2} \begin{pmatrix} E_k & -iE_k \\ -iE_k & E_k \end{pmatrix}$$

Then the transform $Z \mapsto Z\lambda$ transforms X_2 in X_1 . In inhomogeneous coordinates it is the fractional linear matrix transform

$$z \mapsto i(z + iE_k)^{-1}(z - iE_k)$$

It is the matrix version of the classical Cayley transform. Similarly, we can write down the inverse transform.

If $q \neq k$, then there is also an analog of the tube realization. Let $r = q - k > 0$ and

$$H_2 = \begin{pmatrix} 0 & iE_k & 0 \\ -iE_k & 0 & 0 \\ 0 & 0 & -E_r \end{pmatrix}$$

Let us represent the inhomogeneous coordinates as $z = (E_k, w, u), w \in \text{Mat}(k), u \in \text{Mat}(k, r)$. Then the domain X_2 is defined by the condition

$$\frac{1}{i}(w - w^*) - uu^* \gg 0$$

This is an example of Siegel domains of the second kind (Pyatetskii-Shapiro 1969). This domain has a transitive group of affine transformations:

$$\begin{aligned} (w, u) &\mapsto (w + a + 2ub^* + bb^*, u + b) \\ a &\in \text{Herm}(k), \quad b \in \text{Mat}(k, r) \\ (w, u) &\mapsto (cwc^*, cu) \quad c \in \text{GL}(k; \mathbb{C}) \end{aligned}$$

This class of symmetric domains in Grassmannians is called Cartan's domains of the first class. There are similar constructions for minimal flag domains (compact Hermitian symmetric spaces) with other groups. Let us consider the Lagrangian Grassmannian $\text{Gr}_{\mathbb{C}}^L(k; 2k)$ corresponding to the form J above. Here $G_{\mathbb{C}} = \text{Sp}(k, \mathbb{C})$. Its real form $G = \text{Sp}(k; \mathbb{R})$ can be realized as the subgroup of complex symplectic matrices preserving a Hermitian form H of the signature (k, k) . In other words, we intersect the domains from the last example with the Lagrangian Grassmannians. We consider the coordinate chart with inhomogeneous coordinates – symmetric matrices $z \in \text{Sym}(k)$. For H_1 we have the domain of symmetric matrices z with the condition

$$E_k - z\bar{z} \gg 0, z = z^{\top}$$

This bounded realization is called Siegel's disk. For H_2 the real form is the group of real symplectic matrices and X_2 is the domain

$$\Im z = \frac{1}{2i}(z - \bar{z}) \gg 0, \quad z = z^{\top}$$

of complex symmetric matrices with positive imaginary parts; it is called Siegel's half-plane. This is the third class of Cartan's domains. There are Siegel domains of second kind connecting with the cones of positive symmetric matrices; some of them are homogeneous, but they are never symmetric.

There are two more series of classical minimal flag manifolds: the isotropic Grassmannians and quadrics. They both contain the dual bounded symmetric domains (Cartan's domains of second and fourth classes correspondingly). Some of these domains in the isotropic Grassmannians admit the realizations as tubes with the cone of positive Hermitian quaternionic matrices and others as Siegel domains of the second kind corresponding to the same cones.

Symmetric domains in quadrics can be realized as tube domains with the Lorentzian (light) cones.

The corresponding tubes are called the future (past) tube, depending on which light cone was taken. Let us consider this construction. The group of holomorphic automorphisms of these domains is $G = \text{SO}(2; n)$ – the conformal extension of the Lorentz group. To realize this group, let us fix a real symmetric matrix Q of signature $(2, n)$ and the group is the group of linear transformations preserving simultaneously the quadratic symmetric and Hermitian forms with this matrix Q :

$$g^{\top} Q g = Q, \quad g^* Q g = Q$$

The standard realization corresponds to the diagonal matrix Q with the diagonal $(1, 1, -1, \dots, -1)$. Cartan's domains of the fourth class are connected components of the manifold

$$ZQZ^{\top} = 0, \quad ZQZ^* > 0$$

where rows Z are homogeneous coordinates in the projective space $\mathbb{C}P^{n+1}$. In other words, we consider a domain on the quadric in the projective space (which is the complex flat conformal space $\mathbb{C}C^n$). For the standard Q the domain will lie in the coordinate chart; thus it is the bounded realization. For the tube realization, we take

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & E_n \end{pmatrix}$$

Let $Z = (z_0, z_1, w_1, \dots, w_n), w = u + iv, q(s, t) = s_1 t_1 - s_2 t_2 - \dots - s_n t_n$ and we consider the affine chart $\mathbb{C}^{n+1} = \{z_0 = 1\}$. We have

$$\begin{aligned} ZQZ^{\top} &= 2z_1 + q(w, w) = 0 \\ ZQZ^* &= 2\Re z_1 + q(w, \bar{w}) > 0 \end{aligned}$$

The first condition gives $2\Re z_1 = q(v, v) - q(u, u)$ and then the second condition gives the final description of the considered set in \mathbb{C}_w^n :

$$q(v, v) = v_1^2 - v_2^2 - \dots - v_n^2 > 0, \quad w = u + iv$$

as the union of the future and the past tubes ($T_{\pm} = \{v_1 \gtrless 0\}$). The edge \mathbb{R}^n of these tubes ($v = 0$) has the structure of the Minkowski space corresponding to the form q . The parabolic subgroup is the affine conformal group of the Minkowski space. It includes the Poincaré group and is transitive on tubes. The complete group of holomorphic automorphisms of tubes $G = \text{SO}(2, n)$ is the group of all (not only affine) conformal transformations of the Minkowski space. The complete edge of these symmetric domains in the quadric $\mathbb{C}C^n$ is the conformal compactification of the Minkowski space (a compact symmetric R -space with the compact group $S(\text{O}(2) \times \text{O}(n))$ on which the noncompact group $\text{SO}(2, n)$ also acts).

In addition to four Cartan’s classes of classical domains there are two exceptional symmetric domains in the dimensions 27 and 16 (dual to two exceptional compact Hermitian symmetric spaces of these dimensions). The first of them can be realized as the tube domain corresponding to the exceptional cone of positive Hermitian matrices with Cayley numbers of order 3 (the dimension 27) and another can be realized as a Siegel domain of the second kind associated with the eight-dimensional future tube. It is possible, using Γ -function of self-dual homogeneous cones, to write explicit Bergman and Cauchy–Szegő integral formulas.

Noncompact Symmetric R-Spaces

There are several other interesting noncompact symmetric manifolds. Let us mention the noncompact symmetric R -spaces which are real forms of complex symmetric domains. The typical example is the domain of real square matrices $x \in \text{Mat}(k)$:

$$E_k - xx^T \gg 0$$

The condition is that this symmetric matrix is positive. It is the Riemann symmetric space with the group $G = \text{SO}(k, k)$. It can be embedded in the real Grassmannian $\text{Gr}_{\mathbb{R}}(k; 2k)$ with the matrix homogeneous coordinates $X \in \text{Mat}_{\mathbb{R}}(k, 2k)$ and the group $\text{SL}(2k; \mathbb{R})$ acting of X by right multiplications. Let

$$I_1 = \begin{pmatrix} E_k & 0 \\ 0 & -E_k \end{pmatrix}$$

and $\text{SO}(k, k)$ be the subgroup of matrices preserving the quadratic form $I_1: gI_1g^T = I_1$. This group will preserve the domain $XI_1X^T \gg 0$ and, in the inhomogeneous coordinates $X = (E_k, x), x \in \text{Mat}_{\mathbb{R}}(k)$, it will be exactly the same as the domain above. The group $\text{SO}(k, k)$ acts by matrix fractional linear transformations. This domain is the real form on Siegel’s ball. If we replace the form on

$$I_2 = \begin{pmatrix} 0 & E_k \\ E_k & 0 \end{pmatrix}$$

then we realize our symmetric manifold as the domain

$$x + x^T \gg 0$$

So, the symmetric part of the matrix x must be positive. This realization is homogeneous relative to the linear automorphisms: $x \mapsto axa^T + b, a \in \text{GL}(k; \mathbb{R}), b = -b^T$. A similar construction exists for rectangular matrices.

Geometry of Isomorphisms in Small Dimensions

We connected above several local isomorphisms of complex classical groups with some geometrical facts. Let us mention now several similar examples for real groups. We start from isomorphisms of symmetric cones. The cone $\text{Sym}_+(2)$ of symmetric positive matrices of second order is (linearly) isomorphic to the future light cone $L(2)$. The comparison of the groups of automorphisms gives the local isomorphism

$$\text{SL}(2; \mathbb{R}) \cong \text{SO}(1; 2)$$

This isomorphism corresponds also to the isomorphism of two classical realizations of hyperbolic plane – of Poincaré and Klein. Let us also mention that the isomorphism $\text{SL}(2, \mathbb{R}) \cong \text{SU}(1, 1)$ corresponds to the holomorphic equivalency of the disk and the upper half-plane. The isomorphism $\text{Herm}_+(2) = L(3)$ corresponds to the presentation of any Hermitian matrix of the order 2 in Pauli’s coordinates,

$$z = \begin{pmatrix} t - x_1 & x_2 + ix_3 \\ x_2 - ix_3 & t + x_1 \end{pmatrix}$$

Then, $\det z = t^2 - x_1^2 - x_2^2 - x_3^2$. To compare the groups of automorphisms, we receive

$$\text{SL}(2, \mathbb{C}) \cong \text{SO}(1, 3)$$

Similarly, in the quaternionic case, the isomorphism of the cones $\text{Herm}_+(2, \mathbb{H})$ gives the isomorphism

$$\text{SL}(2, \mathbb{H}) \cong \text{SO}(1, 5)$$

The linear isomorphism of cones produces the holomorphic isomorphism of corresponding tubes and their groups of holomorphic automorphisms. So each of these three isomorphisms gives automatically one more isomorphism. Let us give it for the first two cones:

$$\text{Sp}(2, \mathbb{R}) \cong \text{SO}(2, 2), \quad \text{SU}(2, 2) \cong \text{SO}(2, 3)$$

We just compared the descriptions of automorphisms of classical tubes from above.

Considering $\det(x)$ as the quadratic form of signature $(2, 2)$ on $\text{Mat}(2) \simeq \mathbb{R}^4$, we obtain

$$\text{SO}(2, 2) \cong \text{SL}(2, \mathbb{R}) \times \text{SL}(2, \mathbb{R})$$

Each of local isomorphisms in the complex case has different real forms which admit some geometrical interpretations. We mentioned above two real forms of the isomorphism $\text{SL}(4; \mathbb{C}) \cong \text{SO}(6; \mathbb{C})$. The isomorphism for $\text{SO}(2, 2)$ admits another interpretation in the language of Plücker’s coordinates: points of the quadric in $\mathbb{R}P^5$ of the signature $(2, 3)$ can be interpreted as (complex) lines in $\mathbb{C}P^3$ which lie on a Hermitian quadric of the signature $(2, 2)$ (Gindikin

1983). The isomorphism above for the group $SL(2, \mathbb{H})$ also corresponds to Hopf's fibering of CP^3 on complex lines over the sphere S^4 or the isomorphism S^4 and the quaternionic projective line HP^1 . In all these cases, isomorphisms of homogeneous manifolds intertwine the actions of locally isomorphic groups.

Pseudo-Riemann Symmetric Manifolds

We obtain the next broad class of homogeneous manifolds if we preserve conditions that the group G is a real semisimple one, the isotropy subgroup H is involutive, but we remove the restriction that H must be (maximal) compact. Such symmetric manifolds are often called semisimple pseudo-Riemann symmetric manifolds (since there are also pseudo-Riemann symmetric manifolds whose groups are not semisimple). This class of spaces contains symmetric Stein manifolds $X_{\mathbb{C}} = G_{\mathbb{C}}/H_{\mathbb{C}}$. Each semisimple symmetric manifold $X = G/H$ admits complexification as a symmetric Stein manifold. Each real semisimple Lie group G is symmetric relative to the group $G \times G$.

The simplest family of semisimple symmetric manifolds is the family of all hyperboloids of all signatures

$$H_{p,q} = \{x_1^2 + \dots + x_p^2 - x_{p+1}^2 - \dots - x_n^2 = 1\}$$

with the groups $SO(p, q)$. Their complexifications are complex hyperboloids. There are two types of Riemann manifolds in these families: compact ones – spheres and noncompact ones – two-sheeted hyperboloids; all others are pseudo-Riemann.

The Cartan duality holds for pseudo-Hermitian symmetric manifolds: they are domains in compact Hermitian symmetric manifolds (minimal flag manifolds) $Z = G_{\mathbb{C}}/P_{\mathbb{C}}$. They are open orbits of real forms G of the groups of holomorphic automorphisms $G_{\mathbb{C}}$. We construct examples of such manifolds if we consider one of the above-described realizations of noncompact Hermitian symmetric manifolds (through matrix homogeneous coordinates) and replace the condition of positivity with the condition that the symmetric (Hermitian) matrix in the definition has a fixed nondegenerate signature $(i, k - i)$. We can call such pseudo-Hermitian symmetric manifolds satellites of Hermitian ones. Correspondingly, we can consider nonconvex tubes, for example, the set T of such symmetric matrices whose imaginary parts have the signature $(i, n - i)$. This domain is linear homogeneous, but it is not symmetric; to receive the symmetric manifold we need to extend the nonconvex tube by a

manifold of smaller dimension (which plays a role of infinity).

There are pseudo-Hermitian symmetric manifolds which are not satellites of Hermitian ones. Let us give an interesting example. The group $SL(2p, \mathbb{R})$ has two open orbits on the Grassmannian $Gr_{\mathbb{C}}(p; 2p)$ which are both pseudo-Hermitian symmetric spaces. Let us consider as above the Stiefel coordinates $Z \in Mat_{\mathbb{C}}(p, 2p)$ and let $Z = X + iY$. Then the orbits are defined by the conditions

$$\det \begin{pmatrix} X \\ Y \end{pmatrix} \geq 0$$

In the intersection with the coordinate chart $Z = (E, z), z \in Mat_{\mathbb{C}}(p), z = x + iy$, we have the conditions

$$\det y \geq 0$$

Therefore, we obtain (nonconvex) tube domains in $C^N = Mat_{\mathbb{C}}(p), N = p^2$, corresponding to nonconvex homogeneous cones V_{\pm} of real matrices with positive (negative) determinants. These tubes do not coincide with the symmetric manifolds which include also some sets of small dimensions outside of the coordinate chart (on “infinity”). There are other homogeneous nonconvex cones such that corresponding tube domains are Zariski open parts of pseudo-Hermitian symmetric spaces (D'Atri and Gindikin 1993). Between these cones are cones of nondegenerate skew-symmetric matrices, of skew-Hermitian quaternionic matrices. We again observe strong connections with classical mathematics. Not all pseudo-Hermitian symmetric manifolds admit such tube realizations of dense parts. Analysis in pseudo-Hermitian symmetric manifolds is very interesting: we consider there instead of holomorphic functions $\bar{\partial}$ -cohomology of some degree.

Geometric relations between different symmetric manifolds are usually important for analytic applications since they can produce some nontrivial integral transformations. In a broad sense, such transforms are considered in integral geometry (Gelfand et al. 2003). An important example is duality between some compact Hermitian symmetric manifolds (when points in one of them are interpreted as submanifolds in another one). The simplest example is the projective duality between dual copies of projective spaces or, more generally, the realization of points of Grassmannians as projective planes. Such a duality can induce a duality between orbits of real forms of groups. In a special case, it can be a duality between Hermitian and pseudo-Hermitian symmetric manifolds.

Here is one important example. Let us consider in the projective space CP^{2k-1} the domain D which in

homogeneous coordinates – rows $z = (z_0, z_1, \dots, z_n)$ are defined by the equation $zHz^* > 0$, where H is a Hermitian form of the signature (k, k) , for example,

$$|z_0|^2 + \dots + |z_k|^2 - |z_{k+1}|^2 - \dots - |z_n|^2 > 0$$

This domain is $(k - 1)$ -pseudoconcave and it contains $(k - 1)$ -dimensional complex compact cycles, namely $(k - 1)$ -dimensional planes. The manifold of these planes is exactly the domain X in the Grassmannian $\text{Gr}_{\mathbb{C}}(k; 2k)$ (of projective $(k - 1)$ -planes) which is the noncompact Hermitian symmetric space – the orbit of the group $\text{SU}(k, k)$ (see above). This picture is the geometrical basis for a deep analytic construction. In the domain D the spaces of $(k - 1)$ -dimensional $\bar{\partial}$ -cohomology are infinite dimensional for some coefficients. Their integration on $(k - 1)$ -planes (the Penrose transform) gives sections of corresponding vector bundles on X . The images are described by differential equations – generalized massless equations. The basic twistor theory corresponds to $k = 2$ when X is isomorphic to four-dimensional future tube (see above).

Similar dual realizations of Hermitian symmetric manifolds exist only in special cases. The twistor realization of four-dimensional future tube was possible since the Grassmannian $\text{Gr}_{\mathbb{C}}(2; 4)$ is isomorphic to the quadric in $\mathbb{C}P^5$. This does not work for the future tubes of bigger dimensions but there is another possibility (Gindikin 1998). Let us have the quadric $Q_{n-1} \subset \mathbb{C}P^n$ be defined in the homogeneous coordinates by the equation

$$\square(z) = (z_0)^2 - (z_1)^2 - \dots - (z_n)^2 = 0$$

and $z \cdot \zeta$ is the bilinear form. As already mentioned, the set of (nondegenerate) hyperplane sections

$$\zeta \cdot z = 0, \quad \zeta \in \mathbb{C}^{n+1}, \quad \square(\zeta) = 1$$

of Q_{n-1} is the corresponding hyperboloid H_n . Thus, we have the duality between a flag manifold (the quadric Q_{n-1}) and a symmetric Stein manifold (the hyperboloid H_n) with the same group $\text{SO}(n + 1, \mathbb{C})$; they have different dimensions.

The group $\text{SO}(1, n)$ has two orbits on Q_{n-1} : the real quadric $Q_{\mathbb{R}} = \{z \in Q_{n-1}; \Im(z) = 0\}$ and its complement $X = Q_{n-1} \setminus Q_{\mathbb{R}}$. Hyperplane sections which do not intersect $Q_{\mathbb{R}}$ (lie at X) correspond such $\zeta \in H_n$ that

$$\square(\Re(z)) > 0$$

This set has two connected components D_{\pm} which are biholomorphically equivalent to the future and past tubes T_{\pm} of the dimension n . Let us emphasize that their group of automorphisms is $\text{SO}(2, n)$ in

spite of the fact that this group acts neither on X nor on H_n . Such an extension of the symmetry group is a very interesting phenomenon. It happens for several other symmetric manifolds, but is not a general fact. This geometrical construction gives a possibility to construct a multidimensional version of the Penrose transform from $(n - 2)$ -dimensional $\bar{\partial}$ -cohomology with different coefficients into solutions of massless equations on the future (past) tubes.

The last duality is connected with some general geometrical construction. We mentioned that each of the Riemann symmetric manifolds $X = G/K$ admits a canonical embedding in the symmetric Stein manifold $X_{\mathbb{C}} = G_{\mathbb{C}}/K_{\mathbb{C}}$. It turns out that X has in $X_{\mathbb{C}}$ a canonical Stein neighborhood – the complex crown $\Omega(X)$ such that many analytic objects on X can be holomorphically extended on the crown (Gindikin 2002). For example, all solutions of all invariant differential equations on X (which are elliptic) admit such holomorphic extension. In the last example, D_+ is the crown of the Riemann symmetric space which is defined, in H_n , by the condition $\Im(\zeta) = 0, \Re(\zeta_0) > 0$.

Symmetric manifolds are distinguished from most other homogeneous manifolds by a very rich geometry which is a background for deep analytic considerations. There are several important nonsymmetric homogeneous manifolds. We already mentioned flag manifolds and Stein homogeneous manifolds with complex semisimple Lie groups which can be nonsymmetric. Pseudo-Riemann symmetric manifolds are open orbits of real groups on compact Hermitian symmetric spaces. It turns out that open orbits on other flag manifolds also produce interesting homogeneous manifolds. Let $F = G_{\mathbb{C}}/P_{\mathbb{C}}$ be a flag manifold. Flag domains are open orbits of a real form G on F . Of course, pseudo-Hermitian symmetric manifolds are a special case of this construction. Let us consider a simple example with $G_{\mathbb{C}} = \text{SL}(3; \mathbb{C})$ and P – the triangle group. Then points of F are pairs {a point z and a line l passing through it}. Let $G = \text{SU}(2; 1)$; it has two open orbits on $\mathbb{C}P^2$: the complex ball D and its complementary $D^{\mathbb{C}}$. On F , the group G has three open orbits (flag domains): in the first $z \in D, l$ is arbitrary; in the second $l \subset D^{\mathbb{C}}$; in the third $z \in D^{\mathbb{C}}, l$ intersects D . They are all 1-pseudoconcave. In one-dimensional $\bar{\partial}$ -cohomology of these flag domains with coefficients in line bundles, are realized all three discrete series of unitary representations of $\text{SU}(2, 1)$. For arbitrary semisimple Lie groups, all discrete series of representations can also be realized in $\bar{\partial}$ -cohomology of flag domains. Crowns of Riemann symmetric spaces which we just mentioned parametrize cycles (complex compact submanifolds)

in flag domains. Some general version of the Penrose transform connects through the integration along cycles cohomology in flag domains with holomorphic solutions of some differential equations in crowns (generalized massless equations).

See also: Combinatorics: Overview; Compact Groups and their Representations; Lie Groups: General Theory; Pseudo-Riemannian Nilpotent Lie Groups; Several Complex Variables: Compact Manifolds; Stability of Minkowski Space; Symmetry Classes in Random Matrix Theory; Twistor Theory: Some Applications; Twistors.

Further Reading

Akhiezer D (1990) Homogeneous complex manifolds. In: Gindikin S and Henkin G (eds.) *Several Complex Variables IV*, vol. 10, Encyclopaedia of Mathematical Science. New York: Springer.
 D'Atri J and Gindikin S (1993) Siegel domain realization of pseudo-Hermitian symmetric manifolds. *Geometriae Dedicata* 46: 91–126.

Faraut J and Koranyi A (1994) *Analysis on Symmetric Cones*. Oxford: Oxford University Press.
 Gelfand I, Gindikin S, and Graev M (2003) *Selected Topics in Integral Geometry*. Providence, RI: American Mathematical Society.
 Gindikin S (1983) The complex universe of Roger Penrose. *Mathematical Intelligencer* 5(1): 27–35.
 Gindikin S (1998) $SO(1, n)$ -twistors. *Journal of Geometry and Physics* 26: 26–36.
 Gindikin S (2002) Some remarks on complex crowns of real symmetric spaces. *Acta Mathematica Applicata* 73(1–2): 95–101.
 Helgasson S (1978) *Differential Geometry, Lie Groups and Symmetric Spaces*. New York: Academic Press.
 Helgasson S (1994) *Geometric Analysis on Symmetric Spaces*. Providence, RI: American Mathematical Society.
 Onishchik A and Vinberg E (1993) Lie Groups and Lie Algebras I Foundations of Lie Theory. In: Onishchik A (ed.) *Lie Groups and Lie Algebras*. Encyclopaedia of Mathematical Sciences, vol. 20. New York: Springer.
 Pyatetskii-Shapiro I (1969) *Automorphic Functions and Geometry of Classical Domains*. Amsterdam: Gordon and Breach.

Classical r -Matrices, Lie Bialgebras, and Poisson Lie Groups

M A Semenov-Tian-Shansky, Steklov Institute of Mathematics, St. Petersburg, Russia, and Université de Bourgogne, Dijon, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The notion of “classical r -matrices” has emerged as a by-product of the quantum inverse scattering method (which was developed mainly by L D Faddeev and his team in their work at the Steklov Mathematical Institute in Leningrad); it has given a new insight into the study of Hamiltonian structures associated with classical integrable systems solvable by the classical inverse scattering method and its generalizations. Important classification results for classical r -matrices are due to Belavin and Drinfeld. Based on the initial results of Sklyanin, Drinfeld introduced the important concepts of “Poisson Lie groups” and “Lie bialgebras” which arise as a semiclassical approximation in the study of quantum groups.

A Poisson group is a Lie group G equipped with a Poisson bracket such that the multiplication $m: G \times G \rightarrow G$ is a Poisson mapping. A Poisson bracket on G with this property is called multiplicative. More explicitly, let λ_x, ρ_x be the left and right translation operators in $C^\infty(G)$ by an element $x \in G$, $\lambda_x \varphi(y) = \varphi(xy)$, $\rho_x \varphi(y) = \varphi(yx)$.

Multiplication in G is a Poisson mapping, if for any $\varphi, \psi \in C^\infty(G)$, we have

$$\{\varphi, \psi\}(xy) = \{\lambda_x \varphi, \lambda_x \psi\}(y) + \{\rho_y \varphi, \rho_y \psi\}(x) \quad [1]$$

Note that in general, multiplicative brackets are neither left nor right invariant; in other words, for fixed x translation operators λ_x, ρ_x do not preserve Poisson brackets.

Multiplicative Poisson brackets naturally arise in the study of integrable systems which admit the so-called “zero-curvature representation.” The study of zero-curvature equations, and in particular, of the Poisson properties of the associated monodromy map, was the main source of nontrivial examples (associated with classical r -matrices, classical Yang–Baxter equations, and factorizable Lie bialgebras). The special class of multiplicative Poisson brackets encountered in this context is closely related to factorization problems in Lie groups (in particular, the matrix Riemann problem); these problems represent the key tools in constructing solutions of zero-curvature equations.

The equivalent definition of Poisson Lie groups uses the dual language of “Hopf algebras.” Let $A = F(G)$ be the commutative algebra of (smooth) functions on a Lie group G equipped with the standard coproduct $\Delta: A \rightarrow A \otimes A$

$$\Delta \varphi(x, y) = \varphi(xy), \quad \varphi \in F(G), \quad x, y \in G$$

as usual, we identify the (topological) tensor product $F(G) \otimes F(G)$ with $F(G \times G)$. The multiplicative

Poisson bracket on G equips $F(G)$ with the structure of a Poisson–Hopf algebra, that is

$$\Delta\{\varphi, \psi\} = \{\Delta\varphi, \Delta\psi\} \tag{2}$$

Equation [2] is the starting point for the study of relations between Poisson groups and quantum groups. Following the general philosophy of deformation quantization, we can look for a deformation A_b of the commutative Hopf algebra A with the deformation germ determined by the Poisson structure on A satisfying eqn [2]. The fundamental theorem (conjectured by Drinfeld and proved by Etingof and Kazhdan) asserts that any Poisson algebra associated with a Poisson group admits a formal quantization (in the category of Hopf algebras).

Poisson Groups and Lie Bialgebras

Let G be a Lie group with Lie algebra \mathfrak{g} equipped with a multiplicative Poisson bracket. Any Poisson bracket is bilinear in differentials of functions; it is convenient to express it by means of right- or left-invariant differentials. For $\varphi \in F(G)$ set

$$\begin{aligned} \langle \nabla\varphi(x), X \rangle &= (d/dt)_{t=0}\varphi(e^{tX}x), \\ \langle \nabla'\varphi(x), X \rangle &= (d/dt)_{t=0}\varphi(xe^{tX}), \\ X \in \mathfrak{g}, \nabla\varphi(x), \nabla'\varphi(x) &\in \mathfrak{g}^* \end{aligned}$$

Let us define the Poisson operator $\eta: G \rightarrow \text{Hom}(\mathfrak{g}^*, \mathfrak{g})$ by setting

$$\{\varphi, \psi\}(x) = \langle \eta(x)\nabla\varphi(x), \nabla\psi \rangle \tag{3}$$

For a finite-dimensional Lie algebra, we can identify $\text{Hom}(\mathfrak{g}^*, \mathfrak{g})$ with $\mathfrak{g} \otimes \mathfrak{g}$; the skew symmetry of Poisson bracket implies that $\eta \in \mathfrak{g} \wedge \mathfrak{g}$. By an abuse of language, the same identification is traditionally used for infinite-dimensional algebras (e.g., for loop algebras) as well. Of course, in the latter case, the corresponding Poisson tensors are represented by singular kernels which do not lie in the algebraic tensor product and should be regarded as distributions.

Multiplicativity of Poisson bracket on G implies a functional equation for η

$$\eta(xy) = (\text{Ad } x \otimes \text{Ad } x) \cdot \eta(y) + \eta(x) \tag{4}$$

which means that η is a 1-cocycle on G (with values in $\mathfrak{g} \wedge \mathfrak{g}$). By setting

$$\delta(X) = \left(\frac{d}{dt}\right)_{t=0} \eta(e^{tX}), \quad X \in \mathfrak{g}$$

we conclude from eqn [4] that $\delta: \mathfrak{g} \rightarrow \mathfrak{g} \wedge \mathfrak{g}$ is a 1-cocycle on \mathfrak{g} , that is,

$$\begin{aligned} \delta([X, Y]) &= [X \otimes I + I \otimes X, \delta(Y)] \\ &\quad - [Y \otimes I + I \otimes Y, \delta(X)] \end{aligned}$$

Equation [4] implies that $\eta(e) = 0$, that is, a multiplicative Poisson structure is identically zero at the unit element. Its linearization at this point induces the structure of a Lie algebra on the cotangent space $T_e^*G \simeq \mathfrak{g}^*$; namely, for any $\xi, \xi' \in \mathfrak{g}^*$, choose $\varphi, \varphi' \in F(G)$ in such a way that $\nabla_e\varphi = \xi, \nabla_e\varphi' = \xi'$, and set

$$[\xi, \xi']_* = \nabla_e\{\varphi, \varphi'\} \tag{5}$$

It is easy to see that $\langle [\xi, \xi']_*, X \rangle = \langle \xi \wedge \xi', \delta(X) \rangle$, which proves that the bracket is well defined, while eqn [5] implies the Jacobi identity.

Definition 1 Let $\mathfrak{g}, \mathfrak{g}^*$ be a pair of linear spaces set in duality; $(\mathfrak{g}, \mathfrak{g}^*)$ is called a Lie bialgebra if both \mathfrak{g} and \mathfrak{g}^* are Lie algebras and the mapping $\delta: \mathfrak{g} \rightarrow \mathfrak{g} \otimes \mathfrak{g}$ which is dual to the commutator map $[\cdot, \cdot]_*: \mathfrak{g}^* \otimes \mathfrak{g}^* \rightarrow \mathfrak{g}^*$ is a 1-cocycle on \mathfrak{g} .

Thus if G is a Poisson–Lie group, the pair $(\mathfrak{g}, \mathfrak{g}^*)$ is a Lie bialgebra (called the “tangent Lie bialgebra” of G). Poisson–Lie groups form a category in which the morphisms are Lie group homomorphisms, which are also Poisson mappings. A morphism $(\mathfrak{g}, \mathfrak{g}^*) \rightsquigarrow (\mathfrak{h}, \mathfrak{h}^*)$ in the category of Lie bialgebras is a Lie algebra homomorphism $\mathfrak{g} \rightarrow \mathfrak{h}$ such that the dual map $\mathfrak{h}^* \rightarrow \mathfrak{g}^*$ is again a Lie algebra homomorphism. It is easy to see that morphisms of Poisson groups induce morphisms of their tangent bialgebras. The converse is also true.

Theorem 1

- (i) Let $(\mathfrak{g}, \mathfrak{g}^*)$ be a Lie bialgebra, G a connected, simply connected Lie group with Lie algebra \mathfrak{g} . There is a unique multiplicative Poisson bracket on G such that $(\mathfrak{g}, \mathfrak{g}^*)$ is its tangent Lie bialgebra.
- (ii) Morphisms of Lie bialgebras induce Poisson mappings of the corresponding Poisson groups.

Basically, the theorem asserts that a Poisson tensor is uniquely restored from the infinitesimal cocycle on the corresponding Lie algebra; moreover, the obstruction for the Jacobi identity vanishes globally if this is true for its infinitesimal part at the unit element of the group.

It is important to observe that Lie bialgebras possess a remarkable symmetry: if $(\mathfrak{g}, \mathfrak{g}^*)$ is a Lie bialgebra, the same is true for $(\mathfrak{g}^*, \mathfrak{g})$. Hence, the dual group G^* (which corresponds to \mathfrak{g}^*) also carries a multiplicative Poisson bracket. The duality theory for Lie bialgebras, based on the key notion of the Drinfeld double, is discussed in the next section.

Classical r -Matrices and Special Classes of Lie Bialgebras

The general classification problem for Lie bialgebras is unfeasible (e.g., classification of abelian Lie bialgebras includes classification of all Lie algebras). In applications, one mainly deals with important special classes of Lie bialgebras, of which factorizable Lie bialgebras are probably the most important. In a sense, this class may be regarded as exhaustive, since (as explained below) any Lie bialgebra is canonically embedded into a factorizable one. Various other special classes discussed in literature are “coboundary bialgebras,” “triangular bialgebras,” and “quasitriangular bialgebras.”

The Lie bialgebra $(\mathfrak{g}, \mathfrak{g}^*, \delta)$ is called a coboundary bialgebra if the cobracket δ is a trivial 1-cocycle on \mathfrak{g} , that is,

$$\delta(X) = [X \otimes I + I \otimes X, r] \quad \text{for all } X \in \mathfrak{g} \quad [6]$$

the constant element $r \in \mathfrak{g} \wedge \mathfrak{g}$ is called the “classical r -matrix.” If \mathfrak{g} is semisimple, $H^1(\mathfrak{g}, V) = 0$ for any \mathfrak{g} -module V by the classical Whitehead theorem, and hence all Lie bialgebra structures on \mathfrak{g} are of coboundary type. The associated Lie bracket on \mathfrak{g}^* is given by the formula

$$[\xi, \xi']_* = \text{ad}_{\mathfrak{g}}^* r \xi \cdot \xi' - \text{ad}_{\mathfrak{g}}^* r \xi' \cdot \xi \quad [7]$$

where we identified $r \in \mathfrak{g} \wedge \mathfrak{g}$ with a skew-symmetric linear operator $r: \mathfrak{g}^* \rightarrow \mathfrak{g}$. The restrictions imposed on r by the Jacobi identity are formulated in terms of the so-called “Yang–Baxter tensor” $[[r, r]] \in \mathfrak{g} \wedge \mathfrak{g} \wedge \mathfrak{g}$, which is a quadratic expression in r . To define it, let us mark different factors in tensor products, for example, $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$, by fixed numbers 1, 2, 3, ... which indicate their place; for simplicity, we assume that \mathfrak{g} is embedded in an associative algebra \mathcal{A} with a unit. The embeddings are defined as

$$i_{12}, i_{23}, i_{13} : \mathfrak{g} \otimes \mathfrak{g} \longrightarrow \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$$

by setting $i_{12}(X \otimes Y) = X \otimes Y \otimes I$, and similarly in other cases. For $a \in \mathfrak{g} \otimes \mathfrak{g}$, we put $i_{12}(a) = a_{12}$, etc. Set

$$[[r, r]] = [r_{12}, r_{13}] + [r_{12}, r_{23}] + [r_{13}, r_{23}] \quad [8]$$

The commutators in the RHS are computed in the associative algebra $\mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$; it is easy to check that the result does not depend on the choice of the embedding $\mathfrak{g} \hookrightarrow \mathcal{A}$.

Proposition 1 *The Jacobi identity for $[\cdot, \cdot]_*$ is valid if and only if $[[r, r]]$ is $\text{ad } \mathfrak{g}$ -invariant, that is, if*

$$[X \otimes I \otimes I + I \otimes X \otimes I + I \otimes I \otimes X, [[r]]] = 0 \quad \text{for all } X \in \mathfrak{g}$$

A coboundary Lie bialgebra with $[[r, r]] \in (\wedge^3 \mathfrak{g})^{\mathfrak{g}}$ is called “quasitriangular”; it is called “triangular” if r satisfies the classical Yang–Baxter equation $[[r, r]] = 0$. (Both terms come from another name of the classical Yang–Baxter equation, the “classical triangle equation.”)

When a Lie algebra \mathfrak{g} admits a nondegenerate invariant inner product, the class of quasitriangular Lie bialgebra structures on \mathfrak{g} admits an important specialization. Let $\mathfrak{g} \otimes \mathfrak{g}^* \simeq \mathfrak{g} \otimes \mathfrak{g}$ be the natural isomorphism induced by the inner product. Let $I \in \mathfrak{g} \otimes \mathfrak{g}^*$ be the canonical element; its image $t \in \mathfrak{g} \otimes \mathfrak{g}$ under this isomorphism is called the “tensor Casimir element.” Clearly, $t \in (S^2 \mathfrak{g})^{\mathfrak{g}}$ and, moreover, $[t_{12}, t_{23}] \in (\wedge^3 \mathfrak{g})^{\mathfrak{g}}$. When \mathfrak{g} is semisimple, the mapping $(S^2 \mathfrak{g})^{\mathfrak{g}} \rightarrow (\wedge^3 \mathfrak{g})^{\mathfrak{g}} : s \mapsto [s_{12}, s_{23}]$ is an isomorphism; in particular, if \mathfrak{g} is simple, both spaces are one dimensional and generated by a tensor Casimir (which is unique up to a scalar multiple). A Lie bialgebra (\mathfrak{g}, r) is called factorizable if $r \in \mathfrak{g} \wedge \mathfrak{g}$ satisfies the modified classical Yang–Baxter equation

$$[r, r] = c[t_{12}, t_{23}], \quad c = \text{const} \neq 0 \quad [9]$$

The convenient normalization is $c = -1/4$ (it can be achieved by an appropriate normalization of r). Instead of dealing with the modified Yang–Baxter equation, we may relax the antisymmetry condition imposed on r . Set $r_{\pm} = r \pm (1/2)t \in \mathfrak{g} \otimes \mathfrak{g}$. Since t is $\text{ad } \mathfrak{g}$ -invariant, the symmetric part of r_{\pm} drops out from the cobracket; on the other hand, one has $[[r_{\pm}, r_{\pm}]] = 0$. Regarding r_{\pm} as a linear operator, $r_{\pm} \in \text{Hom}(\mathfrak{g}^*, \mathfrak{g})$, we get the following important result:

Proposition 2 *Let $(\mathfrak{g}, \mathfrak{g}^*)$ be a factorizable Lie bialgebra.*

- (i) *The mappings $r_{\pm} \in \text{Hom}(\mathfrak{g}^*, \mathfrak{g})$ are Lie algebra homomorphisms; moreover, $r_+^* = -r_-$.*
- (ii) *The combined mapping*

$$i_r : \mathfrak{g}^* \rightarrow \mathfrak{g} \oplus \mathfrak{g} : X \mapsto (r_+ X, r_- X)$$

is a Lie algebra embedding.

- (iii) *Any $X \in \mathfrak{g}$ admits a unique decomposition $X = X_+ - X_-$ with $(X_+, X_-) \in \text{Im } i_r$.*

The additive decomposition in a factorizable Lie bialgebra gives rise to a multiplicative factorization problem in the associated Lie group. Namely, i_r may be extended to a Lie group embedding $i_r : G^* \rightarrow G \times G$ and any $x \in G$, which is sufficiently close to the unit element, admits a decomposition $x = x_+ x_-^{-1}$ with $(x_+, x_-) \in \text{Im } i_r$.

Any Lie bialgebra $(\mathfrak{g}, \mathfrak{g}^*)$ admits a canonical embedding into a larger Lie bialgebra (called its

“double”) which is already factorizable. Namely, set $\mathfrak{d} = \mathfrak{g} \oplus \mathfrak{g}^*$ as a linear space and equip it with the natural inner product,

$$\langle\langle (X, F), (X', F') \rangle\rangle = \langle F, X' \rangle + \langle F', X \rangle \quad [10]$$

Theorem 2

- (i) *There exists a unique structure of the Lie algebra on \mathfrak{d} such that: (a) $\mathfrak{g}, \mathfrak{g}^* \subset \mathfrak{d}$ are Lie subalgebras. (b) The inner product [10] is invariant.*
- (ii) *Let $P_{\mathfrak{g}}, P_{\mathfrak{g}^*}$ be the projection operators onto $\mathfrak{g}, \mathfrak{g}^* \subset \mathfrak{d}$ parallel to the complementary subalgebra. Set $r_+^{\mathfrak{d}} = P_{\mathfrak{g}}, r_-^{\mathfrak{d}} = -P_{\mathfrak{g}^*}$; then $(\mathfrak{d}, r_{\pm}^{\mathfrak{d}})$ is a factorizable Lie bialgebra.*
- (iii) *The inclusion map $(\mathfrak{g}, \mathfrak{g}^*) \rightsquigarrow (\mathfrak{d}, \mathfrak{d}^*)$ is a homomorphism of Lie bialgebras and the dual inclusion map $(\mathfrak{g}^*, \mathfrak{g}) \rightsquigarrow (\mathfrak{d}, \mathfrak{d}^*)$ is an antihomomorphism.*

Conversely, let α be a Lie algebra equipped with a nondegenerate invariant inner product, $\alpha_{\pm} \subset \alpha$ its Lie subalgebras such that (i) α_{\pm} are isotropic with respect to inner product, (ii) $\alpha = \alpha_+ + \alpha_-$ as a linear space. The triple $(\alpha, \alpha_+, \alpha_-)$ is called a “Manin triple.” Let P_{\pm} be the projection operators onto α_{\pm} in this decomposition. Set $r_{\pm} = \pm P_{\pm}$. Then (α, r_{\pm}) is a factorizable Lie bialgebra; moreover, α_+ and α_- are set into duality by the inner product in α and inherit the structure of a Lie bialgebra, and α is their double.

If $(\mathfrak{g}, \mathfrak{g}^*)$ is itself a factorizable Lie bialgebra, its double admits a simple explicit description. Set $\mathfrak{d} = \mathfrak{g} \oplus \mathfrak{g}$ (direct sum of Lie algebras); let us equip \mathfrak{d} with the inner product

$$\langle\langle (X, X'), (Y, Y') \rangle\rangle = \langle X, Y \rangle - \langle Y, X' \rangle$$

Let $\mathfrak{g}^{\delta} \subset \mathfrak{d}$ be the diagonal subalgebra; we identify \mathfrak{g}^* with the embedded subalgebra $i_r(\mathfrak{g}^*) \subset \mathfrak{d}$.

Proposition 3

- (i) $(\mathfrak{d}, \mathfrak{g}^{\delta}, i_r(\mathfrak{g}^*))$ is a Manin triple.
- (ii) As a Lie algebra, $\mathfrak{d} = \mathfrak{g} \oplus \mathfrak{g}$ is isomorphic to the double of \mathfrak{g} .

Key examples of factorizable Lie bialgebras are associated with semisimple Lie algebras and their loop algebras.

1. Let \mathfrak{k} be a compact semisimple Lie algebra: $\mathfrak{g} = \mathfrak{k}_{\mathbb{C}}$ its complexification regarded as a real Lie algebra, $\sigma \in \text{Aut } \mathfrak{g}$ the Cartan involution which fixes \mathfrak{k} , and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ the associated Cartan decomposition. Fix a real split Cartan subalgebra $\alpha \subset \mathfrak{p}$ and the associated Iwasawa decomposition $\mathfrak{g} = \mathfrak{k} + \alpha + \mathfrak{n}$; put $\mathfrak{s} = \alpha + \mathfrak{n}$. Let B be the complex Killing form on \mathfrak{g} ; let us equip \mathfrak{g} with the real inner product $\langle X, Y \rangle = \text{Im } B(X, Y)$, then $(\mathfrak{g}, \mathfrak{k}, \mathfrak{s})$ is a Manin

triple. Hence, any compact semisimple Lie group K carries a natural Poisson structure; its double $G = D(K)$ is the complex group $G = K_{\mathbb{C}}$ (regarded as a real Lie group). The associated factorization problem in G is the Iwasawa decomposition $G = KAN$, which exists globally.

2. Let \mathfrak{g} be a real split semisimple Lie algebra, \mathfrak{h} its Cartan subalgebra, and Δ_+ a system of positive roots. Fix an invariant inner product on \mathfrak{g} which is positive on \mathfrak{h} , and let $\{e_{\alpha}; \alpha \in \pm\Delta_+\}$ be the root vectors normalized in such a way that $(e_{\alpha}, e_{-\alpha}) = 1$. Let

$$\mathfrak{n}_{\pm} = \bigoplus_{\alpha \in \Delta_+} \mathbb{R} \cdot e_{\pm\alpha}$$

Fix an orthonormal basis $\{H_i\}$ in \mathfrak{h} ; let P_{\pm}, P_0 be the projection operators onto $\mathfrak{n}_{\pm}, \mathfrak{h}$ in the Bruhat decomposition $\mathfrak{g} = \mathfrak{n}_- + \mathfrak{h} + \mathfrak{n}_+$. The standard Lie bialgebra structure on \mathfrak{g} is given by the r -matrices $r_{\pm} = \pm P_{\pm} \pm \frac{1}{2}P_0$. In tensor notation,

$$r_{\pm} = \pm \sum_{\alpha \in \Delta_+} e_{\alpha} \wedge e_{-\alpha} \pm \frac{1}{2} \sum_i H_i \otimes H_i \quad [11]$$

Let $\mathfrak{b}_{\pm} = \mathfrak{h} + \mathfrak{n}_{\pm}$ be the opposite Borel subalgebras; the inner product in \mathfrak{g} sets them into duality, and $(\mathfrak{b}_+, \mathfrak{b}_-)$ is a Lie sub-bialgebra in $(\mathfrak{g}, \mathfrak{g}^*)$. Let G be the connected, simply connected Lie group associated with $\mathfrak{g}, B_{\pm} = HN_{\pm}$ its opposite Borel subgroups which correspond to \mathfrak{b}_{\pm} . Let $p: B_{\pm} \rightarrow B_{\pm}/N_{\pm} \simeq H$ be the canonical projection. The associated factorization problem in $G, \mathfrak{g} = \mathfrak{b}_+ \mathfrak{b}_-^{-1}, (b_+, b_-) \in B_+ \times B_-, p(b_+) = p(b_-)^{-1}$, is closely related to the Bruhat decomposition; it is solvable for all g in the open Bruhat cell $B_+ N_- \subset G$.

3. Let $L\mathfrak{g} = \mathfrak{g} \otimes \mathbb{C}((z))$ be the loop algebra of a finite dimensional semisimple Lie algebra \mathfrak{g} , as usual we denote the ring of formal Laurent series by $\mathbb{C}((z))$. Put $L\mathfrak{g}_+ = \mathfrak{g} \otimes \mathbb{C}[[z]], L\mathfrak{g}_- = \mathfrak{g} \otimes z^{-1}\mathbb{C}[[z^{-1}]]$. Fix an invariant inner product on \mathfrak{g} and equip $L\mathfrak{g}$ with the inner product

$$\langle\langle X, Y \rangle\rangle = \text{Res}_{z=0} \langle X(z), Y(z) \rangle dz$$

Then $(L\mathfrak{g}, L\mathfrak{g}_+, L\mathfrak{g}_-)$ is a Manin triple. The associated classical r -matrix is called “rational r -matrix”; in tensor notation, it is represented by a singular kernel

$$r(z, z') = \frac{t}{z - z'}$$

where $t \in \mathfrak{g} \otimes \mathfrak{g}$ is the tensor Casimir, which is essentially the Cauchy kernel.

4. Let us assume that $\mathfrak{g} = \mathfrak{sl}(n)$; in this case, the loop algebra $L\mathfrak{g}$ admits a nontrivial decomposition

associated with the so-called “elliptic r -matrix.” Set

$$I_1 = \text{diag}(1, \varepsilon, \dots, \varepsilon^{n-1}),$$

$$I_2 = \begin{pmatrix} 0 & 1 & \dots & 0 \\ & 0 & 1 & \\ \vdots & & \ddots & \vdots \\ & & & \ddots & 1 \\ 1 & 0 & \dots & & 0 \end{pmatrix}, \quad \varepsilon = e^{2\pi i/n} \quad [12]$$

Put $\mathbb{Z}_n^2 = \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$; for $a = (a_1, a_2) \in \mathbb{Z}_n^2$, set $I_a = I_1^{a_1} I_2^{a_2}$; matrices I_a define an irreducible projective representation of \mathbb{Z}_n^2 (they form the so-called “finite Heisenberg group”). Let us denote the elliptic curve of modulus τ by $\mathcal{E} = \mathbb{C}/\mathbb{Z} + \tau\mathbb{Z}$ and let $P \rightarrow \mathcal{E}$ be the n -dimensional holomorphic vector bundle with flat connection and with monodromies given by

$$z \mapsto z + 1 : h_1 = \text{Ad } I_1, \quad z \mapsto z + \tau : h_2 = \text{Ad } I_2$$

Let $\mathcal{G}_{\mathcal{E}} \subset L\mathfrak{g}$ be the subspace of Laurent expansions at zero of the global meromorphic sections of P with a unique pole at $0 \in \mathcal{E}$. Then $(L\mathfrak{g}, L\mathfrak{g}_+, \mathcal{G}_{\mathcal{E}})$ is again a Manin triple. The associated classical r -matrix is the kernel of a singular integral operator which associates a meromorphic section of P to its principal part at 0. Explicitly, it is given by

$$r(z - z') = \frac{1}{n} \sum_{a,b=0}^{n-1} \zeta \left(\frac{z - z'}{n} - a - b\tau \right) \times (\text{Ad } I_{a,b} \otimes I) \cdot t \quad [13]$$

where ζ is the Weierstrass zeta function.

- Let \mathfrak{g} be an arbitrary semisimple Lie algebra again. Let us equip the loop algebra $L\mathfrak{g}$ with the inner product

$$\langle \langle X, Y \rangle \rangle_0 = \text{Res}_{z=0} \langle X(z), Y(z) \rangle z^{-1} dz$$

Set $\mathcal{N}_+ = \mathfrak{n}_+ \dot{+} \mathfrak{g} \otimes z\mathbb{C}[[z]]$, $\mathcal{N}_- = \mathfrak{n}_- \dot{+} \mathfrak{g} \otimes z^{-1}\mathbb{C}[[z^{-1}]]$. We have $L\mathfrak{g} = \mathcal{N}_+ \dot{+} \mathfrak{h} \dot{+} \mathcal{N}_-$, where we identify $\mathfrak{h}, \mathfrak{n}_{\pm} \subset \mathfrak{g}$ with the corresponding subalgebras of constant loops in $L\mathfrak{g}$. Let P_{\pm}, P_0 be the projection operators onto $\mathcal{N}_{\pm}, \mathfrak{h}$ in this decomposition and $r_{\pm} = \pm P_{\pm} \pm (1/2)P_0$. The classical r -matrices r_{\pm} define on $L\mathfrak{g}$ the structure of a factorizable Lie bialgebra. The associated tensor kernels are called the trigonometric classical r -matrices.

Classical r -matrices described above are associated with factorization problems in the infinite-dimensional loop groups: matrix Riemann problems or matrix Cousin problems (in the elliptic case). Belavin and

Drinfeld have given a complete classification of factorizable Lie bialgebra structures for semisimple Lie algebras; in the loop algebra case, the problem they solved consists of classification of all meromorphic solutions of the classical Yang–Baxter equation. In other words, we assume that the distribution kernel associated with the classical r -matrix is represented by a meromorphic function (of two complex variables). Up to an equivalence, any such solution depends only on one variable and belongs to the rational, trigonometric, or elliptic type (in the latter case, the underlying Lie algebra is necessarily $\mathfrak{sl}(n)$). Classification of solutions in the elliptic case is completely rigid; in the trigonometric case, the moduli space is finite dimensional and admits an explicit description. In the rational case, the classification is somewhat less explicit (it has been completed by A Stolin under some nondegeneracy condition). Contrary to the popular belief, there are many other structures of a factorizable Lie bialgebra on loop algebras, for which the associated r -matrices are given by more singular distribution kernels.

Poisson Lie Groups

If the tangent Lie bialgebra of a Poisson Lie group is of coboundary type, the cocycle η is also trivial, $\eta(\mathfrak{g}) = r - \text{Ad } \mathfrak{g} \otimes \text{Ad } \mathfrak{g} \cdot r$. Hence, the Poisson bracket on G is given by

$$\{\varphi, \psi\} = \langle r, \nabla' \varphi \wedge \nabla' \psi \rangle - \langle r, \nabla \varphi \wedge \nabla \psi \rangle, \quad r \in \mathfrak{g} \wedge \mathfrak{g}$$

where $\nabla \varphi, \nabla' \varphi \in \mathfrak{g}^*$ are left and right differentials of $\varphi \in C^\infty(G)$. This is the so-called “Sklyanin bracket”. Let us assume that G is a matrix group; its affine ring generated by evaluation functions ϕ_{ij} which assign to $L \in G$ its matrix coefficients, $\phi_{ij}(L) = L_{ij}$. The Poisson bracket on G is completely determined by its values on ϕ_{ij} . Explicitly, we get

$$\{\phi_{ij}, \phi_{km}\}(L) = [r, L \otimes L]_{ikjm} \quad [14]$$

the commutator in the RHS is in $\text{Mat}(n^2)$. By a variation of language, evaluating functions and their values on a generic element $L \in G$ are denoted by the same letter; using tensor notation to suppress matrix indices, we get

$$\{L_1, L_2\} = [r, L_1 L_2], \quad L_1 = L \otimes I, \quad L_2 = I \otimes L \quad [15]$$

In the case of loop algebras, these Poisson bracket relations take the form

$$\{L_1(\lambda), L_2(\mu)\} = [r(\lambda, \mu), L_1(\lambda) L_2(\mu)]$$

Let us assume that G is factorizable and the associated factorization problem is globally solvable. The Poisson bracket on the dual group $G^* \simeq$

$i_r(G^*) \subset G \times G$ may be characterized in terms of the matrix coefficients of $(b_+, b_-) = i_r(b)$, or of their quotient $b = b_+ b_-^{-1}$. Explicitly, we get

$$\{b_{\pm}^1, b_{\pm}^2\}_* = [r, b_{\pm}^1 b_{\pm}^2], \quad \{b_+^1, b_-^2\}_* = [r_+, b_+^1 b_-^2] \quad [16]$$

$$\begin{aligned} \{b_1, b_2\}_* &= r b_1 b_2 + b_1 b_2 r - b_2 r_+ b_1 - b_1 r_- b_2, \\ r &= \frac{1}{2}(r_+ + r_-) \end{aligned} \quad [17]$$

The key question in the geometry of Poisson groups consists in description of symplectic leaves in G, G^* . This question is already nontrivial when G^* is abelian (and hence may be identified with the dual of the Lie algebra $\mathfrak{g} = \text{Lie}(G)$). The Poisson bracket on \mathfrak{g}^* is linear; this is the well-known Lie–Poisson (alias, Beresin–Kirillov–Kostant) bracket. Its symplectic leaves coincide with the orbits of the coadjoint representation of G in \mathfrak{g}^* . The natural way to prove this fundamental result (which goes back to Lie) is to consider first the natural action of G on the cotangent bundle $T^*G \simeq G \times \mathfrak{g}^*$; this action is Hamiltonian, and the coadjoint orbits arise as a result of Hamiltonian reduction associated with this action. The generalization of the theory of coadjoint orbits to the case of arbitrary Poisson groups starts with the notion of symplectic double, which is the nonlinear analog of the cotangent bundle.

Let D be the double of (G, G^*) ; assume for simplicity that $D = G \cdot G^*$ globally and hence the associated factorization problem is always solvable. Let $r_{\mathfrak{d}} = (1/2)(P_{\mathfrak{g}} - P_{\mathfrak{g}^*})$. Set

$$\{\varphi, \psi\}_{\pm} = \langle r_{\mathfrak{d}} \nabla \varphi, \nabla \psi \rangle \pm \langle r_{\mathfrak{d}} \nabla' \varphi, \nabla' \psi \rangle \quad [18]$$

The bracket $\{, \}_-$ is the usual Sklyanin bracket which defines the structure of a Poisson group on D , while $\{, \}_+$ is nondegenerate and defines a symplectic structure on D . Let us denote the copies of D equipped with the bracket $\{, \}_{\pm}$ by D_{\pm} . The bracket on D_+ is not multiplicative, but it is covariant with respect to the action of D_- by left and right translations; in other words, the natural mappings $D_- \times D_+ \rightarrow D_+$ and $D_+ \times D_- \rightarrow D_+$, associated with multiplication in D , preserve Poisson brackets. Since $G, G^* \subset D_-$ are Poisson subgroups, natural actions $G \times D_+ \rightarrow D_+$ and $G^* \times D_+ \rightarrow D_+$ by left and right translations are Poisson mappings. Consider the natural projections

$$\begin{array}{ccc} D_+ & & D_+ \\ \pi \swarrow & & \searrow \pi' \\ G^* \simeq D/G & & G \setminus D \simeq G^* \end{array} \quad \begin{array}{ccc} D_+ & & D_+ \\ p \swarrow & & \searrow p' \\ G \simeq D/G^* & & G^* \setminus D \simeq G \end{array}$$

onto the space of left and right coset classes. It is easy to see that functions on D_+ which are constant on each projection fiber are closed with respect to the Poisson bracket. This means that the quotient spaces inherit

the Poisson structure. Moreover, the maps π, π' and p, p' form the so-called “dual pairs”, that is, the algebras of functions which are constant on the fibers of π and π' (or of p and p') are mutual centralizers of one another in the big Poisson algebra $F(D_+)$. Since $D = G \cdot G^* = G^* \cdot G$, we have $G^*/D \simeq G, G/D \simeq G^*$; it is easy to check that the quotient Poisson structure induced on G, G^* coincides with the original one. Applying the fundamental theorem on dual pairs of Poisson mappings (going back to S. Lie), we conclude that symplectic leaves in G and G^* , respectively, coincide with the orbits of G^* (respectively, G) in these quotient spaces. The actions $G \times G^* \rightarrow G^*, G^* \times G \rightarrow G$ are called “dressing transformations”. Unit elements in G and G^* are fixed points of dressing transformations; their linearizations at the tangent spaces $T_e G^* \simeq \mathfrak{g}^*, T_e G \simeq \mathfrak{g}$ coincide with the coadjoint actions of G and G^* , respectively.

When $D \neq G \cdot G^*$ (i.e., the factorization problem in D is not always solvable), dressing actions are still well defined as global transformations of the quotient spaces; in this case G, G^* may be identified with open cells in $D/G^*, D/G$, respectively, which means that dressing action on G, G^* is, in general, incomplete.

If the group G is factorizable, symplectic leaves in the dual group G^* admit a nice uniform description: since in this case $D = G \times G$ and $G \subset D$ is the diagonal subgroup, the quotient D/G may be modeled on G itself. The quotient Poisson bracket in this realization coincides with [17], while the dressing action coincides with conjugation in G (and is independent of r). Hence, symplectic leaves in D/G coincide with conjugacy classes in G ; the equivalence of this model with G^* (equipped with the bracket [16]) is provided by the factorization map. The description of symplectic leaves in G is more subtle (and already crucially depends on the choice of r !); for semisimple Lie groups with the standard Poisson structure, it is related to the geometry of double Bruhat cells.

For loop groups with rational, trigonometric, or elliptic r -matrices, dressing action is associated with auxiliary factorization problems in the loop group. Roughly speaking, symplectic leaves correspond to rational loops with prescribed singularities. Many important examples have been described in connection with integrable lattice systems, although a complete classification theorem is still not available. For $\mathfrak{g} = \mathfrak{sl}(2)$, the elliptic Manin triple described earlier leads to the Poisson structure on the group of “elliptic loops” with values in $\text{SL}(2)$; its simplest symplectic leaves (corresponding to loops with simple poles) are associated with a remarkable Poisson algebra, the Sklyanin algebra (with four generators and two Casimir functions), which admits an interesting explicit quantization.

Dressing action is a nontrivial example of a Poisson group action. In general, such actions are not Hamiltonian in the usual sense; the appropriate generalization is provided by the notion of the nonabelian moment map. Let $G \times \mathcal{M} \rightarrow \mathcal{M}$ be an action of a Poisson group G on a Poisson manifold $\mathcal{M}, \mathfrak{g} \rightarrow \text{Vect } \mathcal{M}$, the associated homomorphism of Lie algebras. A mapping $\mu: \mathcal{M} \rightarrow G^*$ is called the nonabelian moment map associated with this action, if for any $X \in \mathfrak{g}$ and $\varphi \in F(\mathcal{M})$, we have

$$X \cdot \varphi = \langle \mu^{-1}\{\mu, \varphi\}_{\mathcal{M}}, X \rangle$$

In this case, $G \times \mathcal{M} \rightarrow \mathcal{M}$ is *a fortiori* a Poisson map. Both dressing actions $G^* \times G \rightarrow G$ and $G \times G^* \rightarrow G^*$ admit nonabelian moment maps, which are just the identity maps $\mu = \text{id}_G$ and $\mu^* = \text{id}_{G^*}$. For compact Poisson groups, the nonabelian moment map has good convexity properties, which generalize the convexity properties of the ordinary moment map for Hamiltonian group actions.

The general theory of homogeneous Poisson spaces has some peculiarities. Typically, the G -covariant Poisson structure on a given homogeneous space is not unique (when it exists); this is true already for principal homogeneous spaces (a simple example is provided by the symplectic double D_+). Let G be a Poisson Lie group, $(\mathfrak{g}, \mathfrak{g}^*)$ its tangent Lie bialgebra, \mathfrak{d} its double, U its Lie subgroup, $\mathfrak{u} = \text{Lie } U$. A subalgebra $\mathfrak{l} \subset \mathfrak{d}$ is called Lagrangian if it is isotropic with respect to the canonical inner product in \mathfrak{d} . The general classification result, according to Drinfeld, asserts that there is a bijection between G -covariant Poisson structures on G/U and the set of all Lagrangian subalgebras $\mathfrak{l} \subset \mathfrak{d}$ such that $\mathfrak{l} \cap \mathfrak{g} = \mathfrak{u}$. Various nontrivial examples arise, notably in the study of integrable systems. For instance, the geometric proof of the factorization theorem for lattice zero-curvature equation, which is stated in the following section, uses a different Poisson structure on the double (the so-called “twisted symplectic double”).

Applications to Integrable Systems

The definition of Poisson–Lie groups was motivated by key examples which arise in the theory of integrable systems. In applications, one often deals with nonlinear differential equations which may be written in the form of the so-called “lattice zero curvature equations”

$$\frac{dL_m}{dt} = L_m M_m - M_{m+1} L_m, \quad m \in \mathbb{Z} \quad [19]$$

where L_m, M_m are matrices, possibly depending on an additional parameter (or, more generally, abstract

linear operators). Equations [19] give the compatibility conditions for the auxiliary linear system

$$\psi_{m+1} = L_m \psi_m, \quad \frac{d\psi_m}{dt} = -M_m \psi_m, \quad m \in \mathbb{Z} \quad [20]$$

The use of finite-difference operators associated with a one-dimensional lattice, as in [20], is particularly well suited for the study of “multiparticle” lattice models. Let us assume that the “potential” L_m in [20] is periodic, $L_{m+N} = L_m$; the period N may be interpreted as the number of copies of an “elementary” system. It is natural to presume that “Lax matrices” L_m in [19] are elements of a matrix Lie group G (or of a loop group, if they depend on an extra parameter). The auxiliary linear problem [20] leads to a family of dynamical systems on G^N which remain integrable for any N . Let $T: G^N \rightarrow G$ be the “monodromy map” which assigns to the set L_1, \dots, L_N of local Lax matrices their ordered product $T_L = L_N L_{N-1} \cdots L_1$. Let us assume that G is equipped with the Sklyanin bracket associated with a factorizable r -matrix r . Then T is a Poisson map. Let $I(G)$ be the algebra of central functions on G ; for $\varphi \in I(G)$, set $H_\varphi = \varphi \circ T$. All functions $H_\varphi, \varphi \in I(G)$ are in involution with respect to the product Poisson bracket on G^N and give rise to lattice zero-curvature equations of the same form as [19]; for a given φ , we may choose the M -matrix in either of the two forms:

$$M_m^\pm = r_\pm(\psi_m \nabla \varphi(T_L) \psi_m^{-1}), \quad \psi_m = \prod_{1 \leq k \leq m} L_k$$

Let $L_m(t), m = 1, \dots, N$, be the integral curve of this equation which starts at L_m^0 . The construction of this curve reduces to the factorization problem associated with the chosen r -matrix. Explicitly, we get

$$L_m(t) = g_{m+1}(t)_+^{-1} L_m^0 g_m(t)_+ = g_{m+1}(t)_-^{-1} L_m^0 g_m(t)_-$$

where $(g_m(t)_+, g_m(t)_-)$ is the curve in G^* which solves the factorization problem

$$g_m(t)_+ g_m(t)_-^{-1} = {}^0\psi_m \exp(t \nabla \varphi(T(L^0))) {}^0\psi_m^{-1}, \\ {}^0\psi_m = \psi_m(L^0)$$

This result exhibits the double role of the r -matrix. On the one hand, it serves to define the Poisson structure on G^N which is adapted to the study of lattice zero-curvature equations; in particular, the dynamical flow associated with these equations is automatically confined to symplectic leaves in G^N . (In applications, G is usually a loop group equipped with a factorizable r -matrix; despite the fact that $\dim G = \infty$, it admits plenty finite-dimensional symplectic leaves.) In its second incarnation, the r -matrix serves to define the factorization problem which solves these zero-curvature equations. In the loop

group case, this is a matrix Riemann problem; its explicit solution is based on the study of the spectral curve associated with the “monodromy matrix” T_L and uses the technique of algebraic geometry.

The monodromy map $T : G^N \rightarrow G$ may be regarded as a nonabelian moment map associated with an action of the dual Lie algebra \mathfrak{g}^* on the phase space. This action actually extends to an action of the (local) Lie group G^* which transforms solutions into solutions again. This is the prototype “dressing” action (originally defined by Zakharov and Shabat in their study of zero-curvature equations related to Riemann–Hilbert problems). Dressing provides an effective tool to produce new solutions of zero-curvature equations from the “trivial” ones; it was also the first nontrivial example of a Poisson group action.

See also: Affine Quantum Groups; Bicrossproduct Hopf Algebras and Noncommutative Spacetime; Bi-Hamiltonian Methods in Soliton Theory; Deformations of the Poisson Bracket on a Symplectic Manifold; Functional Equations and Integrable Systems; Hamiltonian Fluid Dynamics; Hopf Algebras and q -Deformation Quantum Groups; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems: Overview; Lie, Symplectic and Poisson Groupoids, and their Lie Algebroids; Multi-Hamiltonian Systems; Poisson Reduction; Recursion Operators in Classical Mechanics; Toda Lattices; Yang–Baxter Equations.

Further Reading

- Babelon O, Bernard D, and Talon M (2003) *Introduction to Classical Integrable Systems*. Cambridge: Cambridge University Press.
- Belavin AA and Drinfel'd VG (1984) Triangle equations and simple Lie algebras. In: *Mathematical physics reviews*, vol. 4, *Soviet Scientific Reviews Section C Mathematical Physics Reviews*, pp. 93–165. Chur: Harwood Academic Publishers, Reprinted in

- 1998, *Classic Reviews in Mathematics and Mathematical Physics*, vol. 1. Amsterdam: Harwood Academic Publishers.
- Chari V and Pressley A (1995) *A Guide to Quantum Groups*. Cambridge: Cambridge University Press.
- Drinfeld VG (1987) Quantum groups. In: *Proceedings of the International Congress of Mathematicians, (Berkeley, Calif., 1986)* vol. 1, pp. 798–820. Providence, RI: American Mathematical Society.
- Etingof P and Schiffman O (1998) *Lectures on Quantum Groups*. Boston: International Press.
- Frenkel E, Reshetikhin N, and Semenov-Tian-Shansky MA (1998) Drinfeld–Sokolov reduction for difference operators and deformations of W-algebras. I. The case of Virasoro algebra. *Communications in Mathematical Physics* 192(3): 605–629.
- Lu J-H (1991) Momentum mappings and reduction of Poisson actions. *Symplectic Geometry, Groupoids, and Integrable Systems (Berkeley, CA, 1989)*, *Mathematical de Sciences Research Institute Publications* vol. 20: 209–226. New York: Springer.
- Lu J-H and Weinstein A (1990) Poisson–Lie groups, dressing transformations, and Bruhat decompositions. *Journal of Differential Geometry* 31(2): 501–526.
- Reshetikhin N (2000) Characteristic systems on Poisson–Lie groups and their quantization. In: *Integrable Systems: From Classical to Quantum (Montréal, QC, 1999)*, CRM Proceedings Lecture Notes, vol. 26, pp. 165–188. Providence, RI: American Mathematical Society.
- Reshetikhin NY and Semenov-Tian-Shansky MA (1990) Central extensions of quantum current groups. *Letters in Mathematical Physics* 19(2): 133–142.
- Reyman AG and Semenov-Tian-Shansky MA (1994) Group-theoretical methods in the theory of finite-dimensional integrable systems. In: *Encyclopaedia of Mathematical Sciences, Dynamical Systems VII*, ch. 2, vol. 16, pp. 116–225. Berlin: Springer.
- Semenov-Tian-Shansky MA (1994) Lectures on R-matrices, Poisson–Lie groups and integrable systems. In: Babelon O, Cartier P, and Kosmann-Schwarzbach Y (eds.) *Lectures on Integrable Systems (Sophia-Antipolis, 1991)*, pp. 269–317. River Edge: World Scientific.
- Terng C-L and Uhlenbeck K (1998) Poisson actions and scattering theory for integrable systems. In: *Surveys in Differential Geometry: Integrable Systems*, pp. 315–402. Lectures on geometry and topology, sponsored by Lehigh University's *Journal of Differential Geometry*. A supplement to the *Journal of Differential Geometry*. Edited by Chuu Lian Terng and Karen Uhlenbeck. Surveys in Differential Geometry IV, Boston: International Press.

Clifford Algebras and Their Representations

A Trautman, Warsaw University, Warsaw, Poland

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Introductory and Historical Remarks

Clifford (1878) introduced his “geometric algebras” as a generalization of Grassmann algebras, complex numbers, and quaternions. Lipschitz (1886) was the first to define groups constructed from “Clifford numbers” and use them to represent rotations in a

Euclidean space. Cartan discovered representations of the Lie algebras $\mathfrak{so}_n(\mathbb{C})$ and $\mathfrak{so}_n(\mathbb{R})$, $n > 2$, that do not lift to representations of the orthogonal groups. In physics, Clifford algebras and spinors appear for the first time in Pauli’s nonrelativistic theory of the “magnetic electron.” Dirac (1928), in his work on the relativistic wave equation of the electron, introduced matrices that provide a representation of the Clifford algebra of Minkowski space. Brauer and Weyl (1935) connected the Clifford and Dirac ideas with Cartan’s spinorial representations of Lie algebras; they found, in any number of dimensions, the spinorial, projective representations of the orthogonal groups.

Clifford algebras and spinors are implicit in Euclid’s solution of the Pythagorean equation $x^2 - y^2 + z^2 = 0$, which is equivalent to

$$\begin{pmatrix} y-x & z \\ z & y+x \end{pmatrix} = 2 \begin{pmatrix} p \\ q \end{pmatrix} (p \ q) \quad [1]$$

and gives $x = q^2 - p^2, y = p^2 + q^2, z = 2pq$. If the numbers appearing in [1] are real, then this equation can be interpreted as providing a representation of a vector $(x, y, z) \in \mathbb{R}^3$, null with respect to a quadratic form of signature (1, 2), as the “square” of a spinor $(p, q) \in \mathbb{R}^2$. The pure spinors of Cartan (1938) provide a generalization of this observation to higher dimensions.

Multiplying the square matrix in [1] on the left by a real, 2×2 unimodular matrix, on the right by its transpose, and taking the determinant, one arrives at the exact sequence of group homomorphisms:

$$1 \rightarrow \mathbb{Z}_2 \rightarrow \text{SL}_2(\mathbb{R}) = \text{Spin}_{1,2}^0 \rightarrow \text{SO}_{1,2}^0 \rightarrow 1$$

Multiplying the same matrix by

$$\varepsilon = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad [2]$$

on the left and computing the square of the product, one obtains

$$\begin{pmatrix} z & x+y \\ x-y & -z \end{pmatrix}^2 = (x^2 - y^2 + z^2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This equation is an illustration of the idea of representing a quadratic form as the square of a linear form in a Clifford algebra. Replacing y by iy , one arrives at complex spinors, the Pauli matrices,

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = i\varepsilon, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$\text{Spin}_3 = \text{SU}_2$, etc.

This article reviews Clifford algebras, the associated groups, and their representations, for quadratic spaces over complex or real numbers. These notions have been generalized by Chevalley (1954) to quadratic spaces over arbitrary number fields.

Notation

If S is a vector space over $K = \mathbb{R}$ or \mathbb{C} , then S^* denotes its dual, that is, the vector space over K of all K -linear maps from S to K . The value of $\omega \in S^*$ on $s \in S$ is sometimes written as $\langle s, \omega \rangle$. The transpose of a linear map $f: S_1 \rightarrow S_2$ is the map $f^*: S_2^* \rightarrow S_1^*$ defined by $\langle s, f^*(\omega) \rangle = \langle f(s), \omega \rangle$ for

every $s \in S_1$ and $\omega \in S_2^*$. If S_1 and S_2 are complex vector spaces, then a map $f: S_1 \rightarrow S_2$ is said to be semilinear if it is \mathbb{R} -linear and $f(is) = -if(s)$. The complex conjugate of a finite-dimensional complex vector space S is the complex vector space \bar{S} of all semilinear maps from S^* to \mathbb{C} . There is a natural semilinear isomorphism (complex conjugation) $S \rightarrow \bar{S}$, $s \mapsto \bar{s}$ such that $\langle \omega, \bar{s} \rangle = \overline{\langle s, \omega \rangle}$ for every $\omega \in S^*$. The space \bar{S} can be identified with S and then $\bar{\bar{s}} = s$. The spaces $(\bar{S})^*$ and \bar{S}^* are identified. If $f: S_1 \rightarrow S_2$ is a complex-linear map, then there is the complex-conjugate map $\bar{f}: \bar{S}_1 \rightarrow \bar{S}_2$ given by $\bar{f}(\bar{s}) = \overline{f(s)}$ and the Hermitian conjugate map $f^\dagger \stackrel{\text{def}}{=} \bar{f}^*: S_1 \rightarrow \bar{S}_2^*$. A linear map $A: S \rightarrow \bar{S}^*$ such that $A^\dagger = A$ is said to be Hermitian. $K(N)$ denotes, for $K = \mathbb{R}, \mathbb{C}$ or \mathbb{H} , the set of all N by N matrices with elements in K .

Real, Complex, and Quaternionic Structures

A real structure on a complex vector space S is a complex-linear map $C: S \rightarrow \bar{S}$ such that $\bar{C}C = \text{id}_S$. A vector $s \in S$ is said to be real if $\bar{s} = C(s)$. The set of all real vectors is a real vector space; its real dimension is the same as the complex dimension of S .

A complex-linear map $C: S \rightarrow \bar{S}$ such that $\bar{C}C = -\text{id}_S$ defines on S a quaternionic structure; a necessary condition for such a structure to exist is that the complex dimension m of S be even, $m = 2n, n \in \mathbb{N}$. The space S with a quaternionic structure can be made into a right vector space over the field \mathbb{H} of quaternions. In the context of quaternions, it is convenient to represent the imaginary unit of \mathbb{C} as $\sqrt{-1}$. Multiplication on the right by the quaternion unit i is realized as the multiplication (on the left) by $\sqrt{-1}$. If j and $k = ij$ are the other two quaternion units and $s \in S$, then one puts $sj = \bar{C}(s)$ and $sk = sij$.

A real vector space S can be complexified by forming the tensor product $\mathbb{C} \otimes_{\mathbb{R}} S = S \oplus iS$.

The realification of a complex vector space S is the real vector space having S as its set of vectors so that $\dim_{\mathbb{R}} S = 2 \dim_{\mathbb{C}} S$. The complexification of a realification of S is the “double” $S \oplus S$ of the original space.

Inner-Product Spaces and Their Groups

Definitions: quadratic and symplectic spaces A bilinear map $B: S \times S \rightarrow K$ on a vector space S over K is said to make S into an inner-product space. To save on notation, one also writes $B: S \rightarrow S^*$ so that $\langle s, B(t) \rangle = B(s, t)$ for all $s, t \in S$. The group of automorphisms of an inner-product space,

$$\text{Aut}(S, B) = \{R \in \text{GL}(S) | R^* \circ B \circ R = B\}$$

is a Lie subgroup of the general linear group $\text{GL}(S)$. An inner-product space (S, B) is said here to be

quadratic (resp., symplectic) if B is symmetric (resp., antisymmetric and nonsingular). A quadratic space is characterized by its quadratic form $s \mapsto B(s, s)$. For $K = \mathbb{C}$, a Hermitian map $A: S \rightarrow \bar{S}^*$ defines a Hermitian scalar product $A(s, t) = \langle \bar{s}, A(t) \rangle$.

An orthogonal space is defined here as a quadratic space (S, B) such that $B: S \rightarrow S^*$ is an isomorphism. The group of automorphisms of an orthogonal space is the orthogonal group $O(S, B)$. The group of automorphisms of a symplectic space is the symplectic group $Sp(S, B)$. The dimension of a symplectic space is even. If $S = K^{2n}$ is a symplectic space over $K = \mathbb{R}$ or \mathbb{C} , then its symplectic group is denoted by $Sp_{2n}(K)$. Two quaternionic symplectic groups appear in the list of spin groups of low-dimensional spaces:

$$Sp_2(\mathbb{H}) = \{a \in \mathbb{H}(2) \mid a^\dagger a = I\}$$

and

$$Sp_{1,1}(\mathbb{H}) = \{a \in \mathbb{H}(2) \mid a^\dagger \sigma_z a = \sigma_z\}$$

Here a^\dagger denotes the matrix obtained from a by transposition and quaternionic conjugation.

Contractions, frames, and orthogonality From now on, unless otherwise specified, (V, g) is a quadratic space of dimension m . Let $\wedge V = \bigoplus_{p=0}^m \wedge^p V$ be its exterior (Grassmann) algebra. For every $v \in V$ and $w \in \wedge V$ there is the contraction $g(v)]w$ characterized as follows. The map $V \times \wedge V \rightarrow \wedge V$, $(v, w) \mapsto g(v)]w$, is bilinear; if $x \in \wedge^p V$, then $g(v)](x \wedge w) = (g(v)]x) \wedge w + (-1)^p x \wedge (g(v)]w)$ and $g(v)]v = g(v, v)$.

A frame (e_μ) in a quadratic space (V, g) is said to be a quadratic frame if $\mu \neq \nu$ implies $g(e_\mu, e_\nu) = 0$.

For every subset W of V there is the orthogonal subspace W^\perp containing all vectors that are orthogonal to every element of W .

If (V, g) is a real orthogonal space, then there is an orthonormal frame (e_μ) , $\mu = 1, \dots, m$, in V such that k frame vectors have squares equal to -1 , l frame vectors have squares equal to 1 and $k + l = m$. The pair (k, l) is the signature of g . The quadratic form g is said to be neutral if the orthogonal space (V, g) admits two maximal totally null subspaces W and W' such that $V = W \oplus W'$. Such a space V is $2n$ -dimensional, either complex or real with g of signature (n, n) . A Lorentzian space has maximal totally null subspaces of dimension 1 and a Euclidean space, characterized by a definite quadratic form, has no null subspaces. The Minkowski space is a Lorentzian space of dimension 4 .

If (V, g) is a complex orthogonal space, then an orthonormal frame (e_μ) , $\mu = 1, \dots, m$, can be

chosen in V so that, defining $g_{\mu\nu} = g(e_\mu, e_\nu)$, one has $g_{\mu\mu} = (-1)^{\mu+1}$ and, if $\mu \neq \nu$, then $g_{\mu\nu} = 0$.

If $A: S \rightarrow \bar{S}^*$ is a Hermitian isomorphism, then there is a (pseudo)unitary frame (e_α) in S such that the matrix $A_{\alpha\beta} = A(e_\alpha, e_\beta)$ is diagonal, has p 1 's and q -1 's on the diagonal, $p + q = \dim S$. If $p = q$, then A is said to be neutral. A is definite if either p or $q = 0$.

Algebras

Definitions An algebra over K is a vector space \mathcal{A} over K with a bilinear map $\mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}$, $(a, b) \mapsto ab$, which is distributive with respect to addition. The algebra is associative if $(ab)c = a(bc)$ holds for all $a, b, c \in \mathcal{A}$. It is commutative if $ab = ba$ for all $a, b \in \mathcal{A}$. An element $1_{\mathcal{A}}$ is the unit of \mathcal{A} if $1_{\mathcal{A}}a = a1_{\mathcal{A}} = a$ holds for every $a \in \mathcal{A}$.

From now on, unless otherwise specified, the bare word algebra denotes a finite-dimensional, associative algebra over $K = \mathbb{R}$ or \mathbb{C} , with a unit element. If S is an N -dimensional vector space over K , then the set $\text{End } S$ of all endomorphisms of S is an N^2 -dimensional algebra over K , the product being defined by composition; if $f, g \in \text{End } S$, then one writes fg instead of $f \circ g$; the unit of $\text{End } S$ is the identity map I . By definition, homomorphisms of algebras map units into units. The map $K \rightarrow \mathcal{A}$, $a \mapsto a1_{\mathcal{A}}$ is injective and one identifies K with its image in \mathcal{A} by this map so that the unit can be represented by $1 \in K \subset \mathcal{A}$. A set $\mathcal{B} \subset \mathcal{A}$ is said to generate \mathcal{A} if every element of \mathcal{A} can be represented as a linear combination of products of elements of \mathcal{B} . For example, if V is a vector space over K , then its tensor algebra

$$T(V) = \bigoplus_{p=0}^{\infty} \otimes^p V$$

is an (infinite-dimensional) algebra over K generated by $K \oplus V$. The algebra of all $N \times N$ matrices with entries in an algebra \mathcal{A} is denoted by $\mathcal{A}(N)$. Its unit element is the unit matrix I . In particular, $\mathbb{R}(N)$, $\mathbb{C}(N)$, and $\mathbb{H}(N)$ are algebras over \mathbb{R} . The algebra $\mathbb{R}(2)$ is generated by the set $\{\sigma_x, \sigma_z\}$. As a vector space, the algebra $\mathbb{R}(2)$ is spanned by the set $\{I, \sigma_x, \varepsilon, \sigma_z\}$.

The direct sum $\mathcal{A} \oplus \mathcal{B}$ of the algebras \mathcal{A} and \mathcal{B} over K is an algebra over K such that its underlying vector space is $\mathcal{A} \times \mathcal{B}$ and the product is defined by $(a, b) \cdot (a', b') = (aa', bb')$ for every $a, a' \in \mathcal{A}$ and $b, b' \in \mathcal{B}$. Similarly, the product in the tensor product algebra $\mathcal{A} \otimes_K \mathcal{B}$ is defined by

$$(a \otimes b) \cdot (a' \otimes b') = aa' \otimes bb' \quad [3]$$

For example, if \mathcal{A} is an algebra over \mathbb{R} , then the tensor product algebra $\mathbb{R}(N) \otimes_{\mathbb{R}} \mathcal{A}$ is isomorphic to $\mathcal{A}(N)$ and

$$K(N) \otimes_K K(N') = K(NN') \tag{4}$$

for $K = \mathbb{R}$ or \mathbb{C} and $N, N' \in \mathbb{N}$. There are isomorphisms of algebras over \mathbb{R} :

$$\begin{aligned} \mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} &= \mathbb{C} \oplus \mathbb{C} \\ \mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} &= \mathbb{C}(2) \\ \mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} &= \mathbb{R}(4) \end{aligned} \tag{5}$$

An algebra over \mathbb{R} can be complexified by complexifying its underlying vector space; it follows from [5] that $\mathbb{C}(2)$ is the complex algebra obtained by complexification of the real algebra \mathbb{H} .

The center of an algebra \mathcal{A} is the set

$$\mathcal{Z}(\mathcal{A}) = \{a \in \mathcal{A} \mid ab = ba \ \forall b \in \mathcal{A}\}$$

The center is a commutative subalgebra containing K . An algebra over K is said to be central if its center coincides with K . The algebras $\mathbb{R}(N)$ and $\mathbb{H}(N)$ are central over \mathbb{R} . The algebra $\mathbb{C}(N)$ is central over \mathbb{C} , but not over \mathbb{R} .

Simplicity and representations Let \mathcal{B}_1 and \mathcal{B}_2 be subsets of the algebra \mathcal{A} . Define $\mathcal{B}_1\mathcal{B}_2 = \{b_1b_2 \mid b_1 \in \mathcal{B}_1, b_2 \in \mathcal{B}_2\}$. A vector subspace \mathcal{B} of \mathcal{A} is said to be a left (resp., right) ideal of \mathcal{A} if $\mathcal{A}\mathcal{B} \subset \mathcal{B}$ (resp., $\mathcal{B}\mathcal{A} \subset \mathcal{B}$). A two-sided ideal – or simply an ideal – is a left and right ideal. An algebra $\mathcal{A} \neq \{0\}$ is said to be simple if its only two-sided ideals are $\{0\}$ and \mathcal{A} .

For example, the algebras $\mathbb{R}(N)$ and $\mathbb{H}(N)$ are simple over \mathbb{R} ; the algebra $\mathbb{C}(N)$ is simple when considered as an algebra over both \mathbb{R} and \mathbb{C} ; every associative, finite-dimensional simple algebra over \mathbb{R} or \mathbb{C} is isomorphic to one of them.

A representation of an algebra \mathcal{A} over K in a vector space S over K is a homomorphism of algebras $\rho: \mathcal{A} \rightarrow \text{End } S$. If ρ is injective, then the representation is said to be faithful. For example, the regular representation $\rho: \mathcal{A} \rightarrow \text{End } \mathcal{A}$ of an algebra \mathcal{A} , defined by $\rho(a)b = ab$ for all $a, b \in \mathcal{A}$, is faithful. A vector subspace T of the vector space S carrying a representation ρ of \mathcal{A} is said to be invariant for ρ if $\rho(a)T \subset T$ for every $a \in \mathcal{A}$; it is proper if distinct from both $\{0\}$ and S . For example, a left ideal of \mathcal{A} is invariant for the regular representation. Given an invariant subspace T of ρ one can reduce ρ to T by forming the representation $\rho_T: \mathcal{A} \rightarrow \text{End } T$, where $\rho_T(a)s = \rho(a)s$ for every $a \in \mathcal{A}$ and $s \in T$. A representation is irreducible if it has no proper invariant subspaces.

A linear map $F: S_1 \rightarrow S_2$ is said to intertwine the representations $\rho_1: \mathcal{A} \rightarrow \text{End } S_1$ and $\rho_2: \mathcal{A} \rightarrow \text{End } S_2$ if $F\rho_1(a) = \rho_2(a)F$ holds for every $a \in \mathcal{A}$. If F is an

isomorphism, then the representations ρ_1 and ρ_2 are said to be equivalent, $\rho_1 \sim \rho_2$. The following two propositions are classical:

Proposition (A)

- (i) *An algebra over K is simple if and only if it admits a faithful irreducible representation in a vector space over K . Such a representation is unique, up to equivalence.*
- (ii) *The complexification of a central simple algebra over \mathbb{R} is a central simple algebra over \mathbb{C} .*

For real algebras, one often considers complex representations, that is, representations in complex vector spaces. Two such representations $\rho_1: \mathcal{A} \rightarrow \text{End } S_1$ and $\rho_2: \mathcal{A} \rightarrow \text{End } S_2$ are said to be complex equivalent if there is a complex isomorphism $F: S_1 \rightarrow S_2$ intertwining the representations; they are real equivalent if there is an isomorphism among the realifications of S_1 and S_2 , intertwining the representations. For example, \mathbb{C} , considered as an algebra over \mathbb{R} , has two complex-inequivalent representations in \mathbb{C} : the identity representation and its complex conjugate. The realifications of these representations, given by $i \mapsto \varepsilon$ and $i \mapsto -\varepsilon$, respectively, are real equivalent: they are intertwined by σ_z . The real algebra \mathbb{H} , being central simple, has only one, up to complex equivalence, representation in \mathbb{C}^2 : every such representation is equivalent to the one given by

$$i \mapsto \sigma_x/\sqrt{-1}, \quad j \mapsto \sigma_y/\sqrt{-1}, \quad k \mapsto \sigma_z/\sqrt{-1}$$

This representation extends to an injective homomorphism of algebras $i: \mathbb{H}(N) \rightarrow \mathbb{C}(2N)$ which is used to define the quaternionic determinant of a matrix $a \in \mathbb{H}(N)$ as $\det_{\mathbb{H}}(a) = \det i(a)$, so that $\det_{\mathbb{H}}(a) \geq 0$ and $\det_{\mathbb{H}}(ab) = \det_{\mathbb{H}}(a)\det_{\mathbb{H}}(b)$ for every $a, b \in \mathbb{H}(N)$. In particular, if $q \in \mathbb{H}$ and $\lambda, \mu \in \mathbb{R}$, then $\det_{\mathbb{H}}(q) = \bar{q}q$ and

$$\det_{\mathbb{H}} \begin{pmatrix} \lambda & q \\ -\bar{q} & \mu \end{pmatrix} = (\lambda\mu + \bar{q}q)^2 \tag{6}$$

There are quaternionic unimodular groups $\text{SL}_N(\mathbb{H}) = \{a \in \mathbb{H}(N) \mid \det_{\mathbb{H}}(a) = 1\}$. For example, the group $\text{SL}_1(\mathbb{H})$ is isomorphic to SU_2 and $\text{SL}_2(\mathbb{H})$ is a noncompact, 15-dimensional Lie group, one of the spin groups in six dimensions.

Antiautomorphisms and inner products An automorphism of an algebra \mathcal{A} is a linear isomorphism $\alpha: \mathcal{A} \rightarrow \mathcal{A}$ such that $\alpha(ab) = \alpha(a)\alpha(b)$. An invertible element $c \in \mathcal{A}$ defines an inner automorphism $\text{Ad}(c) \in \text{GL}(\mathcal{A})$, $\text{Ad}(c)a = cac^{-1}$. Complex conjugation in \mathbb{C} , considered as an algebra over \mathbb{R} , is an automorphism that is not inner. An antiautomorphism of an

algebra \mathcal{A} is a linear isomorphism $\beta: \mathcal{A} \rightarrow \mathcal{A}$ such that $\beta(ab) = \beta(b)\beta(a)$ for all $a, b \in \mathcal{A}$. An (anti)automorphism β is involutive if $\beta^2 = \text{id}$. For example, conjugation of quaternions defines an involutive antiautomorphism of \mathbb{H} .

Let $\rho: \mathcal{A} \rightarrow \text{End } S$ be a representation of an algebra with an involutive antiautomorphism β . There is then the contragredient representation $\check{\rho}: \mathcal{A} \rightarrow \text{End } S^*$ given by $\check{\rho}(a) = (\rho(\beta(a)))^*$. If, moreover, \mathcal{A} is central simple and ρ is faithful irreducible, then there is an isomorphism $B: S \rightarrow S^*$ intertwining ρ and $\check{\rho}$ which is either symmetric, $B^* = B$, or antisymmetric, $B^* = -B$. It defines on S the structure of an inner-product space. This structure extends to $\text{End } S$: there is a symmetric isomorphism $B \otimes B^{-1}: \text{End } S \rightarrow (\text{End } S)^* = \text{End } S^*$ given, for every $f \in \text{End } S$, by $(B \otimes B^{-1})(f) = BfB^{-1}$.

Let $K^\times = K \setminus \{0\}$ be the multiplicative group of the field K . Given a simple algebra \mathcal{A} with an involutive antiautomorphism β , one defines $N(a) = \beta(a)a$ and the group

$$\mathcal{G}(\beta) = \{a \in \mathcal{A} \mid N(a) \in K^\times\}$$

Let $\rho: \mathcal{A} \rightarrow \text{End } S$ be the faithful irreducible representation as above, then, for $a \in \mathcal{A}$ and $s, t \in S$, one has

$$B(\rho(a)s, \rho(a)t) = N(a)B(s, t)$$

If $a \in \mathcal{G}(\beta)$ and $\lambda \in K^\times$, then $\lambda a \in \mathcal{G}(\beta)$ and the norm N satisfies $N(\lambda a) = \lambda^2 N(a)$. The inner product B is invariant with respect to the action of the group

$$\mathcal{G}_1(\beta) = \{a \in \mathcal{G}(\beta) \mid N(a) = 1\}$$

Proposition (B) *Let \mathcal{A} be a central simple algebra over K with an involutive antiautomorphism β and a faithful irreducible representation ρ so that*

$$\check{\rho}(a) = B\rho(a)B^{-1}$$

The map $h: \mathcal{A} \times \mathcal{A} \rightarrow K$ defined by

$$h(a, b) = \text{tr } \rho(\beta(a)b)$$

is bilinear, symmetric, and nondegenerate. The map ρ is an isometry of the quadratic space (\mathcal{A}, h) on its image in the quadratic space $(\text{End } S, B \otimes B^{-1})$.

Graded Algebras

Definitions An algebra \mathcal{A} is said to be \mathbb{Z} -graded (resp., \mathbb{Z}_2 -graded) if there is a decomposition of the underlying vector space $\mathcal{A} = \bigoplus_{p \in \mathbb{Z}} \mathcal{A}^p$ (resp., $\mathcal{A} = \mathcal{A}^0 \oplus \mathcal{A}^1$) such that $\mathcal{A}^p \mathcal{A}^q \subset \mathcal{A}^{p+q}$. In a \mathbb{Z}_2 -graded algebra, it is understood that $p + q$ is reduced mod 2. If $a \in \mathcal{A}^p$, then a is said to be homogeneous of degree p . The exterior algebra $\wedge V$ of a vector space V is \mathbb{Z} -graded. Every \mathbb{Z} -graded algebra becomes \mathbb{Z}_2 -graded

when one reduces the degree of every element mod 2. A graded isomorphism of graded algebras is an isomorphism that preserves the grading.

A \mathbb{Z}_2 -grading of \mathcal{A} is characterized by the involutive automorphism α such that, if $a \in \mathcal{A}^p$, then $\alpha(a) = (-1)^p a$. From now on, grading means \mathbb{Z}_2 -grading unless otherwise specified. The elements of \mathcal{A}^0 (resp., \mathcal{A}^1) are said to be even (resp., odd). It is often convenient to denote the graded algebra as

$$\mathcal{A}^0 \rightarrow \mathcal{A} \tag{7}$$

Given such an algebra over K and $N \in \mathbb{N}$, one constructs the graded algebra $\mathcal{A}^0(N) \rightarrow \mathcal{A}(N)$. Two graded algebras over K , $\mathcal{A}^0 \rightarrow \mathcal{A}$ and $\mathcal{A}'^0 \rightarrow \mathcal{A}'$ are said to be of the same type if there are integers N and N' such that the algebras $\mathcal{A}^0(N) \rightarrow \mathcal{A}(N)$ and $\mathcal{A}'^0(N') \rightarrow \mathcal{A}'(N')$ are graded isomorphic. The property of being of the same type is an equivalence relation in the set of all graded algebras over K .

Given an algebra \mathcal{A} , one constructs two ‘‘canonical’’ graded algebras as follows:

1. the double algebra

$$\mathcal{A} \rightarrow \mathcal{A} \oplus \mathcal{A}$$

graded by the ‘‘swap’’ automorphism, $\alpha(a_1, a_2) = (a_2, a_1)$ for $a_1, a_2 \in \mathcal{A}$;

2. the algebra

$$\mathcal{A} \oplus \mathcal{A} \rightarrow \mathcal{A}(2)$$

is defined by declaring the diagonal (resp., anti-diagonal) elements of $\mathcal{A}(2)$ to be even (resp., odd).

The real algebra $\mathbb{R}(2)$ has also another grading, given by the involutive automorphism α such that $\alpha(a) = \varepsilon a \varepsilon^{-1}$, where $a \in \mathbb{R}(2)$ and ε is as in [2]. In this case, [7] reads

$$\mathbb{C} \rightarrow \mathbb{R}(2)$$

There are also graded algebras over \mathbb{R} :

$$\mathbb{R} \rightarrow \mathbb{C}, \quad \mathbb{C} \rightarrow \mathbb{H}, \quad \text{and} \quad \mathbb{H} \rightarrow \mathbb{C}(2)$$

The grading of the last algebra can be defined by declaring the Pauli matrices and iI to be odd.

Super Lie algebras A super Lie algebra is a graded algebra \mathcal{A} such that the product $(a, b) \mapsto [a, b]$ is super anticommutative, $[a, b] = -(-1)^{pq}[b, a]$, and satisfies the super Jacobi identity,

$$[a, [b, c]] = [[a, b], c] + (-1)^{pq}[b, [a, c]]$$

for every $a \in \mathcal{A}^p$, $b \in \mathcal{A}^q$ and $c \in \mathcal{A}$. To every graded associative algebra \mathcal{A} there corresponds a super Lie algebra \mathcal{GLA} : its underlying vector space and grading are as in \mathcal{A} and the product, for $a \in \mathcal{A}^p$

and $b \in \mathcal{A}^q$, is given as the supercommutator $[a, b] = ab - (-1)^{pq}ba$.

Supercentrality and graded simplicity A graded algebra \mathcal{A} over K is supercentral if $\mathcal{Z}(\mathcal{A}) \cap \mathcal{A}^0 = K$. The algebra $\mathbb{R} \rightarrow \mathbb{C}$ is supercentral, but the real ungraded algebra \mathbb{C} is not central.

A subalgebra \mathcal{B} of a graded algebra \mathcal{A} is said to be a graded subalgebra if $\mathcal{B} = \mathcal{B} \cap \mathcal{A}^0 \oplus \mathcal{B} \cap \mathcal{A}^1$. A graded ideal of \mathcal{A} is an ideal that is a graded subalgebra. A graded algebra $\mathcal{A} \neq \{0\}$ is said to be graded simple if it has no graded ideals other than $\{0\}$ and \mathcal{A} . The double algebra of a simple algebra is graded simple, but not simple.

The graded tensor product Let \mathcal{A} and \mathcal{B} be graded algebras; the tensor product of their underlying vector spaces admits a natural grading, $(\mathcal{A} \otimes \mathcal{B})^p = \bigoplus_q \mathcal{A}^q \otimes \mathcal{B}^{p-q}$. The product defined in [3] makes $\mathcal{A} \otimes \mathcal{B}$ into a graded algebra. There is another “super” product in the same graded vector space given by

$$(a \otimes b) \cdot (a' \otimes b') = (-1)^{pq} a a' \otimes b b'$$

for $a' \in \mathcal{A}^p$ and $b \in \mathcal{B}^q$. The resulting graded algebra is referred to as the graded tensor product and denoted by $\mathcal{A} \hat{\otimes} \mathcal{B}$. For example, if V and W are vector spaces, then the Grassmann algebra $\wedge(V \oplus W)$ is isomorphic to $\wedge V \hat{\otimes} \wedge W$.

Clifford Algebras

Definitions: The Universal Property and Grading

The Clifford algebra associated with a quadratic space (V, g) is the quotient algebra

$$\mathcal{C}\ell(V, g) = \mathcal{T}(V) / \mathcal{J}(V, g) \tag{8}$$

where $\mathcal{J}(V, g)$ is the ideal in the tensor algebra $\mathcal{T}(V)$ generated by all elements of the form $v \otimes v - g(v, v)1_{\mathcal{T}(V)}$, $v \in V$.

The Clifford algebra is associative with a unit element denoted by 1. One denotes by κ the canonical map of $\mathcal{T}(V)$ onto $\mathcal{C}\ell(V, g)$ and by ab the product of two elements $a, b \in \mathcal{C}\ell(V, g)$ so that $\kappa(P \otimes Q) = \kappa(P)\kappa(Q)$ for $P, Q \in \mathcal{T}(V)$. The map κ is injective on $K \oplus V$, and one identifies this subspace of $\mathcal{T}(V)$ with its image under κ . With this identification, for all $u, v \in V$, one has

$$uv + vu = 2g(u, v)$$

Clifford algebras are characterized by their universal property described in the following proposition.

Proposition (C) *Let \mathcal{A} be an algebra with a unit $1_{\mathcal{A}}$ and let $f: V \rightarrow \mathcal{A}$ be a Clifford map, that is, a linear*

map such that $f(v)^2 = g(v, v)1_{\mathcal{A}}$ for every $v \in V$. There then exists a homomorphism $\hat{f}: \mathcal{C}\ell(V, g) \rightarrow \mathcal{A}$ of algebras with units, an extension of f , so that $f(v) = \hat{f}(v)$ for every $v \in V$.

As a corollary, one obtains

Proposition (D) *If f is an isometry of (V, g) into (W, h) , then there is a homomorphism of algebras $\mathcal{C}\ell(f): \mathcal{C}\ell(V, g) \rightarrow \mathcal{C}\ell(W, h)$ extending f so that there is the commutative diagram*

$$\begin{array}{ccc} \mathcal{C}\ell(V, g) & \xrightarrow{\mathcal{C}\ell(f)} & \mathcal{C}\ell(W, h) \\ \uparrow & & \uparrow \\ V & \xrightarrow{f} & W \end{array}$$

For example, the isometry $v \mapsto -v$ extends to the involutive main automorphism α of $\mathcal{C}\ell(V, g)$, defining its \mathbb{Z}_2 -grading:

$$\mathcal{C}\ell(V, g) = \mathcal{C}\ell^0(V, g) \oplus \mathcal{C}\ell^1(V, g)$$

The algebra $\mathcal{C}\ell(V, g)$ admits also an involutive canonical antiautomorphism β characterized by $\beta(1) = 1$ and $\beta(v) = v$ for every $v \in V$.

The Vector Space Structure of Clifford Algebras

Referring to proposition (D), let $\mathcal{A} = \text{End}(\wedge V)$ and, for every $v \in V$ and $w \in \wedge V$, put $f(v)w = v \wedge w + g(v)w$, then $f: V \rightarrow \text{End}(\wedge V)$ is a Clifford map and the map

$$i: \mathcal{C}\ell(V, g) \rightarrow \wedge V \tag{9}$$

given by $i(a) = \hat{f}(a)1_{\wedge V}$ is an isomorphism of vector spaces. This proves

Proposition (E) *As a vector space, the algebra $\mathcal{C}\ell(V, g)$ is isomorphic to the exterior algebra $\wedge V$.*

If V is m -dimensional, then $\mathcal{C}\ell(V, g)$ is 2^m -dimensional. The linear isomorphism [9] defines a \mathbb{Z} -grading of the vector space underlying the Clifford algebra: if $i(a_k) \in \wedge^k V$, then a_k is said to be of Grassmann degree k . Every element $a \in \mathcal{C}\ell(V, g)$ decomposes into its Grassmann components, $a = \sum_{k \in \mathbb{Z}} a_k$. The Clifford product of two elements of Grassmann degrees k and l decomposes as follows: $a_k b_l = \sum_{p \in \mathbb{Z}} (a_k b_l)_p$, and $(a_k b_l)_p = 0$ if $p < |k - l|$ or $p \equiv k - l + 1 \pmod{2}$ or $p > m - |m - k - l|$.

One often uses [9] to identify the vector spaces $\wedge V$ and $\mathcal{C}\ell(V, g)$; this having been done, one can write, for every $v \in V$ and $a \in \mathcal{C}\ell(V, g)$,

$$va = v \wedge a + g(v)a \tag{10}$$

so that $[v, a] = 2g(v)a$, where $[,]$ is the supercommutator. It defines a super Lie algebra structure in the vector space $K \oplus V$. The quadratic form defined by g need not be nondegenerate; for example, if it is the

0-form, then [10] shows that the Clifford and exterior multiplications coincide and $\mathcal{Cl}(V, 0)$ is isomorphic, as an algebra, to the Grassmann algebra.

Complexification of Real Clifford Algebras

Proposition (F) *If (V, g) is a real quadratic space, then the algebras $\mathbb{C} \otimes \mathcal{Cl}(V, g)$ and $\mathcal{Cl}(\mathbb{C} \otimes V, \mathbb{C} \otimes g)$ are isomorphic, as graded algebras over \mathbb{C} .*

From now on, through the end of the article, one assumes that (V, g) is an orthogonal space over $K = \mathbb{R}$ or \mathbb{C} .

The Clifford algebra associated with the orthogonal space \mathbb{C}^m is denoted by \mathcal{Cl}_m . The Clifford algebra associated with the orthogonal space (\mathbb{R}^{k+l}, g) , where g is of signature (k, l) , is denoted by $\mathcal{Cl}_{k,l}$, so that $\mathbb{C} \otimes \mathcal{Cl}_{k,l} = \mathcal{Cl}_{k+l}$.

Relations between Clifford Algebras in Spaces of Adjacent Dimensions

Consider an orthogonal space (V, g) over K and the one-dimensional orthogonal space (K, h_1) , having a unit vector $w \in K$, $h_1(w, w) = \varepsilon$, where $\varepsilon = 1$ or -1 . The map $V \ni v \mapsto vw \in \mathcal{Cl}^0(V \oplus K, g \oplus h_1)$ satisfies $(vw)^2 = -\varepsilon g(v, v)$ and extends to the isomorphism of algebras $\mathcal{Cl}(V, -\varepsilon g) \rightarrow \mathcal{Cl}^0(V \oplus K, g \oplus h_1)$. This proves

Proposition (G) *There are isomorphisms of algebras: $\mathcal{Cl}_m \rightarrow \mathcal{Cl}_{m+1}^0$ and $\mathcal{Cl}_{k,l} \rightarrow \mathcal{Cl}_{k+1,l}^0$.*

Consider the orthogonal space (K^2, h) with a neutral h such that, for $\lambda, \mu \in K$, one has $\langle (\lambda, \mu), h(\lambda, \mu) \rangle = \lambda\mu$. The map

$$K^2 \rightarrow K(2), \quad (\lambda, \mu) \mapsto \begin{pmatrix} 0 & \lambda \\ \mu & 0 \end{pmatrix}$$

has the Clifford property and establishes the isomorphisms represented by the horizontal arrows in the diagram

$$\begin{array}{ccc} \mathcal{Cl}(K^2, h) & \rightarrow & K(2) \\ \uparrow & & \uparrow \\ \mathcal{Cl}^0(K^2, h) & \rightarrow & K \oplus K \end{array} \quad [11]$$

Proposition (H) *If (K^2, h) is neutral and (V, g) is over K , then the algebra $\mathcal{Cl}(V \oplus K^2, g \oplus h)$ is isomorphic to the algebra $\mathcal{Cl}(V, g) \otimes K(2)$. Specifically, there are isomorphisms*

$$\begin{aligned} \mathcal{Cl}_{k+1,l+1} &= \mathcal{Cl}_{k,l} \otimes \mathbb{R}(2) \\ \mathcal{Cl}_{m+2} &= \mathcal{Cl}_m \otimes \mathbb{C}(2) \end{aligned} \quad [12]$$

The Chevalley Theorem and the Brauer-Wall Group

If (V, g) and (W, h) are quadratic spaces over K , then their sum is the quadratic space $(V \oplus W, g \oplus h)$ characterized by $g \oplus h: V \oplus W \rightarrow V^* \oplus W^*$ so that $(g \oplus h)(v, w) = (g(v), h(w))$. By noting that the map $V \oplus W \ni (v, w) \mapsto v \otimes 1 + 1 \otimes w \in \mathcal{Cl}(V, g) \hat{\otimes} \mathcal{Cl}(W, h)$ has the Clifford property, Chevalley proved

Proposition (I) *The algebra $\mathcal{Cl}(V \oplus W, g \oplus h)$ is isomorphic to the algebra $\mathcal{Cl}(V, g) \hat{\otimes} \mathcal{Cl}(W, h)$.*

The type of the (graded) algebra $\mathcal{Cl}(V \oplus W, g \oplus h)$ depends only on the types of $\mathcal{Cl}(V, g)$ and $\mathcal{Cl}(W, h)$. The Chevalley theorem (I) shows that the set of types of Clifford algebras over K forms an abelian group for a multiplication induced by the graded tensor product. The unit of this Brauer-Wall group of K is the type of the algebra $\mathcal{Cl}(K^2, h)$ described in [11]; for a full account with proofs, see Wall (1963).

The Volume Element and the Centers

Let $e = (e_\mu)$ be an orthonormal frame in (V, g) . The volume element associated with e is

$$\eta = e_1 e_2 \cdots e_m$$

If η' is the volume element associated with another orthonormal frame e' in the same orthogonal space, then either $\eta' = \eta$ (e and e' are of the same orientation) or $\eta' = -\eta$ (e and e' are of opposite orientation). For $K = \mathbb{C}$, one has $\eta^2 = 1$; for $K = \mathbb{R}$ and g of signature (k, l) one has

$$\eta^2 = (-1)^{(1/2)(k-l)(k-l+1)} \quad [13]$$

It is convenient to define $\iota \in \{1, i\}$ so that $\eta^2 = \iota^2$. For every $v \in V$ one has $v\eta = (-1)^{m+1}\eta v$. The structure of the centers of Clifford algebras is as follows:

Proposition (J) *If m is even, then $\mathcal{Z}(\mathcal{Cl}(V, g)) = K$ and $\mathcal{Z}(\mathcal{Cl}^0(V, g)) = K \oplus K\eta$. If m is odd, then $\mathcal{Z}(\mathcal{Cl}(V, g)) = K \oplus K\eta$ and $\mathcal{Z}(\mathcal{Cl}^0(V, g)) = K$.*

The graded algebra $\mathcal{Cl}(V, g)$ is supercentral for every m .

The Structure of Clifford Algebras

The complex case Using [4] one obtains from [11] and [12] the isomorphisms of algebras

$$\mathcal{Cl}_{2n+1}^0 = \mathcal{Cl}_{2n} = \mathbb{C}(2^n) \quad [14]$$

$$\mathcal{Cl}_{2n+1} = \mathcal{Cl}_{2n+2}^0 = \mathbb{C}(2^n) \oplus \mathbb{C}(2^n) \quad [15]$$

for $n = 0, 1, 2, \dots$. Therefore, there are only two types of complex Clifford algebras, represented by $\mathbb{C} \rightarrow \mathbb{C} \oplus \mathbb{C}$ and $\mathbb{C} \oplus \mathbb{C} \rightarrow \mathbb{C}(2)$: the Brauer-Wall group of \mathbb{C} is \mathbb{Z}_2 .

The real case In view of proposition (I) and $\mathcal{Cl}_{1,1} = \mathbb{R}(2)$, the algebra $\mathcal{Cl}_{k,l}$ is of the same type as $\mathcal{Cl}_{k-l,0}$ if $k > l$ and of the same type as $\mathcal{Cl}_{0,l-k}$ if $k < l$. Since $\mathcal{Cl}_{k,l} \hat{\otimes} \mathcal{Cl}_{l,k} = \mathcal{Cl}_{k+l,k+l}$, the type of $\mathcal{Cl}_{l,k}$ is the inverse of the type of $\mathcal{Cl}_{k,l}$. The algebra $\mathcal{Cl}_{4,0}^0 \rightarrow \mathcal{Cl}_{4,0}$ is isomorphic to $\mathbb{H} \oplus \mathbb{H} \rightarrow \mathbb{H}(2)$: if $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \subset \mathcal{Cl}_{4,0}$, and $q = ix_1 + jx_2 + kx_3 + x_4 \in \mathbb{H}$, then an isomorphism is obtained from the Clifford map f ,

$$f(x) = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix} \quad [16]$$

In view of [13], the volume element η satisfies $\eta^2 = 1$. By replacing $-\bar{q}$ with \bar{q} in [16], one shows that $\mathcal{Cl}_{0,4}$ is also isomorphic to $\mathbb{H}(2)$. The map $\mathbb{R}^4 \times \mathbb{R}^{k+l} \rightarrow \mathbb{H}(2) \otimes \mathcal{Cl}_{k,l}$ given by $(x, y) \mapsto f(x) \otimes 1 + \eta \otimes y$ has the Clifford property and establishes the isomorphism of algebras $\mathcal{Cl}_{k+4,l} = \mathbb{H} \otimes \mathcal{Cl}_{k,l}$. Since, similarly, $\mathcal{Cl}_{k,l+4} = \mathbb{H} \otimes \mathcal{Cl}_{k,l}$, one obtains the isomorphism

$$\mathcal{Cl}_{k+4,l} = \mathcal{Cl}_{k,l+4}$$

Therefore,

$$\mathcal{Cl}_{k+8,l} = \mathcal{Cl}_{k+4,l+4} = \mathcal{Cl}_{k,l+8} = \mathcal{Cl}_{k,l} \otimes \mathbb{R}(16)$$

and the algebras $\mathcal{Cl}_{k,l}, \mathcal{Cl}_{k+8,l}$, and $\mathcal{Cl}_{k,l+8}$ are all of the same type. This double periodicity of period 8 is subsumed by saying that real Clifford algebras can be arranged on a “spinorial chessboard.” The type of $\mathcal{Cl}_{k,l}^0 \rightarrow \mathcal{Cl}_{k,l}$ depends only on $k - l \pmod 8$; the eight types have the following low-dimensional algebras as representatives: $\mathcal{Cl}_{1,0}, \mathcal{Cl}_{2,0}, \mathcal{Cl}_{3,0}, \mathcal{Cl}_{4,0} = \mathcal{Cl}_{0,4}, \mathcal{Cl}_{0,3}, \mathcal{Cl}_{0,2}$, and $\mathcal{Cl}_{0,1}$. The Brauer–Wall group of \mathbb{R} is \mathbb{Z}_8 , generated by the type of $\mathcal{Cl}_{1,0}^0 \rightarrow \mathcal{Cl}_{1,0}$, that is, by $\mathbb{R} \rightarrow \mathbb{C}$. Bearing in mind the isomorphism $\mathcal{Cl}_{k,l} = \mathcal{Cl}_{k+1,l}^0$ and abbreviating $\mathbb{C} \rightarrow \mathbb{R}(2)$ to $\mathbb{C} \rightarrow \mathbb{R}$, etc., one can arrange the types of real Clifford algebras in the form of a “spinorial clock”:

$$\begin{array}{ccccc} \mathbb{R} & \xrightarrow{7} & \mathbb{R} \oplus \mathbb{R} & \xrightarrow{0} & \mathbb{R} \\ 6 \uparrow & & & & \downarrow 1 \\ \mathbb{C} & & & & \mathbb{C} \\ 5 \uparrow & & & & \downarrow 2 \\ \mathbb{H} & \xleftarrow{4} & \mathbb{H} \oplus \mathbb{H} & \xleftarrow{3} & \mathbb{H} \end{array} \quad [17]$$

Proposition (K) *Recipe for determining $\mathcal{Cl}_{k,l}^0 \rightarrow \mathcal{Cl}_{k,l}$:*

- (i) find the integers μ and ν such that $k - l = 8\mu + \nu$ and $0 \leq \nu < 8$;
- (ii) from the spinorial clock, read off $\mathcal{A}_\nu^0 \rightarrow \nu \mathcal{A}_\nu$ and compute the real dimensions, $\dim \mathcal{A}_\nu^0 = 2^{\tau^0}$ and $\dim \mathcal{A}_\nu = 2^\tau$; and
- (iii) form $\mathcal{Cl}_{k,l}^0 = \mathcal{A}_\nu^0(2^{(1/2)(k+l-1-\tau^0)})$ and $\mathcal{Cl}_{k,l} = \mathcal{A}_\nu(2^{(1/2)(k+l-\tau)})$.

The spinorial clock is symmetric with respect to the reflection in the vertical line through its center; this is a consequence of the isomorphism of algebras $\mathcal{Cl}_{k,l+2} = \mathcal{Cl}_{l,k} \otimes \mathbb{R}(2)$.

Note that the “abstract” algebra $\mathcal{Cl}_{k,l}$ carries, in general, less information than the Clifford algebra defined in [8], which contains V as a distinguished vector subspace with the quadratic form $v \mapsto v^2 = g(v, v)$. For example, the algebras $\mathcal{Cl}_{8,0}, \mathcal{Cl}_{4,4}$, and $\mathcal{Cl}_{0,8}$ are all graded isomorphic.

Theorem on Simplicity

From general theory (Chevalley 1954) or by inspection of [14], [15], and [17], one has

Proposition (L) *Let m be the dimension of the orthogonal space (V, g) over K .*

- (i) *If m is even (resp., odd), then the algebra $\mathcal{Cl}(V, g)$ (resp., $\mathcal{Cl}^0(V, g)$) over K is central simple.*
- (ii) *If $K = \mathbb{C}$ and m is odd (resp., even), then the algebra $\mathcal{Cl}(V, g)$ (resp., $\mathcal{Cl}^0(V, g)$) is the direct sum of two isomorphic complex central simple algebras.*
- (iii) *If $K = \mathbb{R}$ and m is odd (resp., even), then the algebra $\mathcal{Cl}(V, g)$ (resp., $\mathcal{Cl}^0(V, g)$) when $\eta^2 = 1$ is the direct sum of two isomorphic central simple algebras and when $\eta^2 = -1$ is simple with a center isomorphic to \mathbb{C} .*

Representations

The Pauli, Cartan, Dirac, and Weyl Representations

Odd dimensions Let (V, g) be of dimension $m = 2n + 1$ over K . From propositions (A) and (L) it follows that the central simple algebra $\mathcal{Cl}^0(V, g)$ has a unique, up to equivalence, faithful, and irreducible representation in the complex 2^n -dimensional vector space S of Pauli spinors. By putting $\sigma(\eta) = \iota I$ it is extended to a Pauli representation $\sigma: \mathcal{Cl}(V, g) \rightarrow \text{End } S$. Given an orthonormal frame (e_μ) in V , Pauli endomorphisms (matrices if S is identified with \mathbb{C}^{2^n}) are defined as $\sigma_\mu = \sigma(e_\mu) \in \text{End } S$. The representations σ and $\sigma \circ \alpha$ are complex inequivalent. For $K = \mathbb{C}$ none of them is faithful; their direct sum is the faithful Cartan representation of $\mathcal{Cl}(V, g)$ in $S \oplus S$. For $K = \mathbb{R}$ and $(1/2)(k - l - 1)$ even, the representations σ and $\sigma \circ \alpha$ are real equivalent and faithful. On computing $\beta(\eta)$ one finds that the contragredient representation σ^\dagger is equivalent to σ for n even and to $\sigma \circ \alpha$ for n odd.

Even dimensions Similarly, for (V, g) of dimension $m = 2n$ over K , the central simple algebra $\mathcal{Cl}(V, g)$ has a unique, up to equivalence, faithful, and

irreducible representation $\gamma: \mathcal{C}\ell(V, g) \rightarrow \text{End } S$ in the 2^n -dimensional complex vector space S of Dirac spinors. The Dirac endomorphisms (matrices) are $\gamma_\mu = \gamma(e_\mu)$. Put $\Gamma = \iota\gamma(\eta)$ so that $\Gamma^2 = I$: the matrix Γ generalizes the familiar γ_5 . The Dirac representation γ restricted to $\mathcal{C}\ell^0(V, g)$ decomposes into the sum $\gamma_+ \oplus \gamma_-$ of two irreducible representations in the vector spaces

$$S_\pm = \{s \in S \mid \Gamma s = \pm s\}$$

of Weyl (chiral) spinors. The elements of S_+ are said to be of opposite chirality with respect to those of S_- . The transpose Γ^* defines a similar split of S^* . The representations γ_+ and γ_- are never complex-equivalent, but they are real equivalent and faithful for $K = \mathbb{R}$ and $(1/2)(k - l)$ odd.

The representations $\gamma \circ \alpha$ and $\check{\gamma}$ are both equivalent to γ . It is convenient to describe simultaneously the properties of the transpositions of the Pauli and Dirac matrices; let ρ_μ be either the Pauli matrices for V of dimension $2n + 1$ or the Dirac matrices for V of dimension $2n$. There is a complex isomorphism $B: S \rightarrow S^*$ such that

$$\rho_\mu^* = (-1)^n B \rho_\mu B^{-1} \tag{18}$$

In the case of the Dirac matrices, the factor $(-1)^n$ in [18] implies that this equation also holds for Γ in place of ρ_μ . The isomorphism B preserves (resp., changes) the chirality of Weyl spinors for n even (resp., odd). Every matrix of the form $B\gamma_{\mu_1} \dots \gamma_{\mu_p}$, where

$$1 \leq \mu_1 < \dots < \mu_p \leq 2n \tag{19}$$

is either symmetric or antisymmetric, depending on p and the symmetry of B . A simple argument, based on counting the number of such products of one symmetry, leads to the equation

$$B^* = (-1)^{(1/2)n(n+1)} B$$

valid in dimensions $2n$ and $2n + 1$.

Inner products on spinor spaces Let S be the complex vector space of Dirac or Pauli spinors associated with (V, g) over K . The isomorphism $B: S \rightarrow S^*$ defines on S an inner product $B(s, t) = \langle s, B(t) \rangle$, $s, t \in S$, which is orthogonal for $m \equiv 0, 1, 6$, or $7 \pmod 8$ and symplectic for $m \equiv 2, 3, 4$, or $5 \pmod 8$. For $m \equiv 0 \pmod 4$, this product restricts to an inner product on the space of Weyl spinors that is orthogonal for $m \equiv 0 \pmod 8$ and symplectic for $m \equiv 4 \pmod 8$. For $m \equiv 2 \pmod 4$, the map B defines the isomorphisms $B_\pm: S_\pm \rightarrow S_\mp^*$.

Example One of the most used representations $\gamma: \mathcal{C}\ell_{3,1} \rightarrow \mathbb{C}(4)$ is given by the Dirac matrices

$$\begin{aligned} \gamma_1 &= \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix}, & \gamma_2 &= \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix} \\ \gamma_3 &= \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix}, & \gamma_4 &= \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \end{aligned} \tag{20}$$

Change Conjugation and Majorana Spinors

Throughout this section and next, one assumes $K = \mathbb{R}$ so that, given a representation $\rho: \mathcal{C}\ell(V, g) \rightarrow \text{End } S$, one can form the complex- (“charge”) conjugate representation $\bar{\rho}: \mathcal{C}\ell(V, g) \rightarrow \text{End } \bar{S}$ defined by $\bar{\rho}(a) = \overline{\rho(a)}$ and the Hermitian conjugate representation $\rho^\dagger: \mathcal{C}\ell(V, g) \rightarrow \text{End } S^*$, where $\rho^\dagger(a) = \overline{\rho(a)}$.

Even dimensions The representations $\bar{\gamma}$ and γ are equivalent: there is an isomorphism $C: S \rightarrow \bar{S}$ such that

$$\bar{\gamma}_\mu = C \gamma_\mu C^{-1} \tag{21}$$

The automorphism $\bar{C}C$ is in the commutant of γ ; it is, therefore, proportional to I and, by a change of scale, one can achieve $\bar{C}C = I$ for $k - l \equiv 0$ or $6 \pmod 8$ and $\bar{C}C = -I$ for $k - l \equiv 2$ or $4 \pmod 8$.

The spinor $s_c = C^{-1}\bar{s} \in S$ is the charge conjugate of $s \in S$. If $\psi: V \rightarrow S$ is a solution of the Dirac equation

$$(\gamma^\mu(\partial_\mu - iqA_\mu) - \kappa)\psi = 0$$

for a particle of electric charge q , then ψ_c is a solution of the same equation with the opposite charge. Since

$$\bar{\Gamma} = \iota^2 C \Gamma C^{-1}$$

charge conjugation preserves (resp., changes) the chirality of Weyl spinors for $(1/2)(k - l)$ even (resp., odd).

If $\bar{C}C = I$, then

$$\text{Re } S = \{s \in S \mid s_c = s\}$$

is a real vector space of dimension 2^n , the space of Dirac–Majorana spinors. The representation γ is real: restricted to $\text{Re } S$ and expressed with respect to a frame in this space, it is given by real $2^n \times 2^n$ matrices. For $k - l \equiv 0 \pmod 8$ the representations γ_+ and γ_- are both real: in this case there are Weyl–Majorana spinors.

Odd dimensions On computing $\overline{\sigma(\eta)}$ one finds that the conjugate representation $\bar{\sigma}$ is equivalent to σ

(resp., $\sigma \circ \alpha$) if $\eta^2 = 1$ (resp., $\eta^2 = -1$). There is an isomorphism $C:S \rightarrow \bar{S}$ such that

$$\bar{\sigma}_\mu = (-1)^{(1/2)(k-l+1)} C\sigma_\mu C^{-1} \tag{22}$$

and $\bar{C}C = I$ (resp., $\bar{C}C = -I$) for $k - l \equiv 1$ or $7 \pmod 8$ (resp., $k - l \equiv 3$ or $5 \pmod 8$). For $k - l \equiv 1 \pmod 8$, the restriction of the Pauli representation to $\mathcal{C}\ell_{k,l}^0$ is real and the Pauli matrices are pure imaginary; for $k - l \equiv 7 \pmod 8$, the Pauli representations of $\mathcal{C}\ell_{k,l}$ are both real and so are the Pauli matrices. In both these cases there are Pauli–Majorana spinors.

Hermitian Scalar Products and Multivectors

For $m = k + l$ odd and C as in [22], the map $A = \bar{B}C:S \rightarrow \bar{S}^*$ intertwines the representations σ^\dagger and σ (resp., $\sigma \circ \alpha$) for k even (resp., odd),

$$\sigma_\mu^\dagger = (-1)^k A\sigma_\mu A^{-1}$$

By rescaling of B , the map A can be made Hermitian. The corresponding Hermitian form $s \mapsto A(s, s)$ is definite if and only if k or $l = 0$; otherwise, it is neutral.

For $m = k + l$ even, the representations γ^\dagger and γ are equivalent and one can define a Hermitian isomorphism $A:S \rightarrow \bar{S}^*$ so that

$$\gamma_\mu^\dagger = A\gamma_\mu A^{-1} \tag{23}$$

The isomorphism $A' = A\Gamma$ intertwines the representations γ^\dagger and $\gamma \circ \alpha$; it can also be made Hermitian by rescaling. The Hermitian form $A(s, s)$ is definite for $k = 0$ and $A'(s, s)$ is definite for $l = 0$; otherwise, these forms are neutral. For example, in the familiar representation [20], one has $A = \gamma_4$, a neutral form.

For $p = 0, 1, \dots, m = 2n$, two spinors s and $t \in S$ define the p -vector with components

$$A_{\mu_1 \dots \mu_p}(s, t) = \langle \bar{s}, A\gamma_{\mu_1} \dots \gamma_{\mu_p} t \rangle \tag{24}$$

where the indices are as in [19]. The Hermiticity of A and [23] imply

$$\overline{A_{\mu_1 \dots \mu_p}(s, t)} = (-1)^{(1/2)p(p-1)} A_{\mu_1 \dots \mu_p}(t, s)$$

In view of $\Gamma^\dagger = (-1)^k A\Gamma A^{-1}$, the map A defines, for k even, a nondegenerate Hermitian scalar product on the spaces S_\pm whereas $A(s, t) = 0$ if s and t are Weyl spinors of opposite chiralities. For k odd, A changes the chirality.

The Radon–Hurwitz Numbers

Proposition (M) *For every integer $m > 0$, the algebra $\mathcal{C}\ell_{m,0}$ has an irreducible real representation*

ρ of dimension $2^{\chi(m)}$, where $\chi(m)$ is the m th Radon–Hurwitz number given by

$$\begin{matrix} m = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \chi(m) = & 1 & 2 & 2 & 3 & 3 & 3 & 3 & 4 \end{matrix}$$

and $\chi(m + 8) = \chi(m) + 4$. The matrices $\rho_\mu \in \mathbb{R}(2^{\chi(m)})$, $\mu = 1, \dots, m$, defining these representations satisfy

$$\rho_\mu \rho_\nu + \rho_\nu \rho_\mu = -2\delta_{\mu\nu} I$$

and can be chosen so as to be antisymmetric. In all dimensions other than $m \equiv 3 \pmod 4$ the representations are faithful.

For $m \equiv 2$ and $4 \pmod 8$ (resp., $m \equiv 1, 3$, and $5 \pmod 8$) the representations ρ are the realifications of the corresponding Dirac (resp., Pauli) representations. In dimensions $m \equiv 0$ and $6 \pmod 8$ (resp., $m \equiv 7 \pmod 8$) the Dirac (resp., Pauli) representations themselves are real.

Inductive Construction of Representations

An inductive construction of the Pauli representations

$$\sigma : \mathcal{C}\ell_{n-1,n} \rightarrow \mathbb{R}(2^{n-1}), \quad n = 1, 2, \dots$$

and of the Dirac representations

$$\gamma : \mathcal{C}\ell_{n,n} \rightarrow \mathbb{R}(2^n), \quad n = 1, 2, \dots$$

is as follows.

1. In dimension 1, put $\sigma_1 = 1$.
2. Given $\sigma_\mu \in \mathbb{R}(2^{n-1})$, $\mu = 1, \dots, 2n - 1$, define

$$\gamma_\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \sigma_\mu & 0 \end{pmatrix} \quad \text{for } \mu = 1, \dots, 2n - 1$$

and

$$\gamma_{2n} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

3. Given $\gamma_\mu \in \mathbb{R}(2^n)$, $\mu = 1, \dots, 2n$, define $\sigma_\mu = \gamma_\mu$ for $\mu = 1, \dots, 2n$, and $\sigma_{2n+1} = \gamma_1 \dots \gamma_{2n}$.

All entries of these matrices are either 0, 1, or -1 ; therefore, they can be used to construct representations of Clifford algebras of orthogonal spaces over any commutative field of characteristic $\neq 2$.

By induction, one has $\sigma_\mu^* = (-1)^{\mu+1} \sigma_\mu$. Therefore, the isomorphisms appearing in [18] are $B = \gamma_2 \gamma_4 \dots \gamma_{2n}$ for both $m = 2n$ and $2n + 1$.

By multiplying some of the matrices σ_μ or γ_μ by the imaginary unit, one obtains complex representations of the Clifford algebras associated with the quadratic

forms of other signatures. For example, in dimension 3, $(\sigma_1, i\sigma_2, \sigma_3)$ are the Pauli matrices. In dimension 4, multiplying γ_2 by i one obtains the Dirac matrices for g of signature $(1, 3)$, in the “chiral representation”:

$$\begin{aligned} \gamma_1 &= \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, & \gamma_2 &= \begin{pmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{pmatrix} \\ \gamma_3 &= \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, & \gamma_4 &= \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \end{aligned} \tag{25}$$

To obtain the real Majorana representation one uses the following fact:

Proposition (N) *If the matrix $C \in \mathbb{R}(2^n)$ is such that $C^2 = I$ and [21] holds, then the matrices $(I + iC)\gamma_\mu(I + iC)^{-1}$, $\mu = 1, \dots, 2n$, {it are real}.*

For the matrices [25], one can take $C = \gamma_1\gamma_3\gamma_4$ to obtain

$$\begin{aligned} \gamma'_1 &= \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, & \gamma'_2 &= \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \\ \gamma'_3 &= \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, & \gamma'_4 &= \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \end{aligned}$$

The real representations described in proposition (M) can be obtained by the following direct inductive construction. Consider the following seven real anti-symmetric and anticommuting 8×8 matrices:

$$\begin{aligned} \rho_1 &= \sigma_z \otimes I \otimes \varepsilon, & \rho_2 &= \sigma_z \otimes \varepsilon \otimes \sigma_x \\ \rho_3 &= \sigma_x \otimes \varepsilon \otimes \sigma_z, & \rho_4 &= \sigma_x \otimes \varepsilon \otimes I \\ \rho_5 &= \sigma_x \otimes \sigma_x \otimes \varepsilon, & \rho_6 &= \sigma_x \otimes \sigma_z \otimes \varepsilon \\ \rho_7 &= \varepsilon \otimes I \otimes I \end{aligned} \tag{26}$$

For $m = 4, 5, 6,$ and 7 the matrices ρ_1, \dots, ρ_m generate the representations of $Cl_{m,0}$ in \mathbb{R}^8 . The eight matrices $\theta_\mu = \sigma_x \otimes \rho_\mu, \mu = 1, \dots, 7$, and $\theta_8 = \varepsilon \otimes I \otimes I$ give the required representation of $Cl_{8,0}$ in \mathbb{R}^{16} . By dropping the first factor in ρ_1, ρ_2, ρ_3 , one obtains the matrices generating a representation of $Cl_{3,0}$ in \mathbb{R}^4 , etc. The symmetric matrix $\Theta = \theta_1 \cdots \theta_8 = \sigma_z \otimes I \otimes I \otimes I$ anticommutes with all the θ s and $\Theta^2 = I$. If the matrices $\rho_\mu \in \mathbb{R}(2^{\chi(m)})$ correspond to a representation of $Cl_{m,0}$, then the $m + 8$ matrices $\Theta \otimes \rho_1, \dots, \Theta \otimes \rho_m, \theta_1 \otimes I, \dots, \theta_8 \otimes I$ generate the required representation of $Cl_{m+8,0}$.

Vector Fields on Spheres and Division Algebras

It is known that even-dimensional spheres have no nowhere-vanishing tangent vector fields. All such

fields on odd-dimensional spheres can be constructed with the help of the representation ρ described in proposition (M). Given a positive even integer N , let m be the largest integer such that $N = 2^{\chi(m)}p$, where p is an odd integer. Consider the unit sphere $S_{N-1} = \{x \in \mathbb{R}^N \mid \|x\| = 1\}$ of dimension $N - 1$. For $v \in \mathbb{R}^m$, put $\rho'(v) = \rho(v) \otimes I$, where $I \in \mathbb{R}(p)$ is the unit matrix. Since $\rho(v)$ is antisymmetric, so is the matrix $\rho'(v) \in \mathbb{R}(N)$. Therefore, for every $x \in S_{N-1}$, the vector $\rho'(v)x$ is orthogonal to x . The map $x \mapsto \rho'(v)x$ defines a vector field on S_{N-1} that vanishes nowhere unless $v = 0$: the $(N-1)$ -sphere admits a set of m tangent vector fields which are linearly independent at every point. Using methods of algebraic topology, it has been shown that this method gives the maximum number of linearly independent tangent vector fields on spheres.

If $m = 1, 3,$ or 7 , then $m + 1 = 2^{\chi(m)}$ and, for these values of m , the sphere S_m is parallelizable. Moreover, one can then introduce in \mathbb{R}^{m+1} the structure of an algebra A_m as follows. Put $\rho_0 = I$. If $e_0 \in \mathbb{R}^{m+1}$ is a unit vector and $e_\mu = \rho_\mu(e_0)$, then (e_0, e_1, \dots, e_m) is an orthonormal frame in \mathbb{R}^{m+1} . The product of $x = \sum_{\mu=0}^m x_\mu e_\mu$ and $y = \sum_{\mu=0}^m y_\mu e_\mu$ is defined to be

$$x \cdot y = \sum_{\mu, \nu=0}^m x_\mu y_\nu \rho_\mu(e_\nu)$$

so that e_0 is the unit element for this product. Defining $\text{Re}x = x_0 e_0, \text{Im}x = x - \text{Re}x, \bar{x} = \text{Re}x - \text{Im}x$, one has $\bar{x} \cdot x = e_0 \|x\|^2$ and $\bar{x} \cdot (x \cdot y) = (\bar{x} \cdot x) \cdot y$, so that $x \cdot y = 0$ implies $x = 0$ or $y = 0$: A_m is a normed algebra without zero divisors. The algebras A_1 and A_3 are isomorphic to \mathbb{C} and \mathbb{H} , respectively, and A_7 is, by definition, the algebra \mathbb{O} of octonions discovered by Graves and Cayley. The algebra \mathbb{O} is nonassociative; its multiplication table is obtained with the help of [26].

Spinor Groups

Let (V, g) be a quadratic space over K . If $u \in V$ is not null, then it is invertible as an element of $Cl(V, g)$ and the map $v \mapsto -uvu^{-1}$ is a reflection in the hyperplane orthogonal to u . The orthogonal group $O(V, g) = O(V, -g) = \{R \in GL(V) \mid R^* \circ g \circ R = g\}$ is generated by the set of all such reflections. A spinor group G is a subset of $Cl(V, g)$ that is a group with respect to multiplication induced by the product in the algebra, with a homomorphism $\rho: G \rightarrow GL(V)$ whose image contains the connected component $SO^0(V, g)$ of the group of rotations of (V, g) . In the case of real quadratic spaces, one considers also spinor groups that are subsets of $\mathbb{C} \otimes Cl(V, g)$ with similar properties. By restriction, every

representation of $Cl(V, g)$ or $\mathbb{C} \otimes Cl(V, g)$ gives spinor representations of the spinor groups it contains.

Pin Groups

It is convenient to define a unit vector $v \in V \subset Cl(V, g)$ to be such that $v^2 = 1$ for V complex and $v^2 = 1$ or -1 for V real. The group $Pin(V, g)$ is defined as the subgroup of $Cpin(V, g)$ consisting of products of all finite sequences of unit vectors. Defining now the twisted adjoint representation Ad by $Ad(a)v = \alpha(a)va^{-1}$, one obtains the exact sequence

$$1 \rightarrow \mathbb{Z}_2 \rightarrow Pin(V, g) \xrightarrow{\widetilde{Ad}} O(V, g) \rightarrow 1 \quad [27]$$

If $\dim V$ is even, then the adjoint representation $Ad(a)v = av a^{-1}$ also yields an exact sequence like [27]; if it is odd, then the image of Ad is $SO(V, g)$ and the kernel is the four-element group $\{1, -1, \eta, -\eta\}$.

Given an orthonormal frame (e_μ) in (V, g) and $a \in Pin(V, g)$, one defines the orthogonal matrix $R(a) = (R_\mu^\nu(a))$ by

$$\widetilde{Ad}(a)e_\mu = e_\nu R_\mu^\nu(a) \quad [28]$$

If (V, g) is complex, then the algebras $Cl(V, g)$ and $Cl(V, -g)$ are isomorphic; this induces an isomorphism of the groups $Pin(V, g)$ and $Pin(V, -g)$. If $V = \mathbb{C}^m$, then this group is denoted by $Pin_m(\mathbb{C})$. If $V = \mathbb{R}^{k+l}$ and g of signature (k, l) , then one writes $Pin(V, g) = Pin_{k,l}$. A similar notation is used for the groups $spin$, see below.

Spin Groups

The spin group $Spin(V, g) = Pin(V, g) \cap Cl^0(V, g)$ is generated by products of all sequences of an even number of unit vectors. Since the algebras $Cl^0(V, g)$ and $Cl^0(V, -g)$ are isomorphic, so are the groups $Spin(V, g)$ and $Spin(V, -g)$. Since $\alpha(a) = a$ for $a \in Spin(V, g)$, the twisted adjoint representation reduces to the adjoint representation and yields the exact sequence

$$1 \rightarrow \mathbb{Z}_2 \rightarrow Spin(V, g) \xrightarrow{Ad} SO(V, g) \rightarrow 1 \quad [29]$$

For $V = \mathbb{C}^m$, the spin group is denoted by $Spin_m(\mathbb{C})$. Since $Spin_m(\mathbb{C}) \subset G_1(\beta)$, the bilinear form B is invariant with respect to the action of this group.

Spin⁰ Groups

The connected component $Spin^0(V, g)$ of the group $Spin(V, g)$ coincides with $Spin(V, g)$ if either the quadratic space (V, g) is complex or real and $kl = 0$. In signature (k, l) , the connect group $Spin_{k,l}^0$ is generated in $Cl_{k,l}^0$ by all products of the form

$u_1 \dots u_{2p} v_1 \dots v_{2q}$ such that $u_i^2 = -1$ and $v_j^2 = 1$. The connected groups $Spin_{m,0}$ and $Spin_{0,m}$ are isomorphic and denoted by $Spin_m^0$. Since $Spin_{k,l}^0 \subset G_1(\beta)$, the Hermitian form A and the bilinear form B are invariant with respect to the action of this group. Moreover, for $k+l$ even, from [24] and [28] there follows the transformation law of multivectors formed from pairs of spinors,

$$A_{\mu_1 \dots \mu_p}(\gamma(a)s, \gamma(a)t) = A_{\nu_1 \dots \nu_p}(s, t) R_{\mu_1}^{\nu_1}(a^{-1}) \dots R_{\mu_p}^{\nu_p}(a^{-1})$$

Consider $Spin^0(V, g)$ and assume that either V is complex of dimension ≥ 2 or real with k or $l \geq 2$. Then there are two unit orthogonal vectors $e_1, e_2 \in V$ such that $(e_1, e_2)^2 = -1$. The vector $u(t) = e_1 \cos t + e_2 \sin t$ is obtained from e_1 by rotation in the plane $\text{span}\{e_1, e_2\}$ by the angle $t \in \mathbb{R}$. The curve $t \mapsto e_1 u(t)$, $0 \leq t \leq \pi$, connects the elements 1 and -1 of $Spin^0(V, g)$. Its image in $SO^0(V, g)$, that is, the curve $t \mapsto Ad(e_1 u(t))$, $0 \leq t \leq \pi$, is closed: $Ad(1) = Ad(-1)$. This fact is often expressed by saying that “a spinor undergoing a rotation by 2π changes sign.” There is no homomorphism – not even a continuous map $-f : SO^0(V, g) \rightarrow Spin^0(V, g)$ such that $Ad \circ f = \text{id}$.

Spin^c Groups

For the purposes of physics, to describe charged fermions, and in the theory of the Seiberg–Witten invariants, one needs the $Spin^c$ groups that are spinorial extensions of the real orthogonal groups by the group U_1 of “phase factors.” Assume V to be real and g of signature (k, l) so that the sequence [29] can be written as

$$1 \rightarrow \mathbb{Z}_2 \rightarrow Spin_{k,l} \rightarrow SO_{k,l} \rightarrow 1$$

Define the action of $\mathbb{Z}_2 = \{1, -1\}$ in $Spin_{k,l} \times U_1$ so that $(-1)(a, z) = (-a, -z)$. The quotient $(Spin_{k,l} \times U_1)/\mathbb{Z}_2 = Spin_{k,l}^c$ yields the extensions

$$1 \rightarrow U_1 \rightarrow Spin_{k,l}^c \rightarrow SO_{k,l} \rightarrow 1$$

and

$$1 \rightarrow Spin_{k,l} \rightarrow Spin_{k,l}^c \rightarrow U_1 \rightarrow 1$$

For example, $Spin_3 = SU_2$ and $Spin_3^c = U_2$.

Spin Groups in Dimensions ≤ 6

The connected components of spin groups associated with orthogonal spaces of dimension ≤ 6 are isomorphic to classical groups. They can be explicitly described starting from the following observations.

Consider the four-dimensional vector space (of twistors) T over K , with a volume element $\text{vol} \in \wedge^4 T$. The six-dimensional vector space $V = \wedge^2 T$ has a scalar product g defined by $g(u, v)\text{vol} = 2u \wedge v$ for $u, v \in V$. The quadratic form $g(u, u)$ is the Pfaffian, $\text{Pf}(u)$. If $u \in V$ is represented by the corresponding isomorphism $T^* \rightarrow T$ and $a \in \text{End } T$, then $\text{Pf}(aua^*) = \det a \text{Pf}(u)$. The last formula shows $\text{Spin}^0(V, g) = \text{SL}(T)$, so that $\text{Spin}_6(\mathbb{C}) = \text{SL}_4(\mathbb{C})$. For $K = \mathbb{R}$, the Pfaffian is of signature $(3, 3)$, so that $\text{Spin}_{3,3}^0 = \text{SL}_4(\mathbb{R})$. A non-null vector $v \in V$ defines a symplectic form on T^* . The five-dimensional vector space $v^\perp \subset V$ is invariant with respect to the symplectic group $\text{Sp}(T^*, u) = \text{Spin}^0(v^\perp, \text{Pf}|_{v^\perp})$. This shows that $\text{Spin}_5(\mathbb{C}) = \text{Sp}_4(\mathbb{C})$ and $\text{Spin}_{2,3}^0 = \text{Sp}_4(\mathbb{R})$. Spin groups for other signatures in real dimensions 6 and 5 are obtained by considering appropriate real subspaces of \mathbb{C}^6 and \mathbb{C}^5 , respectively. For example, [6] is used to show that $\text{Spin}_{1,5}^0 = \text{SL}_2(\mathbb{H})$.

Spin groups in dimensions 4 and lower are similarly obtained from the observation that \det is a quadratic form on the four-dimensional space $K(2)$ and $\text{Cl}^0(K(2), \det) = K(2) \oplus K(2)$.

Several spin groups are listed below.

The complex spin groups

$$\begin{aligned}\text{Spin}_2(\mathbb{C}) &= \mathbb{C}^\times, & \text{Spin}_3(\mathbb{C}) &= \text{SL}_2(\mathbb{C}) \\ \text{Spin}_4(\mathbb{C}) &= \text{SL}_2(\mathbb{C}) \times \text{SL}_2(\mathbb{C}) \\ \text{Spin}_5(\mathbb{C}) &= \text{Sp}_4(\mathbb{C}) \\ \text{Spin}_6(\mathbb{C}) &= \text{SL}_4(\mathbb{C})\end{aligned}$$

The real, compact spin groups

$$\begin{aligned}\text{Spin}_2 &= \text{U}_1, & \text{Spin}_3 &= \text{SU}_2 \\ \text{Spin}_4 &= \text{SU}_2 \times \text{SU}_2, & \text{Spin}_5 &= \text{Sp}_2(\mathbb{H}) \\ \text{Spin}_6 &= \text{SU}_4\end{aligned}$$

The groups $\text{Spin}_{k,l}^0$ for $1 \leq k \leq l$ and $k + l \leq 6$

$$\begin{aligned}\text{Spin}_{1,1}^0 &= \mathbb{R}^\times, & \text{Spin}_{1,2}^0 &= \text{SL}_2(\mathbb{R}) \\ \text{Spin}_{1,3}^0 &= \text{SL}_2(\mathbb{C}) \\ \text{Spin}_{2,2}^0 &= \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R}) \\ \text{Spin}_{1,4}^0 &= \text{Sp}_{1,1}(\mathbb{H}) \\ \text{Spin}_{2,3}^0 &= \text{Sp}_4(\mathbb{R}), & \text{Spin}_{1,5}^0 &= \text{SL}_2(\mathbb{H}) \\ \text{Spin}_{2,4}^0 &= \text{SU}_{2,2} \\ \text{Spin}_{3,3}^0 &= \text{SL}_4(\mathbb{R})\end{aligned}$$

See also: Dirac Operator and Dirac Field; Index Theorems; Relativistic Wave Equations Including Higher Spin Fields; Spinors and Spin Coefficients; Twistors.

Further Reading

- Adams JF (1981) Spin (8), triality, F_4 and all that. In: Hawking SW and Roček M (eds.) *Superspace and Supergravity*. Cambridge: Cambridge University Press.
- Atiyah MF, Bott R, and Shapiro A (1964) Clifford modules. *Topology* 3(suppl. 1): 3–38.
- Baez JC (2002) The octonions. *Bulletin of the American Mathematical Society* 39: 145–205.
- Brauer R and Weyl H (1935) Spinors in n dimensions. *American Journal of Mathematics* 57: 425–449.
- Budinich P and Trautman A (1988) *The Spinorial Chessboard*, Trieste Notes in Physics. Berlin: Springer.
- Cartan É (1938) *Théorie des spineurs*. Actualités Scientifiques et Industrielles, No. 643 et 701. Paris: Hermann (English transl.: *The Theory of Spinors*. Paris: Hermann, 1966).
- Chevalley C (1954) *The Algebraic Theory of Spinors*. New York: Columbia University Press.
- Clifford WK (1878) Applications of Grassmann's extensive algebra. *American Journal of Mathematics* 1: 350–358.
- Clifford WK (1882) On the classification of geometric algebras. In: Tucker R (ed.) *Mathematical Papers by William Kingdon Clifford*, pp. 397–401. London: Macmillan.
- Dirac PAM (1928) The quantum theory of the electron. *Proceedings of the Royal Society of London A* 117: 610–624.
- Eckmann B (1942) Gruppentheoretische Beweis des Satzes von Hurwitz–Radon über die Komposition quadratischer Formen. *Commentarii Mathematici Helvetici* 15: 358–366.
- Karoubi M (1968) Algèbres de Clifford et K -théorie. *Annales Scientifiques de l'École Normale Supérieure 4ème sér* 1: 161–270.
- Lipschitz RO (1886) *Untersuchungen über die Summen von Quadraten*. Berlin: Max Cohen und Sohn.
- Lounesto P (2001) *Clifford Algebras and Spinors*, 2nd edn. London Math. Soc. Lecture Note Series, vol. 286. Cambridge: Cambridge University Press.
- Pauli W (1927) Zur Quantenmechanik des magnetischen Elektrons. *Z. Physik* 43: 601–623.
- Penrose R and MacCallum MAH (1973) Twistor theory: an approach to the quantisation of fields and space-time. *Physics Report* 6C(4): 241–316.
- Porteous IR (1995) *Clifford Algebras and the Classical Groups*, Cambridge Studies in Advanced Mathematics, vol. 50. Cambridge: Cambridge University Press.
- Postnikov MM (1986) *Lie groups and Lie algebras*. Mir: Moscow.
- Sudbery A (1987) Division algebras (pseudo)orthogonal groups and spinors. *Journal of Physics A* 17: 939–955.
- Trautman A (1997) Clifford and the “square root” ideas. *Contemporary Mathematics* 203: 3–24.
- Trautman A and Trautman K (1994) Generalized pure spinors. *Journal of Geometry and Physics* 15: 1–22.
- Wall CTC (1963) Graded Brauer groups. *Journal für die Reine und Angewandte Mathematik* 213: 187–199.

Cluster Expansion

R Kotecký, Charles University, Prague, Czech Republic, and the University of Warwick, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The method of cluster expansions in statistical physics provides a systematic way of computing power series for thermodynamic potentials (logarithms of partition functions) as well as correlations. It originated from the works of Mayer and others devoted to expansions for dilute gas.

Mayer Expansion

Consider a system of interacting particles with Hamiltonian

$$\begin{aligned} H_N(\mathbf{p}_1, \dots, \mathbf{p}_N, \mathbf{r}_1, \dots, \mathbf{r}_N) \\ = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_{i,j=1}^N \Phi(\mathbf{r}_i - \mathbf{r}_j) \end{aligned} \quad [1]$$

where Φ is a *stable* and *regular* pair potential. Namely, we assume that there exists $B \geq 0$ such that

$$\sum_{i,j=1}^N \Phi(\mathbf{r}_i - \mathbf{r}_j) \geq -BN \quad [2]$$

for all $N=2, 3, \dots$ and all $(\mathbf{r}_1, \dots, \mathbf{r}_N) \in \mathbb{R}^{3N}$, and that

$$C(\beta) = \int |e^{-\beta\Phi(\mathbf{r})} - 1| d^3\mathbf{r} < \infty \quad [3]$$

for some $\beta > 0$ (and hence all $\beta > 0$). Basic thermodynamic quantities are given in terms of the *grand-canonical partition function*

$$\begin{aligned} Z(\beta, \lambda, V) &= \sum_{N=0}^{\infty} \frac{z^N}{N!} \int_{\mathbb{R}^{3N} \times V^N} e^{-\beta H_N} \frac{\prod d^3\mathbf{p}_i \prod d^3\mathbf{r}_i}{h^{3N}} \\ &= \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} \int_{V^N} e^{-\beta \sum_{i,j} \Phi(\mathbf{r}_i - \mathbf{r}_j)} \prod d^3\mathbf{r}_i \end{aligned} \quad [4]$$

In the second expression we absorbed the factor resulting from the integration over impulses into (configurational) activity $\lambda = (2\pi m / \beta h^2)^{3/2} z$. In particular, the *pressure* p and the *density* ρ are defined by the thermodynamic limits (with $V \rightarrow \infty$ in the sense of Van Hove)

$$p(\beta, \lambda) = \frac{1}{\beta} \lim_{V \rightarrow \infty} \frac{1}{|V|} \log Z(\beta, \lambda, V) \quad [5]$$

and

$$\rho(\beta, \lambda) = \lim_{V \rightarrow \infty} \frac{1}{|V|} \lambda \frac{\partial}{\partial \lambda} \log Z(\beta, \lambda, V) \quad [6]$$

Mayer series are the expansions of p and ρ in powers of λ :

$$\beta p(\beta, \lambda) = \sum_{n=1}^{\infty} b_n \lambda^n \quad [7]$$

and

$$\rho(\beta, \lambda) = \sum_{n=1}^{\infty} n b_n \lambda^n \quad [8]$$

Mayer's idea for a systematic computation of coefficients b_n was based on a reformulation of partition function $Z(\beta, \lambda, V)$ in terms of *cluster integrals*. Introducing the function

$$f(\mathbf{r}) = e^{-\beta\Phi(\mathbf{r})} - 1 \quad [9]$$

and using $\mathcal{G}[N]$ to denote the set of all graphs on N vertices $\{1, \dots, N\}$, we get

$$\begin{aligned} Z(\beta, \lambda, V) &= \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} \int_{V^N} \prod_{i,j=1}^N (1 + f(\mathbf{r}_i - \mathbf{r}_j)) \prod d^3\mathbf{r}_i \\ &= \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} \sum_{g \in \mathcal{G}[N]} w(g) \end{aligned} \quad [10]$$

where

$$w(g) = \int_{V^N} \prod_{\{i,j\} \in g} f(\mathbf{r}_i - \mathbf{r}_j) \prod d^3\mathbf{r}_i \quad [11]$$

Observing that the weight w is multiplicative in connected components (clusters) g_1, \dots, g_k of the graph g ,

$$w(g) = \prod_{\ell=1}^k w(g_\ell) \quad [12]$$

we can rewrite

$$Z(\beta, \lambda, V) = \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} \sum_{\{g_i\}} \prod w(g) \quad [13]$$

with the sum running over all disjoint collections $\{g_i\}$ of connected graphs with vertices in $\{1, \dots, N\}$. A straightforward exponential expansion can be used to show that, at least in the sense of formal power series,

$$\log Z(\beta, \lambda, V) = \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \sum_{g \in \mathcal{C}[n]} w(g) \quad [14]$$

where $\mathcal{C}[n]$ is the set of all connected graphs on n vertices. Using $b_n^{(V)}$ to denote the coefficients

$$b_n^{(V)} = \frac{1}{|V|} \frac{1}{n!} \sum_{g \in \mathcal{C}[n]} w(g) \tag{15}$$

and observing that the limits $\lim_{V \rightarrow \infty} (1/|V|)w(g)$ of cluster integrals exist, we get $b_n = \lim_{V \rightarrow \infty} b_n^{(V)}$. The convergence of Mayer series can be controlled directly by combinatorial estimates on the coefficients $b_n^{(V)}$. As a result, the diameter of convergence of the series [7] and [8] can be proved to be at least $(C(\beta)e^{2\beta B+1})^{-1}$. A less direct proof is based on an employment of linear integral Kirkwood–Salsburg equations in a suitable Banach space of correlation functions.

Similar combinatorial methods are available also for evaluation of coefficients of the *virial expansion* of pressure in powers of gas density,

$$\beta p(\beta, \rho) = \sum_{n=1}^{\infty} \beta_n \rho^n \tag{16}$$

obtained by inverting [8] (notice that $b_1 = 1$) and inserting it into [7]. One is getting $\beta_n = \lim_{V \rightarrow \infty} \beta_n^{(V)}$ with

$$\beta_n^{(V)} = \frac{1}{|V|} \frac{1}{n!} \sum_{g \in \mathcal{B}[n]} w(g) \tag{17}$$

where $\mathcal{B}[n] \subset \mathcal{C}[n]$ is the set of all 2-connected graphs on $\{1, \dots, n\}$; namely, those graphs that cannot be split into disjoint subgraphs by erasing one vertex (and all adjacent edges). The diameter of convergence of the virial expansion turns out to be no less than $(C(\beta)e^{(e^{2\beta B} + 1)})^{-1}$.

Abstract Polymer Models

An application of the ideas of Mayer expansions to lattice models is based on a reformulation of the partition function in terms of a *polymer model*, a formulation akin to [13] above. Namely, the partition function is rewritten as a sum over collections of pairwise compatible geometric objects – polymers. Most often, the compatibility means simply their disjointness.

While the reformulation of “physical partition function” in terms of a polymer model (including the definition of compatibility) depends on particularities of a given lattice model and on the considered region of parameters – high-temperature, low-temperature, large external fields, etc. – the essence and results of cluster expansion may be conveniently formulated in terms of an *abstract polymer model*.

Let $G = (V, E)$ be any (possibly infinite) countable graph and suppose that a map $w: V \rightarrow \mathbb{C}$ is given.

Vertices $v \in V$ are called *abstract polymers*, with two abstract polymers connected by an edge in the graph G called *incompatible*. We shall refer to $w(v)$ as to the *weight* of the abstract polymer v . For any finite $W \subset V$, we consider the induced subgraph $G[W]$ of G spanned by W and define

$$Z_W(w) = \sum_{I \subset W} \prod_{v \in I} w(v) \tag{18}$$

Here the sum runs over all collections I of compatible abstract polymers – or, in other words, the sum is over all *independent sets* I of vertices in W (no two vertices in I are connected by an edge).

The partition function $Z_W(w)$ is an entire function in $w = \{w(v)\}_{v \in W} \in \mathbb{C}^{|W|}$ and $Z_W(0) = 1$. Hence, it is nonvanishing in some neighborhood of the origin $w = 0$ and its logarithm is, on this neighbourhood, an analytic function yielding a convergent Taylor series

$$\log Z_W(w) = \sum_{X \in \mathcal{X}(W)} a_W(X) w^X \tag{19}$$

Here, $\mathcal{X}(W)$ is the set of all multi-indices $X: W \rightarrow \{0, 1, \dots\}$ and $w^X = \prod_v w(v)^{X(v)}$. Inspecting the formula for $a_W(X)$ in terms of corresponding derivatives of $\log Z_W(w)$, it is easy to show that the Taylor coefficients $a_W(X)$ actually do not depend on $W: a_W(X) = a_{\text{supp } X}(X)$, where $\text{supp } X = \{v \in V: X(v) \neq 0\}$. As a result, one is getting the existence of coefficients $a(X)$ such that

$$\log Z_W(w) = \sum_{X \in \mathcal{X}(W)} a(X) w^X \tag{20}$$

for every finite $W \subset V$.

The coefficients $a(X)$ can be obtained explicitly. One can pass from [18] to [20] in a similar way as passing from [10] to [13]. The starting point is to replace the restriction to compatible collections of abstract polymers in the sum [18] by the factor $\prod_{v, v' \in W} (1 + F(v, v'))$ with

$$F(v, v') = \begin{cases} 0 & \text{if } v \text{ and } v' \text{ are compatible} \\ -1 & \text{otherwise } (v \text{ and } v' \\ & \text{connected by an edge from } G) \end{cases} \tag{21}$$

and to expand the product afterwards. The resulting formula is

$$a(X) = (X!)^{-1} \sum_{H \subset G(X)} (-1)^{|E(H)|} \tag{22}$$

Here, $G(X)$ is the graph with $|X| = \sum |X(v)|$ vertices induced from $G[\text{supp } X]$ by replacing each of its vertices v by the complete graph on $|X(v)|$ vertices and $X!$ is the multifactorial $X! = \prod_{v \in \text{supp } X} |X(v)|!$. The sum is over all connected subgraphs $H \subset G(X)$ spanned by the set of vertices of $G(X)$ and $|E(H)|$ is the number of edges of the graph H .

A useful property of the coefficients $a(X)$ is their *alternating sign*,

$$(-1)^{|X|+1} a(X) \geq 0 \tag{23}$$

More important than an explicit form of the coefficients $a(X)$ are the convergence criteria for the series [20]. One way to proceed is to find direct combinatorial bounds on the coefficients as expressed by [22]. While doing so, one has to take into account the cancelations arising in view of the presence of terms of opposite signs in [22]. Indeed, disregarding them would lead to a failure since, as it is easy to verify, the number of connected graphs on $|X|$ vertices is bounded from below by $2^{(|X|-1)(|X|-2)/2}$. An alternative approach is to prove the convergence of [20] on polydisks $\mathcal{D}_{W,R} = \{w : |w(v)| \leq R(v) \text{ for } v \in W\}$ by induction in $|W|$, once a proper condition on the set of radii $R = \{R(v); v \in V\}$ is formulated. The most natural for the inductive proof (leading in the same time to the strongest claim) turns out to be the Dobrushin condition:

There exists a function $r : V \rightarrow [0, 1)$ such that, for each $v \in V$

$$R(v) \leq r(v) \prod_{v' \in \mathcal{N}(v)} (1 - r(v')) \tag{24}$$

Here $\mathcal{N}(v)$ is the set of vertices $v' \in V$ adjacent in graph G to the vertex v .

Using \mathcal{X} to denote the set of all multi-indices $X : V \rightarrow \{0, 1, \dots\}$ with finite $|X| = \sum |X(v)|$ and saying that $X \in \mathcal{X}$ is a cluster if the graph $G(\text{supp } X)$ is connected, we can summarize the cluster expansion claim for an abstract polymer model in the following way:

Theorem (Cluster expansion). *There exists a function $a : \mathcal{X} \rightarrow \mathbb{R}$ that is nonvanishing only on clusters, so that for any sequence of diameters R satisfying the condition [24] with a sequence $\{r(v)\}$, the following holds true:*

- (i) *For every finite $W \subset V$, and any contour weight $w \in \mathcal{D}_{W,R}$, one has $Z_W(w) \neq 0$ and*

$$\log Z_W(w) = \sum_{X \in \mathcal{X}(W)} a(X) w^X$$

- (ii) $\sum_{X \in \mathcal{X} : \text{supp } X \ni v} |a(X)| |w|^X \leq -\log(1 - r(v))$.

Notice that, we have got not only an absolute convergence of the Taylor series of $\log Z_W$ in the closed polydisk $\mathcal{D}_{W,R}$, but also the bound (ii) (uniform in W) on the sum over all terms containing a fixed vertex v . Such a bound turns out to be very useful in applications of cluster expansions. It yields, eventually, bounds on various error terms, avoiding a need of an explicit evaluation of the number of clusters of “given size.”

The restriction to compatible collections of polymers can be actually relaxed. Namely, replacing [25] by

$$Z_W(w) = \sum_{W' \subset W} \prod_{v \in W'} w(v) \prod_{v, v' \in W'} U(v, v') \tag{25}$$

with $U(v, v') \in [0, 1]$ (soft repulsive interaction), and the condition [24] by

$$R(v) \leq r(v) \prod_{v' \neq v} \frac{1 - r(v')}{1 - U(v, v')r(v')} \tag{26}$$

one can prove that the partition function $Z_W(w)$ does not vanish on the polydisk $\mathcal{D}_{W,R}$ implying thus that the power series of $\log Z_W(w)$ converges absolutely on $\mathcal{D}_{W,R}$.

Polymers that arise in typical applications are geometric objects endowed with a “support” in the considered lattice, say $\mathbb{Z}^d, d \geq 1$, and their weights satisfy the condition of translation invariance. Cluster expansions then yield an explicit power series for the *pressure* (resp. free energy) in the thermodynamic limit as well as its finite-volume approximation.

To formulate it for an abstract polymer model, we assume that for each $x \in \mathbb{Z}^d$, an isomorphism $\tau_x : G \rightarrow G$ is given and that with each abstract polymer $v \in V$ a finite set $\Lambda(v) \subset \mathbb{Z}^d$ is associated so that $\Lambda(\tau_x(v)) = \Lambda(v) + x$ for every $v \in V$ and every $x \in \mathbb{Z}^d$. For any finite $W \subset V$ and any multi-index X , let $\Lambda(W) = \cup_{v \in W} \Lambda(v)$ and $\Lambda(X) = \Lambda(\text{supp}(X))$. On the other hand, for any finite $\Lambda \subset \mathbb{Z}^d$, let $W(\Lambda) = \{v \in V : \Lambda(v) \subset \Lambda\}$. Assuming also that the weight $w : V \rightarrow \mathbb{C}$ is translation invariant – that is, $w(v) = w(\tau_x(v))$ for every $v \in V$ and every $x \in \mathbb{Z}^d$ – we get an explicit expression for the “pressure” of abstract polymer model in the thermodynamic limit

$$p = \lim_{\Lambda \rightarrow \infty} \frac{1}{|\Lambda|} \log Z_{W(\Lambda)}(w) = \sum_{X : \Lambda(X) \ni 0} \frac{a(X) w^X}{|\Lambda(X)|} \tag{27}$$

In addition, the finite-volume approximation can be explicitly evaluated, yielding

$$\begin{aligned} \log Z_{W(\Lambda)}(w) &= p|\Lambda| + \sum_{X : \Lambda(X) \cap \Lambda^c \neq \emptyset} a(X) w^X \frac{|\Lambda(X) \cap \Lambda|}{|\Lambda(X)|} \end{aligned} \tag{28}$$

Using the claim (ii), the second term can be bounded by $\text{const. } |\partial\Lambda|$.

Cluster Expansions for Lattice Models

There is a variety of applications of cluster expansions to lattice models. As noticed above, the first step is always to rewrite the model in terms of a polymer representation.

High-Temperature Expansions

Let us illustrate this point in the simplest case of the Ising model. Its partition function in volume $\Lambda \subset \mathbb{Z}^d$, with free boundary conditions and vanishing external field, is

$$Z_\Lambda(\beta) = \sum_{\sigma_\Lambda} \exp \left\{ \sum_{\substack{x,y \in \Lambda \\ |x-y|=1}} \sigma_x \sigma_y \right\} \quad [29]$$

Using the identity

$$e^{\beta \sigma_x \sigma_y} = \cosh \beta + \sigma_x \sigma_y \sinh \beta \quad [30]$$

it can be rewritten in the form

$$Z_\Lambda(\beta) = 2^{|\Lambda|} (\cosh \beta)^{|B(\Lambda)|} \sum_B (\tanh \beta)^{|B|} \quad [31]$$

Here, the sum runs over all subsets B of the set $B(\Lambda)$ of all bonds in Λ (pairs of nearest-neighbor sites from Λ) such that each site is contained in an even number of bonds from B . Using $\Lambda(B)$ to denote the set of sites contained in bonds from B , we say that $B_1, B_2 \subset B(\Lambda)$ are disjoint if $\Lambda(B_1) \cap \Lambda(B_2) = \emptyset$. Splitting now B into a collection $\mathcal{B} = \{B_1, \dots, B_k\}$ of its connected components called (high-temperature) *polymers* and using $\mathcal{B}(\Lambda)$ to denote the set of all polymers in Λ , we are getting

$$Z_\Lambda(\beta) = 2^{|\Lambda|} (\cosh \beta)^{|B(\Lambda)|} \sum_{\mathcal{B} \subset \mathcal{B}(\Lambda)} \prod_{B \in \mathcal{B}} (\tanh \beta)^{|B|} \quad [32]$$

with the sum running over all collections \mathcal{B} of mutually disjoint polymers. This expression is exactly of the form [18], once we define compatibility of polymers by their disjointness. Introducing the weights

$$w(B) = (\tanh \beta)^{|B|} \quad [33]$$

and taking the set $\mathcal{B}(\Lambda)$ of all polymers in Λ for W , we get the *polymer representation* $Z_\Lambda(\beta) = 2^{|\Lambda|} (\cosh \beta)^{|B(\Lambda)|} Z_{\mathcal{B}(\Lambda)}(w)$.

To apply the cluster expansion theorem, we have to find a function r such that the right-hand side of [24] is positive and yields thus the radius of a polydisk of convergence. Taking $r(B) = \epsilon^{|B|}$ with a suitable ϵ , we get

$$\prod_{B' \in \mathcal{N}(B)} (1 - r(B')) \geq e^{-2|B|} \quad [34]$$

allowing to choose $R(B) = r(B)e^{-2|B|} = (\epsilon e^{-2})^{|B|}$. Indeed, to verify [34] we just notice that the number of polymers of size n containing a fixed site is bounded by κ^n with a suitable constant κ . Thus,

$$\sum_{B': \Lambda(B') \ni x} \epsilon^{|B'|} \leq \sum_{n=1}^{\infty} \kappa^n \epsilon^n \leq 1 \quad [35]$$

once ϵ is sufficiently small, and thus

$$\sum_{B' \in \mathcal{N}(B)} \epsilon^{|B'|} \leq |\Lambda(B)| \leq |B| \quad [36]$$

yielding [34] ($1 - t > e^{-2t}$ for $t < 1/2$). To have $w \in \mathcal{D}_{W,R}$ (for any W) is, for $R(B) = (\epsilon e^{-2})^{|B|}$, sufficient to take $\beta \leq \beta_0$ with $\tanh \beta_0 = \epsilon e^{-2}$.

As a consequence, for $\beta \leq \beta_0$ we can use the cluster expansion theorem to obtain a convergent power series in powers of $\tanh \beta$. In particular, using $\Lambda(X) = \cup_{B \in \text{supp} X} \Lambda(B)$, we get the pressure by the explicit formula

$$\beta p(\beta) = \log 2 + d \log(\cosh \beta) + \sum_{X: \Lambda(X) \ni x} \frac{a(X)}{|\Lambda(X)|} w^X \quad [37]$$

for any fixed $x \in \mathbb{Z}^d$ (by translation invariance of the contributing terms, the choice of x is irrelevant). The function $\beta p(\beta)$ is analytic on the region $\beta \leq \beta_0$ since it is obtained as a uniformly absolutely convergent series of analytic terms $(\tanh \beta)^{|X|}$.

This type of *high-temperature cluster expansion* can be extended to a large class of models with Boltzmann factor in the form $\exp\{-\beta \sum_A U_A(\phi)\}$, where $\phi = (\phi_x; x \in \mathbb{Z}^d)$ is the configuration with *a priori* on-site probability distribution $\nu(d\phi_x)$ and U_A , for any finite $A \subset \mathbb{Z}^d$, are the multi-site interactions (depending only on $(\phi_x; x \in A)$). Using the Mayer trick we can rewrite

$$\exp\left\{-\beta \sum_{A \subset \Lambda} U_A(\phi)\right\} = \prod_A (1 + f_A(\phi)) \quad [38]$$

with $f_A(\phi) = \exp\{-\beta U_A(\phi)\} - 1$. Expanding the product we will get a polymer representation with polymers \mathcal{A} consisting of connected collections $\mathcal{A} = (A_1, \dots, A_k)$ with weights

$$w(\mathcal{A}) = \int \prod_{A \in \mathcal{A}} f_A(\phi) \prod_{x \in \cup_{A \in \mathcal{A}} A} \nu(d\phi_x) \quad [39]$$

under appropriate bounds on the interactions U_A and for β small enough, using $\Lambda(\mathcal{A})$ to denote the set $\cup_{A \in \mathcal{A}} A$, we get,

$$\sum_{\mathcal{A}: \Lambda(\mathcal{A}) \ni x} |w(\mathcal{A})| \leq 1 \quad [40]$$

This assumption allows, as before in the case of the high-temperature Ising model, to apply the cluster expansion theorem yielding an explicit series expansion for the pressure.

Correlations

Cluster expansions can be applied for evaluation of decay of correlations. Let us consider, for the class of models discussed above, the expectation

$$\langle \Psi \rangle_\Lambda = \frac{1}{Z_\Lambda} \int \Psi(\phi) e^{-\beta H_\Lambda(\phi)} \prod_{x \in \Lambda} \nu(d\phi_x) \quad [41]$$

with $H_\Lambda(\phi) = \sum_{A \subset \Lambda} U_A(\phi)$ and a function Ψ depending only on variables ϕ_x on sites x from a finite set $S \subset \Lambda \subset \mathbb{Z}^d$.

A convenient way of evaluating the expectation starts with introduction of the modified partition function

$$Z_{\Lambda, \Psi}(\alpha) = Z_\Lambda + \alpha Z_{\Lambda, \Psi} = Z_\Lambda(1 + \alpha \langle \Psi \rangle_\Lambda) \quad [42]$$

Clearly,

$$\langle \Psi \rangle_\Lambda = \left. \frac{d \log Z_{\Lambda, \Psi}(\alpha)}{d\alpha} \right|_{\alpha=0} \quad [43]$$

Thus, one may get an expression for the expectation $\langle \Psi \rangle_\Lambda$, by forming a polymer representation of $Z_{\Lambda, \Psi}(\alpha)$ and isolating terms linear in α in the corresponding cluster expansion. For the first step, in the just cited high-temperature case with general multi-site interactions, we first enlarge the original set $\mathcal{A}(\Lambda)$ of all polymers in Λ (consisting of connected collections $\mathcal{A} = (A_1, \dots, A_k)$) to $\mathcal{W}_S(\Lambda) = \mathcal{A}(\Lambda) \cup \mathcal{A}_S(\Lambda)$, where $\mathcal{A}_S(\Lambda)$ is the set of all collections $(\mathcal{A}_1, \dots, \mathcal{A}_k)$ of polymers such that each of them intersects the set S (polymers (A_1, \dots, A_k) are “glued” by S into a single entity). Compatibility is defined as before by disjointness; in addition, any two collections from $\mathcal{A}_S(\Lambda)$ are declared to be incompatible as well as any polymer \mathcal{A} from $\mathcal{A}(\Lambda)$ intersecting S is considered to be incompatible with any collection from $\mathcal{A}_S(\Lambda)$. Defining now $w_\alpha(\mathcal{A}) = w(\mathcal{A})$ for $\mathcal{A} \in \mathcal{A}(\Lambda)$ and

$$w_\alpha(\mathcal{A}) = \alpha \int \Psi(\phi) e^{-\beta H_\Lambda(\phi)} \prod_{x \in \cup_{A \in \mathcal{A}_1 \cup \dots \cup \mathcal{A}_k} A \cup S} \nu(d\phi_x) \quad [44]$$

for $\mathcal{A} = (\mathcal{A}_1, \dots, \mathcal{A}_k) \in \mathcal{A}_S(\Lambda)$, we get $Z_{\Lambda, \Psi}(\alpha)$ exactly in the form [18],

$$Z_{\Lambda, \Psi}(\alpha) = \sum_{\mathcal{I} \subset \mathcal{W}_S(\Lambda)} \prod_{\mathcal{A} \in \mathcal{I}} w_\alpha(\mathcal{A}) \quad [45]$$

As a result, we have

$$\log Z_{\Lambda, \Psi}(\alpha) = \sum_{X \in \mathcal{X}(\mathcal{W}_S(\Lambda))} a(X) w_\alpha^X \quad [46]$$

allowing easily to isolate terms linear in α : namely, the terms with multi-indices X with $\text{supp } X \cap \mathcal{A}_S(\Lambda)$ consisting of a single collection, say \mathcal{A}_0 , that occurs with multiplicity one, $X(\mathcal{A}_0) = 1$. Explicitly, using

$$\begin{aligned} \mathcal{X}_{S, \mathcal{A}_0}(\Lambda) &= \{X \in \mathcal{X}(\mathcal{W}_S(\Lambda)) : \text{supp } X \cap \mathcal{A}_S(\Lambda) \\ &= \{\mathcal{A}_0\}, X(\mathcal{A}_0) = 1\} \end{aligned} \quad [47]$$

we get

$$\langle \Psi \rangle_\Lambda = \sum_{\mathcal{A}_0 \in \mathcal{A}_S(\Lambda)} \sum_{X \in \mathcal{X}_{S, \mathcal{A}_0}(\Lambda)} a(X) w^X \quad [48]$$

It is easy to show that, for sufficiently small β , the series on the right-hand side is absolutely convergent even if

we extend $\mathcal{A}_S(\Lambda)$ to $\mathcal{A}_S = \cup_\Lambda \mathcal{A}_S(\Lambda)$ and $\mathcal{X}_{S, \mathcal{A}_0}(\Lambda)$ to $\mathcal{X}_{S, \mathcal{A}_0} = \cup_\Lambda \mathcal{X}_{S, \mathcal{A}_0}(\Lambda)$. As a result, we have an explicit expression for the limiting expectation $\langle \Psi \rangle$ in terms of an absolutely convergent power series. This can be immediately applied to show that $|\langle \Psi \rangle - \langle \Psi \rangle_\Lambda|$ decay exponentially in distance between S and the complement of Λ . Indeed, it suffices to find a suitable bound on $\sum_X |a(X)| |w|^X$ with the sum running over all clusters X reaching from the set S to Λ^c . To this end one does not need to evaluate explicitly the number of clusters of given “diameter” $\text{diam}(X) = \sum_{\mathcal{A}} X(\mathcal{A}) \text{diam}(\Lambda(\mathcal{A})) = m$ with $m \geq \text{dist}(S, \Lambda^c)$. The needed estimate is actually already contained in the condition (ii) from the cluster expansion theorem. It just suffices to choose a suitable k and assume that β is small enough to assure validity of (40) in a stronger form, $\sum_{\mathcal{A}: \Lambda(\mathcal{A}) \ni x} |w(\mathcal{A})| K^{|\Lambda(\mathcal{A})|} \leq 1$, yielding eventually

$$\begin{aligned} \sum_{X: \text{diam}(X) \geq \text{dist}(S, \Lambda^c)} |a(X)| |w|^X &\leq K^{-\text{dist}(S, \Lambda^c)} |S| \\ &\sum_{X: \cup_{\mathcal{A} \in \text{supp } X} \Lambda(\mathcal{A}) \ni x} |a(X)| |w|^X K^{\sum X(\mathcal{A}) |\Lambda(\mathcal{A})|} \\ &\leq |S| K^{-\text{dist}(S, \Lambda^c)} \end{aligned} \quad [49]$$

Exponential decay of correlations $\langle \Psi_1; \Psi_2 \rangle_\Lambda = \langle \Psi_1 \Psi_2 \rangle_\Lambda - \langle \Psi_1 \rangle_\Lambda \langle \Psi_2 \rangle_\Lambda$ (and the limiting $\langle \Psi_1; \Psi_2 \rangle$) in distance between supports of Ψ_1 and Ψ_2 can be established in a similar way by isolating terms proportional to $\alpha_1 \alpha_2$ in the cluster expansion of $\log Z_{\Lambda, \Psi_1, \Psi_2}(\alpha_1, \alpha_2)$ with

$$\begin{aligned} Z_{\Lambda, \Psi_1, \Psi_2}(\alpha_1, \alpha_2) \\ = Z_\Lambda(1 + \alpha_1 \langle \Psi_1 \rangle_\Lambda + \alpha_2 \langle \Psi_2 \rangle_\Lambda + \alpha_1 \alpha_2 \langle \Psi_1 \Psi_2 \rangle_\Lambda) \end{aligned} \quad [50]$$

The resulting claim can be readily generalized to one about the decay of the correlation $\langle \Psi_1; \dots; \Psi_k \rangle$ in terms of the shortest tree connecting supports S_1, \dots, S_k of the functions Ψ_1, \dots, Ψ_k .

Low-Temperature Expansions

Finally, in some models with symmetries, we can apply cluster expansion also at low temperatures. Let us illustrate it again in the case of Ising model. This time, we take the partition function $Z_\Lambda^+(\beta)$ with plus boundary conditions. First, let us define for each nearest-neighbor bond $\langle x, y \rangle$ its dual as the $(d - 1)$ -dimensional closed unit hypercube orthogonal to the segment from x to y and bisecting it at its center. For a given configuration σ_Λ , we consider the boundary of the regions of constant spins consisting of the union $\partial(\sigma_\Lambda)$ of all hypercubes that are dual to nearest-neighbor bonds $\langle x, y \rangle$ for which $\sigma_x \neq \sigma_y$. The contours corresponding to σ_Λ are now defined as the connected components of $\partial(\sigma_\Lambda)$. Notice that, under the fixed boundary condition, there is a one-to-one correspondence between configurations σ_Λ and sets Γ of mutually compatible (disconnected) contours in Λ .

Observing that the number of faces in $\partial(\sigma_\Lambda)$ is just the sum of the areas $|\gamma|$ of the contours $\gamma \in \Gamma$, we get the polymer representation

$$Z_\Lambda^+(\beta) = e^{\beta E(\Lambda)} \sum_{\Gamma} \exp\left(-\beta \sum_{\gamma \in \Gamma} |\gamma|\right) \quad [51]$$

where the sum is over all collections of disjoint contours in Λ . Here $E(\Lambda)$ is the set of all bonds $\langle x, y \rangle$ with at least one endpoint x, y in Λ .

The condition [24] with $r(\gamma) = \epsilon^\gamma$ yields a similar bound on the weights $w(\gamma) = e^{-\beta|\gamma|}$ as in the high-temperature expansion. To verify it, for β sufficiently large, boils down to the evaluation of number of contours of size n that contain a fixed site.

As a result, we can employ the cluster expansion theorem to get

$$\log Z_\Lambda^+(\beta) = \beta|E(\Lambda)| + \sum_{X: X \in \mathcal{X}(C(\Lambda))} a(X)w^X \quad [52]$$

with an explicit formula for the limit

$$\beta p(\beta) = \beta d + \sum_{X: A(X) \ni 0} \frac{a(X)}{|A(X)|} w^X \quad [53]$$

Here, $A(X)$ is the set of sites attached to contours from $\text{supp } X$,

$$A(X) = \cup_{\gamma \in \text{supp } X} A(\gamma) \quad [54]$$

with

$$A(\gamma) = \{x \in \mathbb{Z}^d \mid \text{such that } \text{dist}(x, \gamma) \leq 1/2\} \quad [55]$$

As a consequence of the fact that [53] is, for large β , an absolutely convergent sum of analytic terms $a(X)w^X = a(X)e^{-\beta \sum_{\gamma \in X} |\gamma|}$ (considered as functions of β), the function $\beta p(\beta)$ is, for large β , analytic in β .

The fact that one can explicitly express the difference $\log Z_\Lambda^+(\beta) - |\Lambda|\beta p(\beta)$ (cf. [28]) found numerous applications in situations where one needs an accurate evaluation of the influence of the boundary of the region Λ on the partition function. One such example is a study of microscopic behavior of interfaces. The main idea is to use the explicit expression in the form

$$\begin{aligned} Z_\Lambda^+(\beta) &= \exp\{\beta p(\beta)|\Lambda|\} \exp\left\{ \sum_{X: A(X) \cap \Lambda^c \neq \emptyset} a(X)w^X \frac{|A(X) \cap \Lambda|}{|A(X)|} \right\} \\ &= \exp\{\beta p(\beta)|\Lambda|\} \prod_{X: A(X) \cap \Lambda^c \neq \emptyset} (1 + f_X) \end{aligned} \quad [56]$$

Noticing that

$$f_X = \exp\left\{ a(X)w^X \frac{|A(X) \cap \Lambda|}{|A(X)|} \right\} - 1$$

does not vanish only if $A(X) \cap \Lambda \neq \emptyset$, we can expand the product to obtain “decorations” of the boundary $\partial\Lambda$ by clusters f_X . In the case of interface these clusters can be incorporated into the weight of interface, while on a fixed boundary they yield a “wall free energy.”

The possibility of the (low-temperature) polymer representation of the partition function in terms of contours is based on the $+ \leftrightarrow -$ symmetry of the Ising model. In absence of such a symmetry, cluster expansions can still be used, but in the framework of Pirogov–Sinai theory (see Pirogov–Sinai Theory).

Bibliographical Notes

Cluster expansions originated from the works of Ursell, Yvon, Mayer, and others and were first studied in terms of formal power series. The combinatorial and enumeration problems considered in this framework were summarized in Uhlenbeck and Ford (1962). For related topics in modern language, see Bergeron *et al.* (1998). The convergence results for Mayer and virial expansions for dilute gas were first proved in the works of Penrose, Lebowitz, Groenvelde, and Ruelle (see Ruelle (1969) for a detailed survey). General polymer models on lattice were discussed by Gruber and Kunz (1971) (see also Simon (1993) for discussion of high-temperature and low-temperature cluster expansions of lattice models). Abstract polymer models were introduced in Kotecký and Preiss (1986). An elegant proof of a general claim presented by Dobrushin (1996) was further extended and summarized by Scott and Sokal (2005). We follow their reformulation of the Dobrushin condition. Cluster expansions with a view on applications in quantum field theory are reviewed in Brydges (1986).

See also: Phase Transitions in Continuous Systems; Pirogov–Sinai Theory; Wulff Droplets.

Further Reading

- Bergeron F, Labelle G, and Leroux P (1998) *Combinatorial Species and Tree-Like Structures*, Coll. Encyclopaedia of Mathematics and Its Applications, vol. 67. Cambridge, MA: Cambridge University Press.
- Brydges DC (1986) A short course on cluster expansions. In: Osterwalder K and Stora R (eds.) *Critical Phenomena, Random Systems, Gauge Theories*, pp. 129–183. Les Houches, Session XLIII, 1984. Amsterdam/New York: Elsevier.
- Dobrushin RL (1996) Estimates of semi-invariants for the Ising model at low temperatures. In: Dobrushin RL, Minlos RA, Shukin MA, and Vershik AM (eds.) *Topics in Statistical and Theoretical Physics*, pp. 59–81. Providence, RI: American Mathematical Society.
- Gruber C and Kunz H (1971) General properties of polymer systems. *Communications Mathematical Physics* 22: 133–161.
- Kotecký R and Preiss D (1986) Cluster expansion for abstract polymer models. *Communications in Mathematical Physics* 103: 491–498.

- Ruelle D (1969) *Statistical Mechanics: Rigorous Results*, The Mathematical Physics Monograph Series. Reading, MA: Benjamin.
- Scott AD and Sokal AD (2005) The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *Journal of Statistical Physics* 118: 1151–1261.

- Simon B (1993) *The Statistical Mechanics of Lattice Gases*, Princeton Series in Physics, vol. 1. Princeton: Princeton University Press.
- Uhlenbeck GE and Ford GW (1962) The theory of linear graphs with applications to the theory of the virial development of the properties of gases. In: de Boer J and Uhlenbeck GE (eds.) *Studies in Statistical Mechanics*, vol. I, Amsterdam: North-Holland.

Coherent States

S T Ali, Concordia University, Montreal, QC, Canada

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Very generally, a family of coherent states is a set of continuously labeled quantum states, with specific mathematical and physical properties, in terms of which arbitrary quantum states can be expressed as linear superpositions. Since coherent states are continuously labeled, they form overcomplete sets of vectors in the Hilbert space of states. Originally these states were introduced into physics by Schrödinger (1926), as a family of quantum states in terms of which the transition from quantum to classical mechanics could be conveniently studied. These states have the minimal uncertainty property, in the sense that they saturate the Heisenberg uncertainty relations. The name coherent state was applied when these states were rediscovered in the context of quantum optical radiation by Glauber, Klauder, and Sudarshan. It was demonstrated that in these states the correlation functions of the quantum optical field factorize as they do in classical optics, so that the optical field has a near-classical behavior, with the optical beam being coherent. In this article, we shall refer to these originally studied coherent states as canonical coherent states (CCS).

The canonical coherent states, apart from their use in quantum optics, have also been found to be extremely useful in computations in atomic and molecular physics, in quantum statistical mechanics, and in certain areas of mathematics and mathematical physics, including harmonic analysis, symplectic geometry, and quantization theory. Their wide applicability has prompted the search for other families of states sharing similar mathematical and physical properties. These other families of states are usually called generalized coherent states, even when there is no link to optical coherence in such studies.

Some Properties of CCS

In addition to the minimal uncertainty property, the canonical coherent states have a number of analytical

and group-theoretical properties which are taken as starting points in looking for generalizations. We now define the canonical coherent states mathematically and enumerate a few of these properties.

Suppose that the vectors $|0\rangle, |1\rangle, \dots, |n\rangle, \dots$, correspond to quantum states of $0, 1, \dots, n, \dots$, excitons, respectively. The Hilbert space of these states, in which they form an orthonormal basis, is often known as Fock space. The canonical coherent states are then defined in terms of this basis, for each complex number z , by the analytic expansion:

$$|z\rangle = e^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle \quad [1]$$

The states $|z\rangle$ are normalized to unity: $\langle z|z\rangle = 1$. They satisfy the formal eigenvalue equation

$$a|z\rangle = z|z\rangle \quad [2]$$

where a is the annihilation operator for excitons, which acts on the basis vectors (Fock states) $|n\rangle$ as follows:

$$a|n\rangle = \sqrt{n}|n-1\rangle \quad [3]$$

Its adjoint a^\dagger has the action

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad [4]$$

and

$$[a, a^\dagger] = aa^\dagger - a^\dagger a = I \quad [5]$$

I being the identity operator on Fock space. Introducing the self-adjoint operators Q and P , of position and momentum, respectively,

$$Q = \frac{a + a^\dagger}{\sqrt{2}}, \quad P = \frac{a - a^\dagger}{i\sqrt{2}} \quad [6]$$

it is possible to demonstrate the minimal uncertainty property referred to above (we take $\hbar = 1$):

$$\langle \Delta Q \rangle \langle \Delta P \rangle = \frac{1}{2} \quad [7]$$

where for any observable A ,

$$\langle \Delta A \rangle = \left[\langle z|A^2|z\rangle - \langle z|A|z\rangle^2 \right]^{1/2}$$

is its dispersion in the state $|z\rangle$.

One can also prove the resolution of the identity,

$$\int_{\mathbb{C}} |z\rangle\langle z| \frac{dq dp}{2\pi} = I \tag{8}$$

where $z = (1/\sqrt{2})(q - ip)$ has been written in terms of its real and imaginary parts $(1/\sqrt{2})q$ and $(1/\sqrt{2})p$, respectively. The above operator integral is to be understood in the weak sense, as will be explained later. Equation [8] incorporates the mathematical fact that the set of vectors $|z\rangle$ is overcomplete in the Hilbert space. Indeed, using [8] any vector $|\phi\rangle$ in the Hilbert space can be written as a linear (integral) superposition of these states:

$$|\phi\rangle = \int_{\mathbb{C}} \overline{\Psi(z)} |z\rangle \frac{dq dp}{2\pi}$$

where Ψ is the component function, $\Psi(z) = \langle \phi | z \rangle$. Thus, the coherent states $|z\rangle$ form a continuously labeled total set of vectors in the Hilbert space and since this space is separable, they are an overcomplete set.

Analytic properties of the vectors $|z\rangle$ emerge when the scalar product $\langle \phi | z \rangle$ is taken with respect to an arbitrary vector $|\phi\rangle$ in Fock space. From [1] it is clear that

$$F(z) = \langle \phi | z \rangle = e^{-|z|^2/2} f(z)$$

where f is an entire analytic function in the complex variable z . Moreover, the mapping $\phi \mapsto f$ is an isometric embedding of the Fock space onto the Hilbert space of analytic functions, with respect to the norm

$$\|f\| = \left[\int_{\mathbb{C}} |f(z)|^2 d\mu(z, \bar{z}) \right]^{1/2} \tag{9}$$

defined by the measure $d\mu(z, \bar{z}) = (1/2\pi)e^{-|z|^2} dq dp$.

Group-theoretical properties of the CCS can be demonstrated by noting that

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}} |0\rangle \text{ and } a|0\rangle = 0$$

using which [1] can be recast into the form

$$\begin{aligned} |z\rangle &= e^{-|z|^2/2} e^{za^\dagger} |0\rangle = U(z)|0\rangle \\ U(z) &= e^{za^\dagger - \bar{z}a} \end{aligned} \tag{10}$$

The vectors $|z\rangle$ and the unitary operator $U(z)$ can be reexpressed in terms of the real variables q, p and the operators Q, P as

$$\begin{aligned} |z\rangle &= |q, p\rangle = U(q, p)|0\rangle \\ U(q, p) &= e^{i(pQ - qP)} \end{aligned} \tag{11}$$

The operators $U(q, p)$ realize a (projective) unitary, irreducible representation of the Weyl–Heisenberg group, which is the group whose Lie algebra has the generators Q, P , and I , obeying the commutation relations $[Q, P] = iI$. The existence of the resolution of the identity [8] is the statement of the fact that this representation is square integrable (a notion which will be elaborated upon in the section “Some examples”) which gives us the next paradigm for building coherent states, namely by the action, on a fixed vector, of the unitary operators of a square-integrable representation of a locally compact group.

The above range of properties, which are enjoyed by the CCS, cannot all be expected to hold when looking for generalizations. It then becomes necessary to adopt one or other of these properties as the starting point and to proceed from there. In so doing, it is best first to set down a general definition of coherent states, involving a minimal mathematical structure. Motivated more by possible applications to physics, we do this in the following section.

General Definition

Let \mathfrak{H} be an abstract, separable Hilbert space over the complexes, X a locally compact space and $d\nu$ a measure on X . Let $|x, i\rangle$ be a family of vectors in \mathfrak{H} , defined for each x in X and $i = 1, 2, 3, \dots, N$, where N is usually a finite integer, although it could also be infinite. We assume that this set of vectors possesses the following properties:

1. For each i , the mapping $x \mapsto |x, i\rangle$ is weakly continuous, that is, for each vector $|\phi\rangle$ in \mathfrak{H} , the function $\Psi_i(x) = \langle x, i | \phi \rangle$ is continuous (in the topology of X).
2. For each x in X , the vectors $|x, i\rangle, i = 1, 2, \dots, N$, are linearly independent.
3. The resolution of the identity

$$\sum_{i=1}^N \int_X |x, i\rangle\langle x, i| d\nu(x) = I_{\mathfrak{H}} \tag{12}$$

holds in the weak sense on the Hilbert space \mathfrak{H} , that is, for any two vectors $|\phi\rangle, |\psi\rangle$ in \mathfrak{H} , the following equality holds:

$$\sum_{i=1}^N \int_X \langle \phi | x, i \rangle \langle x, i | \psi \rangle d\nu(x) = \langle \phi | \psi \rangle$$

A set of vectors $|x, i\rangle$ satisfying the above three properties is called a family of generalized vector coherent states. In case $N = 1$, the set is called a family of generalized coherent states. Sometimes the resolution of the identity condition is replaced by a weaker

condition, with the vectors $|x, i\rangle$ simply forming a total set in \mathfrak{H} and the functions $F_i(x) = \langle x, i|\phi\rangle$, as $|\phi\rangle$ runs through \mathfrak{H} , forming a reproducing kernel Hilbert space. Alternatively, the identity on the right-hand side of [12] could also be replaced by a bounded, positive operator T with bounded inverse. In this case, the term frame is also used for the family of generalized coherent states. For physical applications, however, the resolution of the identity condition is always assumed to hold, although the measure $d\nu$ could be of a very general nature (possibly also singular). The objective in all these cases is to ensure that an arbitrary vector $|\phi\rangle$ be expressible as a linear (integral) combination of these vectors. Indeed, [12] is immediately seen to imply that

$$|\phi\rangle = \sum_{i=1}^N \int_X \Psi_i(x) |x, i\rangle d\nu(x) \quad [13]$$

where $\Psi_i(x) = \langle x, i|\phi\rangle$.

Associated to a family of generalized coherent states on a Hilbert space \mathfrak{H} , there is an intrinsic isomorphism between this space and a Hilbert space of (in general, vector valued) continuous functions over X . Using this isomorphism, it is always possible to look upon coherent states as a family of continuous functions which are square integrable with respect to the measure $d\nu$. To demonstrate this, we note that, in view of [12], for each vector $|\phi\rangle$ in \mathfrak{H} , the vector-valued function $\Psi(x)$ on x , with components $\Psi_i(x) = \langle x, i|\phi\rangle$, $i = 1, 2, \dots, N$, satisfies the norm condition

$$\sum_{i=1}^N \int_X |\Psi_i(x)|^2 d\nu(x) = \|\phi\|_{\mathfrak{H}}^2$$

This means that the set of vectors Ψ , as $|\phi\rangle$ runs through \mathfrak{H} , is a closed subspace of the Hilbert space $L_{\mathbb{C}^N}^2(X, d\nu)$ of N -vector-valued functions on x . Let us denote this subspace by \mathfrak{H}_K and note that this space is a reproducing kernel Hilbert space with a matrix-valued kernel $K(x, y)$ having matrix elements

$$K(x, y)_{ij} = \langle x, i|y, j\rangle, \quad i, j = 1, 2, \dots, N \quad [14]$$

and enjoying the properties

$$K(x, y)_{ij} = \overline{K(y, x)_{ji}}, \quad K(x, x)_{ii} > 0 \quad [15]$$

and

$$\sum_{\ell=1}^N \int_X K(x, z)_{i\ell} K(z, y)_{\ell j} d\nu(z) = K(x, y)_{ij} \quad [16]$$

If e^i , $i = 1, 2, \dots, N$, are the vectors constituting the canonical basis of \mathbb{C}^N , then for each x in X and $i = 1, 2, \dots, N$, the vector-valued function ξ_x^i on X ,

defined by $\xi_x^i(y) = K(y, x)e^i$, is the image in \mathfrak{H}_K of the generalized vector coherent state $|x, i\rangle$, under the above-mentioned isometry. The vectors ξ_x^i span the space \mathfrak{H}_K and for an arbitrary element Ψ of this Hilbert space, the reproducing property [16] of the kernel implies the relation

$$\int_X K(x, y) \Psi(y) d\nu(y) = \Psi(x) \quad [17]$$

Conversely, given any reproducing kernel Hilbert space, with a kernel satisfying the relations [15] and [16], generalized coherent states can be constructed as above in terms of this kernel. Mathematically, therefore, generalized coherent states are just the set of vectors naturally defined by the kernel in a reproducing kernel Hilbert space.

Some Examples

We present in this section some of the more commonly used types of coherent states, as illustrations of the general structure given above.

A large class of generalizations of the canonical coherent states [1] is obtained by a simple modification of their analytic structure. Let $x_1 \leq x_2 \leq \dots \leq x_n \leq \dots$ be an infinite sequence of positive numbers ($x_1 \neq 0$). Define $x_n! = x_1 x_2 \dots x_n$ and by convention set $x_0! = 1$. In the same Fock space in which the CCS were described, we now define the related deformed or nonlinear coherent states via the analytic expansion

$$|z\rangle = \mathcal{N}(|z|^2)^{-1/2} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{x_n!}} |n\rangle \quad [18]$$

The normalization factor $\mathcal{N}(|z|^2)$ is chosen so that $\langle z|z\rangle = 1$. These generalized coherent states are overcomplete in the Fock space and satisfy a resolution of the identity of the type

$$\int_{\mathcal{D}} |z\rangle \langle z| \mathcal{N}(|z|^2) d\nu(z, \bar{z}) = I \quad [19]$$

\mathcal{D} being an open disk in the complex plane of radius L , the radius of convergence of the series $\sum_{n=0}^{\infty} (z^n / \sqrt{x_n!})$. (In the case of the CCS, $L = \infty$.) The measure $d\nu$ is generically of the form $d\theta d\lambda(r)$ (for $z = re^{i\theta}$), where $d\lambda$ is related to the $x_n!$ through the moment condition

$$\frac{x_n!}{2\pi} = \int_0^L r^{2n} d\lambda(r), \quad n = 0, 1, 2, \dots \quad [20]$$

This means that once the quantities $x_n!$ are specified, the measure $d\lambda$ is to be determined by solving the

moment problem [20], which of course may not always have a solution. This puts a constraint on the type of sequences $\{x_n\}$ which may be used in the construction.

Once again, we see that for an arbitrary vector $|\phi\rangle$ in the Fock space, the function $F(z) = \langle \phi | z \rangle$, of the complex variable z , is of the form $F(z) = \mathcal{N}(|z|^2)^{-1/2} f(z)$, where f is an analytic function on the domain \mathcal{D} . The reproducing kernel associated to these coherent states is

$$K(\bar{z}, z') = \langle z | z' \rangle = \left[\mathcal{N}(|z|^2) \mathcal{N}(|z'|^2) \right]^{-1/2} \sum_{n=0}^{\infty} \frac{(\bar{z} z')^n}{x_n!} \quad [21]$$

By analogy with [2], one can define a generalized annihilation operator A by its action on the vectors $|z\rangle$,

$$A|z\rangle = z|z\rangle \quad [22]$$

and its adjoint operator A^\dagger . These act on the Fock states $|n\rangle$ as follows:

$$\begin{aligned} A|n\rangle &= \sqrt{x_n} |n-1\rangle \\ A^\dagger|n\rangle &= \sqrt{x_{n+1}} |n+1\rangle \end{aligned} \quad [23]$$

Depending on the exact values of the quantities x_n , these two operators, together with the identity I and all their commutators, could generate a wide range of algebras including various deformed quantum algebras. The term nonlinear, as often applied to these generalized coherent states, comes again from quantum optics, where many such families of states are used in studying the interaction between the radiation field and atoms, and the strength of the interaction itself depends on the frequency of radiation. Of course, these coherent states will not in general have either the group-theoretical or the minimal uncertainty properties of the CCS.

The following is an example of generalized coherent states of the above type, built over the unit disk, $\mathcal{D} = \{z \in \mathbb{C} \mid |z| < 1\}$: on the Fock space, we define the states

$$|z\rangle = (1 - r^2)^\kappa \sum_{n=0}^{\infty} \left[\frac{(2\kappa)_n}{n!} \right]^{1/2} z^n |n\rangle \quad r = |z| \quad [24]$$

where $\kappa = 1, 3/2, 2, 5/2, \dots$, and

$$\begin{aligned} (a)_m &= \frac{\Gamma(a+m)}{\Gamma(a)} \\ &= a(a+1)(a+2) \cdots (a+m-1) \end{aligned}$$

Comparing [24] with [18] we see that $x_n = n/(2\kappa + n - 1)$ so that $\lim_{n \rightarrow \infty} x_n = 1$. Thus, the infinite sum is convergent for any z lying in the unit disk. These

generalized coherent states arise from representations of the group $SU(1, 1)$ belonging to the discrete series, each irreducible representation being labeled by a specific value of the index κ . The associated Hilbert space of functions, analytic on the unit disk, is a subspace of $L^2(\mathcal{D}, d\mu_\kappa)$, with

$$d\mu_\kappa(z, \bar{z}) = (2\kappa - 1) \frac{(1 - r^2)^{2\kappa-2}}{\pi} r dr d\theta$$

$$z = r e^{i\theta}$$

which can be obtained by solving the moment problem [20]. The resolution of the identity satisfied by these states is

$$\frac{2\kappa - 1}{\pi} \int_{\mathcal{D}} |z\rangle \langle z| \frac{r dr d\theta}{(1 - r^2)^2} = I \quad [25]$$

The associated generalized creation and annihilation operators are

$$\begin{aligned} A|n\rangle &= \sqrt{\frac{n}{2\kappa + n - 1}} |n-1\rangle \\ A^\dagger|n\rangle &= \sqrt{\frac{n+1}{2\kappa + n}} |n+1\rangle \end{aligned} \quad [26]$$

so that, clearly, $[A, A^\dagger] \neq I$.

Operators A and A^\dagger of the general type defined in [23] are also known as ladder operators. When such operators appear as generators of representations of Lie algebras, their eigenvectors (see [22]) are usually called Barut–Girardello coherent states. As an example, the representation of the Lie algebra of $SU(1,1)$ on the Fock space is generated by the three operators K_+ , K_- , and K_3 , which satisfy the commutation relations

$$[K_3, K_\pm] = \pm K_\pm, \quad [K_-, K_+] = 2K_3 \quad [27]$$

They act on the vectors $|n\rangle$ as follows:

$$\begin{aligned} K_-|n\rangle &= \sqrt{n(2\kappa + n - 1)} |n-1\rangle \\ K_+ &= K_-^\dagger \\ K_3|n\rangle &= (\kappa + n)|n\rangle \end{aligned} \quad [28]$$

Thus, $K_-|0\rangle = 0$ and

$$|n\rangle = \frac{1}{\sqrt{n!(2\kappa)_n}} K_+^n |0\rangle$$

The Barut–Girardello coherent states $|z\rangle$ are now defined as the formal eigenvectors of the ladder operator K_- :

$$K_-|z\rangle = z|z\rangle, \quad z \in \mathbb{C} \quad [29]$$

They have the analytic form

$$|z\rangle = \frac{|z|^{2\kappa-1}}{\sqrt{I_{2\kappa-1}(2|z|)}} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!(2\kappa + n - 1)!}} |n\rangle \quad [30]$$

where $I_\nu(x)$ is the order- ν modified Bessel function of the first kind. These coherent states satisfy the resolution of the identity,

$$\frac{2}{\pi} \int_{\mathbb{C}} |z\rangle\langle z| K_{2\kappa-1}(2r) I_{2\kappa-1}(2r) r \, dr \, d\theta = I \tag{31}$$

$$z = r e^{i\theta}$$

where again, $K_\nu(x)$ is the order- ν modified Bessel function of the second kind.

A nonanalytic extension of the expression [18] is often used to define generalized coherent states associated to physical Hamiltonians having pure point spectra. These coherent states, known as Gazeau–Klauder coherent states, are labeled by action–angle variables. Suppose that we are given the physical Hamiltonian $H = \sum_{n=0}^\infty E_n |n\rangle\langle n|$, with $E_0 = 0$, that is, it has the energy eigenvalues E_n and eigenvectors $|n\rangle$, which we assume to form an orthonormal basis for the Hilbert space of states \mathfrak{H} . Let us write the eigenvalues as $E_n = \omega \epsilon_n$ by introducing a sequence of dimensionless quantities $\{\epsilon_n\}$ ordered as: $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < \dots$. Then, for all $J \geq 0$ and $\gamma \in \mathbb{R}$, the Gazeau–Klauder coherent states are defined as

$$|J, \gamma\rangle = \mathcal{N}(J)^{-1/2} \sum_{k=0}^\infty \frac{J^{n/2} e^{-i\epsilon_n \gamma}}{\sqrt{\epsilon_n!}} |n\rangle \tag{32}$$

where again \mathcal{N} is a normalization factor, which turns out to be dependent on J only. These coherent states satisfy the temporal stability condition

$$e^{-iHt} |J, \gamma\rangle = |J, \gamma + \omega t\rangle \tag{33}$$

and the action identity

$$\langle J, \gamma | H | J, \gamma \rangle_{\mathfrak{H}} = \omega J \tag{34}$$

While these generalized coherent states do form an overcomplete set in \mathfrak{H} , the resolution of the identity is generally not given by an integral relation of the type [12].

For the second set of examples of generalized coherent states, we take the group-theoretical structure of the CCS as the point of departure. Let G be a locally compact group and suppose that it has a continuous, irreducible representation on a Hilbert space \mathfrak{H} by unitary operators $U(g)$, $g \in G$. This representation is called square integrable if there exists a nonzero vector $|\psi\rangle$ in \mathfrak{H} for which the integral

$$c(\psi) = \int_G |\langle \psi | U(g) \psi \rangle|^2 \, d\mu(g) \tag{35}$$

converges. Here $d\mu$ is a Haar measure of G , which for definiteness, we take to be the left-invariant measure. (The value of the above integral is

independent of whether the left- or the right-invariant measure is used, so we could just as well have used the right-invariant measure.) A vector $|\psi\rangle$, satisfying [35], is said to be admissible, and it can be shown that the existence of one such vector guarantees the existence of an entire dense set of such vectors in \mathfrak{H} . Moreover, if the group G is unimodular, that is, if the left- and the right-invariant measures coincide, then the existence of one admissible vector implies that every vector in \mathfrak{H} is admissible. Given a square-integrable representation and an admissible vector $|\psi\rangle$, let us define the vectors

$$|g\rangle = \frac{1}{\sqrt{c(\psi)}} U(g) |\psi\rangle \tag{36}$$

for all g in the group G . These vectors are to be seen as the analogs of the canonical coherent states [11], written there in terms of the representation of the Weyl–Heisenberg group. Next, it can be shown that the resolution of the identity

$$\int_G |g\rangle\langle g| \, d\mu(g) = I_{\mathfrak{H}} \tag{37}$$

holds on \mathfrak{H} . Thus, the vectors $|g\rangle$ constitute a family of generalized coherent states. The functions $F(g) = \langle g | \phi \rangle$ for all vectors $|\phi\rangle$ in \mathfrak{H} are square integrable with respect to the measure $d\mu$ and the set of such functions, which in fact are continuous in the topology of G , forms a closed subspace of $L^2(G, d\mu)$. Furthermore, the mapping $\phi \mapsto F$ is a linear isometry between \mathfrak{H} and $L^2(G, d\mu)$ and under this isometry the representation U gets mapped to a subrepresentation of the left regular representation of G on $L^2(G, d\mu)$.

A typical example of the above construction is provided by the affine group, G_{Aff} . This is the group of all 2×2 matrices of the type

$$g = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \tag{38}$$

a and b being real numbers with $a \neq 0$. We shall also write $g = (b, a)$. This group is nonunimodular, with the left-invariant measure being given by $d\mu(b, a) = (1/a^2) \, db \, da$. (The right-invariant measure is $(1/a) \, db \, da$.) The affine group has a unitary irreducible representation on the Hilbert space $L^2(\mathbb{R}, dx)$. Vectors in $L^2(\mathbb{R}, dx)$ are measurable functions $\phi(x)$ of the real variable x and the (unitary) operators $U(b, a)$ of this representation act on them in the manner

$$(U(b, a)\phi)(x) = \frac{1}{\sqrt{|a|}} \phi\left(\frac{x-b}{a}\right) \tag{39}$$

If ψ is a function in $L^2(\mathbb{R}, dx)$ such that its Fourier transform $\widehat{\psi}$ satisfies the condition

$$\int_{\mathbb{R}} \frac{|\widehat{\psi}(k)|^2}{|k|} dk < \infty \quad [40]$$

then it can be shown to be an admissible vector, that is,

$$c(\psi) = \int_{G_{\text{Aff}}} |\langle \psi | U(b, a)\psi \rangle|^2 \frac{db da}{a^2} < \infty$$

Thus, following the general construction outlined above, the vectors

$$|b, a\rangle = \frac{1}{\sqrt{c(\psi)}} U(b, a)\psi, \quad (b, a) \in G_{\text{Aff}} \quad [41]$$

define a family of generalized coherent states and one has the resolution of the identity

$$\int_{G_{\text{Aff}}} |b, a\rangle \langle b, a| \frac{db da}{a^2} = I \quad [42]$$

on $L^2(\mathbb{R}, dx)$.

In the signal-analysis literature a vector satisfying the admissibility condition [40] is called a mother wavelet and the generalized coherent states [41] are called wavelets. Signals are then identified with vectors $|\phi\rangle$ in $L^2(\mathbb{R}, dx)$ and the function

$$F(b, a) = \langle b, a | \phi \rangle \quad [43]$$

is called the continuous wavelet transform of the signal ϕ .

There exist alternative ways of constructing generalized coherent states using group representations. For example, the Perelomov method is based on the observation that the vector $|0\rangle$, appearing in the construction of the canonical coherent states in [10] and [11] using the representation of the Weyl–Heisenberg group, is invariant up to a phase, under the action of its center. Consequently, the coherent states $|z\rangle$, as written in [10], are labeled, not by elements of the group itself, but only by the points in the quotient space of the group by its (central) phase subgroup. Generally, let G be a locally compact group and U a unitary irreducible representation of it on the Hilbert space \mathfrak{H} . We do not assume U to be square integrable. We fix a vector $|\psi\rangle$ in \mathfrak{H} , of unit norm and denote by H the subgroup of G consisting of all elements h for which

$$U(h)|\psi\rangle = e^{i\omega(h)}|\psi\rangle \quad [44]$$

where ω is a real-valued function of h . Let $X = G/H$ be the left-coset space and x an arbitrary element in X .

Choosing a coset representative $g(x) \in G$, for each coset x , we define the vectors

$$|x\rangle = U(g(x))|\psi\rangle \quad [45]$$

in \mathfrak{H} . The dependence of these vectors on the specific choice of the coset representative $g(x)$, is only through a phase. Thus, if instead of $g(x)$ we took a different representative $g(x)' \in G$ for the same coset x , then since $g(x)' = g(x)h$ for some $h \in H$, in view of [44] we would have $U(g(x)')|\psi\rangle = e^{i\omega(h)}|x\rangle$. Hence, quantum mechanically, both $|x\rangle$ and $U(g(x)')|\psi\rangle$ represent the same physical state and in particular, the projection operator $|x\rangle\langle x|$ depends only on the coset. Vectors $|x\rangle$, defined in this manner, are called Gilmore–Perelomov coherent states. Since U is assumed to be irreducible, the set of all these vectors as x runs through G/H is dense in \mathfrak{H} . In this definition of generalized coherent states, no resolution of the identity is postulated. However, if X carries an invariant measure, under the natural action of G , and if the formal operator B defined as

$$B = \int_X |x\rangle\langle x| d\mu(x)$$

is bounded, then it is necessarily a multiple of the identity and a resolution of the identity is again retrieved.

The Perelomov construction can be used to define coherent states for any locally compact group. On the other hand, there exist other constructions of generalized coherent states, using group representations, which generalize the notion of square integrability to homogeneous spaces of the group. Briefly, in this approach one starts with a unitary irreducible representation U and attempts to find a vector $|\psi\rangle$, a subgroup H and a section $\sigma: G/H \rightarrow G$ such that

$$\int_{G/H} |x\rangle\langle x| d\mu(x) = T \quad [46]$$

where $|x\rangle = U(\sigma(x))|\psi\rangle$, T is a bounded, positive operator with bounded inverse and $d\mu$ is a quasi-invariant measure on $X = G/H$. It is not assumed that $|\psi\rangle$ be invariant up to a phase under the action of H and clearly, the best situation is when T is a multiple of the identity. Although somewhat technical, this general construction is of enormous versatility for semidirect product groups of the type $\mathbb{R}^n \rtimes K$, where K is a closed subgroup of $GL(n, \mathbb{R})$. Thus, it is useful for many physically important groups, such as the Poincaré or the Euclidean group, which do not have square-integrable representations in the sense of the earlier definition (see eqn [35]). The integral condition [46] ensures that any vector $|\phi\rangle$ in \mathfrak{H} can be written in terms of the $|x\rangle$. Indeed, it

is easy to see that one has the integral representation of a vector,

$$|\phi\rangle = \int_X \Psi(x)|x\rangle d\mu(x)$$

$$\Psi(x) = \langle x|T^{-1}\phi\rangle$$

in terms of the generalized coherent states.

The canonical coherent states satisfy the minimal uncertainty relation [7]. It is possible to build families of coherent states by generalizing from this condition. To do this, one typically starts with two self-adjoint generators in the Lie algebra of a particular group representation and then looks for appropriate eigenvectors of a complex combination of these two generators. For two self-adjoint operators B and C on a Hilbert space \mathfrak{H} , satisfying the commutation relation $[B, C] = iD$ and any normalized vector ϕ in \mathfrak{H} , one can prove the Heisenberg uncertainty relation

$$(\Delta B)^2(\Delta C)^2 \geq \frac{\langle D \rangle^2}{4} \tag{47}$$

where $\langle X \rangle = \langle \phi|X\phi\rangle$ and $(\Delta X)^2 = \langle X^2 \rangle - \langle X \rangle^2$, for any operator X on \mathfrak{H} . More generally, one can prove the Schrödinger–Robertson uncertainty relation

$$(\Delta B)^2(\Delta C)^2 \geq \frac{1}{4} [\langle D \rangle^2 + \langle F \rangle^2] \tag{48}$$

where $\langle F \rangle = \langle BC + CB \rangle - 2\langle B \rangle\langle C \rangle$ measures the correlation between B and C in the state ϕ . If $\langle F \rangle = 0$, the above relation reduces to the Heisenberg uncertainty relation. On the other hand, if $\langle D \rangle = 0$, the Heisenberg uncertainty relations become redundant. Suppose now that B and C are two self-adjoint elements of the Lie algebra in the unitary irreducible representation of a Lie group and we look for states $|\phi\rangle$ which minimize the uncertainty relation [48], that is, for which the equality holds. It turns out that such states can be found by considering the linear combination $B + i\lambda C$, for a fixed complex number λ , and solving the formal eigenvalue equation

$$[B + i\lambda C]|z, \lambda\rangle = z|z, \lambda\rangle$$

with $z = \langle B \rangle + i\lambda\langle C \rangle$ [49]

Solutions to this equation for which $|\lambda| = 1$ are called squeezed states, since in this case $\Delta B \neq \Delta C$. Generally, the states $|z, \lambda\rangle$ are known as intelligent states. As an example, for the operators Q and P in [6], for which one has

$$(\Delta Q)^2(\Delta P)^2 \geq \frac{1}{4} [1 + \langle F \rangle^2]$$

taking the combination $Q + i\lambda P$, one obtains the minimal uncertainty states,

$$|z, \lambda\rangle = \mathcal{N}(z, \lambda)^{-1/2} e^{-w(a^\dagger)^2/2} e^{(z/\sqrt{2})(1+w)a^\dagger} |0\rangle \tag{50}$$

$\mathcal{N}(z, \lambda)$ being a normalization constant and $w = (1 - \lambda)/(1 + \lambda)$. The case $\lambda = -1$ does not lead to any solutions, while $\lambda = 1$ gives the canonical coherent states [10]. For real $\lambda \neq 1$ the above states are the well-known squeezed states of quantum optics.

Our final example is that of a family of vector coherent states, which will be obtained essentially by replacing the complex variable z in [18] by a matrix variable. We choose the domain $\Omega = \mathbb{C}^{2 \times 2}$ (all 2×2 complex matrices), equipped with the measure

$$d\nu(\mathfrak{Z}, \mathfrak{Z}^\dagger) = \frac{e^{-\text{tr}[\mathfrak{Z}\mathfrak{Z}^\dagger]}}{\pi^4} \prod_{k,j=1}^2 dx_{kj} \wedge dy_{kj}$$

where \mathfrak{Z} is an element of Ω and $z_{kj} = x_{kj} + iy_{kj}$ are its entries. One can then prove the matrix orthogonality relation

$$\int_{\Omega} \mathfrak{Z}^k \mathfrak{Z}^{\dagger \ell} d\nu(\mathfrak{Z}, \mathfrak{Z}^\dagger)$$

$$= \frac{1}{2} \int_{\Omega} \text{tr}[\mathfrak{Z}^k \mathfrak{Z}^{\dagger \ell}] d\nu(\mathfrak{Z}, \mathfrak{Z}^\dagger) \mathbb{I}_2$$

$$= b(k) \mathbb{I}_2, \quad k, \ell = 0, 1, 2, \dots, \infty \tag{51}$$

\mathbb{I}_2 being the 2×2 identity matrix and

$$b(k) = \frac{(k+3)!}{2(k+1)(k+2)}$$

$$k = 1, 2, 3, \dots, \quad b(0) = 1 \tag{52}$$

Consider the Hilbert space $\tilde{\mathfrak{H}} = L^2_{\mathbb{C}^2}(\Omega, d\nu)$ of square integrable, two-component vector-valued functions on Ω and in it consider the vectors $|\Psi_k^i\rangle, i = 1, 2, k = 0, 1, 2, \dots, \infty$, defined by the \mathbb{C}^2 -valued functions,

$$|\Psi_k^i(\mathfrak{Z}^\dagger)\rangle = \frac{1}{\sqrt{b(k)}} \mathfrak{Z}^{\dagger k} \chi^i \tag{53}$$

where the vectors $\chi^i, i = 1, 2$, form an orthonormal basis of \mathbb{C}^2 . By virtue of [51], the vectors $|\Psi_k^i\rangle$ constitute an orthonormal set in $\tilde{\mathfrak{H}}$, that is,

$$\langle \Psi_k^i | \Psi_\ell^j \rangle_{\tilde{\mathfrak{H}}} = \delta_{k\ell} \delta_{ij}$$

Denote by \mathfrak{H}_K the Hilbert subspace of $\tilde{\mathfrak{H}}$ generated by this set of vectors. This can be shown to be a reproducing kernel Hilbert space of analytic

functions in the variable \mathfrak{Z}^\dagger , with the matrix valued kernel $K : \Omega \times \Omega \rightarrow C^{2 \times 2}$:

$$\begin{aligned}
 K(\mathfrak{Z}^\dagger, \mathfrak{Z}) &= \sum_{i=1}^2 \sum_{k=0}^\infty \Psi_k^i(\mathfrak{Z}^\dagger) \Psi_k^i(\mathfrak{Z})^\dagger \\
 &= \sum_{i=1}^2 \sum_{k=0}^\infty \frac{\mathfrak{Z}^{\dagger k} \mathfrak{Z}^k}{b(k)} \tag{54}
 \end{aligned}$$

Vector coherent states in \mathfrak{H}_K are then naturally associated to this kernel and are given by

$$\begin{aligned}
 |\mathfrak{Z}, i\rangle &= \sum_{j=1}^2 \sum_{k=0}^\infty \frac{\chi^j \mathfrak{Z}^k \chi^i}{\sqrt{b(k)}} |\Psi_k^j\rangle \\
 \text{that is, } |\mathfrak{Z}, i\rangle \langle \mathfrak{Z}^\dagger, i| &= K(\mathfrak{Z}^\dagger, \mathfrak{Z}) \chi^i \tag{55}
 \end{aligned}$$

for $i = 1, 2$ and all \mathfrak{Z} in Ω . They satisfy the resolution of the identity

$$\sum_{i=1}^2 \int_\Omega |\mathfrak{Z}, i\rangle \langle \mathfrak{Z}, i| d\nu(\mathfrak{Z}, \mathfrak{Z}^\dagger) = I_{\mathfrak{H}_K} \tag{56}$$

The expression for the $|\mathfrak{Z}, i\rangle$ in [55], involving the sum, should be compared to [18], of which it is a direct analog.

Some Applications of Coherent States

Generalized coherent states have many applications in physics, signal analysis, and mathematics, of which we mention a few here. As an example of an application of deformed coherent states, we take

$$x_n = \left[\frac{q^n - q^{-n}}{q - q^{-1}} \right]^{1/2}, \quad q > 0 \tag{57}$$

in the definition of these states in [18]. It is then easy to see that the operators A and A^\dagger , defined in [23], satisfy the q -deformed commutation relation

$$AA^\dagger - qA^\dagger A = q^{-N} \tag{58}$$

where N is the usual number operator, which acts on the Fock states as $N|n\rangle = n|n\rangle$. Clearly, in the limit as $q \rightarrow 1$, these q -deformed coherent states go over to the canonical coherent states, with the operators A and A^\dagger becoming the usual creation and annihilation operators a and a^\dagger , respectively. The operators A and A^\dagger and the commutation relation [58] describe a system of q -deformed oscillators, which have been used to describe, for example, the vibrations of polyatomic molecules. The potential energy between the atoms of such a molecule has anharmonic terms, leading to a deformation of the usual oscillator algebra, generated by the operators a and a^\dagger .

As already mentioned, generalized coherent states are widely used in signal analysis. The wavelet transform $F(b, a) = \langle b, a | \phi \rangle$, introduced in [43], is a time–frequency transform, in which the parameter b is identified with time and $1/a$ with frequency. Wavelet transforms are used extensively to analyze, encode, and reconstruct signals arising in many different branches of physics, engineering, seismography, electronic data processing, etc. Similarly, the canonical coherent states, as written in [11], give rise to the transform $F(q, p) = \langle q, p | \phi \rangle$. Again, if q is interpreted as time and p as frequency, then this is just the windowed Fourier transform, also used extensively in signal processing. More general wavelets, from higher-dimensional affine groups, are used to analyze higher-dimensional signals, while wavelet like transforms from other groups have been used to study signals exhibiting different geometries. In particular, wavelet transforms from spherical geometries have been applied to the study of brain signals and to astrophysical data.

Our final example is taken from quantization theory. A quantization technique is a method for performing the transition from a given classical mechanical system to its quantum counterpart. Many methods have been developed to accomplish this and the use of coherent states is one of them. Suppose that we are given a family of coherent states $|x\rangle$ in a Hilbert space \mathfrak{H} , where the set X from which x is taken is a classical phase space. This means that X is a symplectic manifold with an associated 2-form ω , which defines a Poisson bracket on the set of observables of the classical system, which are real-valued functions on X . There is a natural measure $d\omega$, defined on X by the 2-form ω . Let us assume that the coherent states $|x\rangle$ satisfy a resolution on the identity with respect to this measure:

$$\int_X |x\rangle \langle x| d\omega(x) = I_{\mathfrak{H}}$$

In this case, the coherent states may be used to quantize the observables of the classical system in the following way: let f be a real-valued function on X , representing a classical observable and suppose that the formal operator

$$\hat{f} = \int_X f(x) |x\rangle \langle x| d\omega(x) \tag{59}$$

is well defined as a self-adjoint operator on \mathfrak{H} . Then we may take the operator \hat{f} to be the quantized observable corresponding to the classical observable f . Suppose that we have two such operators, \hat{f} and \hat{g} ,

corresponding to the two classical observables f and g , which have the Poisson bracket $\{f, g\}$, defined via the 2-form ω . We then check if the quantization condition

$$\widehat{\{f, g\}} = \frac{2\pi}{i\hbar} [\widehat{f}, \widehat{g}] \tag{60}$$

where \hbar is Planck's constant, is satisfied. Generally this will be the case for a certain number of classical observables. This method of quantization has been most successfully used for manifolds X which have a (complex) Kähler structure. Over such a manifold, one can define a Hilbert space of analytic functions, which has a reproducing kernel and hence a naturally associated set of coherent states. As a specific example, we take the case of canonical coherent states [11]. We can identify the complex plane \mathbb{C} with the phase space \mathbb{R}^2 of a free classical particle having a single degree of freedom. The measure $d\omega$ in this case is just $(1/2\pi)dq dp$. If we now quantize the classical observables $f(q, p) = q$ and $f(q, p) = p$, of position and momentum, respectively, using the canonical coherent states, we obtain the two operators

$$\begin{aligned} Q &= \int_{\mathbb{R}^2} q |q, p\rangle \langle q, p| \frac{dq dp}{2\pi} \\ P &= \int_{\mathbb{R}^2} p |q, p\rangle \langle q, p| \frac{dq dp}{2\pi} \end{aligned} \tag{61}$$

It can be verified that these two operators satisfy the canonical commutation relations $[Q, P] = iI_{\mathfrak{H}}$, as required.

See also: Solitons and Kac–Moody Lie Algebras; Wavelets: Mathematical Theory.

Further Reading

Ali ST, Antoine J-P, and Gazeau J-P (2000) *Coherent States, Wavelets and Their Generalizations*. New York: Springer.
 Ali ST and Engliš M (2005) Quantization methods – a guide for physicists and analysts. *Reviews in Mathematical Physics* 17: 391–490.
 Brif C (1997) SU(2) and SU(1,1) algebra eigenstates: a unified analytic approach to coherent and intelligent states. *International Journal of Theoretical Physics* 36: 1651–1682.
 Klauder JR and Sudarshan ECG (1968) *Fundamentals of Quantum Optics*. New York: Benjamin.
 Klauder JR and Skagerstam BS (1985) *Coherent States – Applications in Physics and Mathematical Physics*. Singapore: World Scientific.
 Perelomov AM (1986) *Generalized Coherent States and their Applications*. Berlin: Springer.
 Schrödinger E (1926) Der stetige Übergang von der Mikro- zur Makromechanik. *Naturwissenschaften* 14: 664–666.
 Sivakumar S (2000) Studies on nonlinear coherent states. *Journal of Optics B: Quantum Semiclass. Opt.* 2: R61–R75.
 Zhang W-M, Feng DH, and Gilmore RG (1990) Coherent states: theory and some applications. *Reviews of Modern Physics* 62: 867–927.

Cohomology Theories

U Tillmann, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The origins of cohomology theory are found in topology and algebra at the beginning of the last century but since then it has become a tool of nearly every branch of mathematics. It's a way of life! Naturally, this article can only give a glimpse at the rich subject. We take here the point of view of algebraic topology and discuss only the cohomology of spaces.

Cohomology reflects the global properties of a manifold, or more generally of a topological space. It has two crucial properties: it only depends on the homotopy type of the space and is determined by local data. The latter property makes it in general computable.

To illustrate the interplay between the local and global structure, consider the Euler characteristic of a compact manifold; as will be explained below, cohomology is a refinement of the Euler characteristic. For simplicity, assume that the manifold M is a surface and that we have chosen a way of dividing the surface into triangles. The Euler characteristic is then defined to be

$$\chi(M) = F - E + V$$

where F denotes the number of faces, E the number of edges, and V the number of vertices in the triangulation. Remarkably, this number does not depend on the triangulation. Yet, this simple, easy to compute number can already distinguish the different types of closed, oriented surfaces: for the sphere we have $\chi = 2$, the torus $\chi = 0$, and in general for any surface M_g of genus g

$$\chi(M_g) = 2 - 2g$$

The Euler characteristic also tells us something about the geometry and analysis of the manifold. For example, the total curvature of a surface is equal to its Euler characteristic. This is the Gauss–Bonnet theorem and an analogous result holds in higher dimensions. Another striking result is the Poincaré–Hopf theorem which equates the Euler characteristic with the total index of a vector field and thus gives strong restrictions on what kind of vector fields can exist on a manifold. This interplay between global analysis and topology has been one of the most exciting and fruitful research areas and is most powerfully expressed in the celebrated Atiyah–Singer index theorem, which determines the analytic index of an elliptic operator, such as the Dirac operator on a spin manifold, in terms of cohomology classes.

Chain Complexes and Homology

There are several different geometric definitions of the cohomology of a topological space. All share some basic algebraic structure which we will explain first.

A “chain complex” (C_*, ∂_*)

$$\cdots C_{i+1} \xrightarrow{\partial_{i+1}} C_i \xrightarrow{\partial_i} C_{i-1} \cdots \xrightarrow{\partial_1} C_0 \quad [1]$$

is a collection of vector spaces (or R -modules more generally) $C_i, i \geq 0$, and linear maps (R -module maps) $\partial_i: C_i \rightarrow C_{i-1}$ with the property that for all i

$$\partial_i \circ \partial_{i+1} = 0 \quad [2]$$

The scalar fields one tends to consider are the rationals \mathbb{Q} , reals \mathbb{R} , complex numbers \mathbb{C} , or a primary field \mathbb{Z}_p , while the most important ring R is the ring of integers \mathbb{Z} though we will also consider localizations such as $\mathbb{Z}[1/p]$, which has the effect of suppressing any p -primary torsion information.

Of particular interest are the elements in C_i that are mapped to zero by ∂_i , the i -dimensional “cycles,” and those that are in the image of ∂_{i+1} , the i -dimensional “boundaries.” Because of [2], every boundary is a cycle, and we may define the quotient vector space (R -module), the i th-dimensional homology,

$$H_i(C_*; \partial_*) := \frac{\ker \partial_i}{\text{im } \partial_{i+1}} \quad [3]$$

(C_*, ∂_*) is “exact” if all its cycles are boundaries. Homology thus measures to what extent the sequence [1] fails to be exact.

Simplicial Homology

A triangulation of a surface gives rise to its “simplicial” chain complex: Taking coefficients in

$\mathbb{Z}, C_2, C_1, C_0$ are the free abelian groups generated by the set of faces, edges, and vertices, respectively; $C_i = \{0\}$ for $i \geq 3$. The map ∂_2 assigns to a triangle the sum of its edges; ∂_1 maps an edge to the sum of its endpoints. If we are working with \mathbb{Z}_2 coefficients, this defines for us a chain complex as [2] is clearly satisfied; in general, one needs to keep track of the orientations of the triangles and edges and take sums with appropriate signs (cf. [6] below). An easy calculation shows that for an oriented, closed surface M_g of genus g , we have

$$\begin{aligned} H_0(M_g; \mathbb{Z}) &= \mathbb{Z} \\ H_1(M_g; \mathbb{Z}) &= \mathbb{Z}^{2g} \\ H_2(M_g; \mathbb{Z}) &= \mathbb{Z} \\ H_i(M_g; \mathbb{Z}) &= 0 \quad \text{for } i \geq 3 \end{aligned} \quad [4]$$

Note that the Euler characteristic can be recovered as the alternating sum of the rank of the homology groups:

$$\chi(M) = \sum_{i=0}^{\dim M} (-1)^i \text{rk } H_i(M; \mathbb{Z}) \quad [5]$$

Every smooth manifold M has a triangulation, so that its simplicial homology can be defined just as above. More generally, simplicial homology can be defined for any simplicial space, that is, a space that is built up out of points, edges, triangles, tetrahedra, etc. Formula [5] remains valid for any compact manifold or simplicial space.

Singular Homology

Let X be any topological space, and let Δ^n be the oriented n -simplex $[v_0, \dots, v_n]$ spanned by the standard basis vectors v_i in \mathbb{R}^{n+1} . The set of singular n -chains $S_n(X)$ is the free abelian group on the set of continuous maps $\sigma: \Delta^n \rightarrow X$. The boundary of σ is defined by the alternating sum of the restriction of σ to the faces of Δ^n :

$$\partial_n(\sigma) := \sum_{i=0}^n (-1)^{-i} \sigma|_{[v_0, \dots, \hat{v}_i, \dots, v_n]} \quad [6]$$

One easily checks that the boundary of a boundary is zero, and hence $(S_*(X), \partial_*)$ defines a chain complex. Its homology is by definition the singular homology $H_*(X; \mathbb{Z})$ of X . For any simplicial space, the inclusion of the simplicial chains into the singular chains induces an isomorphism of homology groups. In particular, this implies that the simplicial homology of a manifold, and hence its Euler characteristic do not depend on its triangulation.

If in the definition of simplicial and singular homology we take free R -modules (where R may

also be a field) instead of free abelian groups, we get the homology $H_*(X; R)$ of X with coefficients in R . The “universal-coefficient theorem” describes the homology with arbitrary coefficients in terms of the homology with integer coefficients. In particular, if R is a field of characteristic zero,

$$\dim H_n(X; R) = \text{rk } H_n(X; \mathbb{Z})$$

Basic Properties of Singular Homology

While simplicial homology (and the more efficient cellular homology which we will not discuss) is easier to compute and easier to understand geometrically, singular homology lends itself more easily to theoretical treatment.

1. *Homotopy invariance.* Any continuous map $f: X \rightarrow Y$ induces a map on homology $f_*: H_*(X; R) \rightarrow H_*(Y; R)$ which only depends on the homotopy class of f .

In particular, a homotopy equivalence $f: X \rightarrow Y$ induces an isomorphism in homology. So, for example, the inclusion of the circle S^1 into the punctured plane $\mathbb{C} \setminus \{0\}$ is a homotopy equivalence, and thus

$$\begin{aligned} H_i(\mathbb{C} \setminus \{0\}; R) &\simeq H_i(S^1; R) \\ &= \begin{cases} \mathbb{Z} & \text{for } i = 0, 1 \\ 0 & \text{for } i \geq 2 \end{cases} \end{aligned}$$

For the one point space we have $H_0(\text{pt}; R) = R$. Define reduced homology by $\tilde{H}_*(X; R) := \ker(H_*(X; R) \rightarrow H_*(\text{pt}; R))$.

2. *Dimension axiom.* $\tilde{H}_i(\text{pt}; R) = 0$ for all i .

More generally, it follows immediately from the definition of simplicial homology that the homology of any n -dimensional manifold is zero in dimensions larger than n .

We mentioned in the introduction that homology depends only on local data. This is made precise by the

3. *Mayer–Vietoris theorem.* Let $X = A \cup B$ be the union of two open subspaces. Then the following sequence is exact:

$$\begin{aligned} \cdots \longrightarrow H_n(A \cap B; R) &\longrightarrow H_n(A; R) \oplus H_n(B; R) \\ &\longrightarrow H_n(X; R) \xrightarrow{\partial} H_{n-1}(A \cap B; R) \\ &\longrightarrow \cdots \longrightarrow H_0(X; R) \longrightarrow 0 \end{aligned}$$

On the level of chains, the first map is induced by the diagonal inclusion, while the second map takes the difference between the first and second summands. Finally, ∂ takes a cycle $c = a + b$ in the chains of X that can be expressed as the sum of a chain a in A

and b in B to $\partial c := \partial_n a = -\partial_n b$. For example, consider two cones, A and B , on a space X and identify them at the base X to define the suspension ΣX of X . Then $\Sigma X = A \cup B$ with $A, B \simeq \text{pt}$ and $A \cap B \simeq X$. The boundary map ∂ is then an isomorphism:

$$\tilde{H}_n(X; R) \simeq H_{n+1}(\Sigma X; R) \quad \text{for all } n \geq 0 \quad [7]$$

From this one can easily compute the homology of a sphere. First note that

$$\tilde{H}_0(X; \mathbb{Z}) = \mathbb{Z}^{k-1}$$

where k is the number of connected components in X . Also, $S^n \simeq \Sigma S^{n-1} \simeq \cdots \simeq \Sigma^n S^0$. Thus, by [7],

$$H_n(S^n; \mathbb{Z}) \simeq \mathbb{Z} \quad \text{and} \quad \tilde{H}_*(S^n; \mathbb{Z}) = 0 \quad \text{for } * \neq n \quad [8]$$

If Y is a subspace of X , relative homology groups $H_*(X, Y; R)$ can be defined as the homology of the quotient complex $S_*(X)/S_*(Y)$. When Y has a good neighborhood in X (i.e., it is a neighborhood deformation retract in X), then, by the “excision theorem,”

$$H_*(X, Y; R) \simeq \tilde{H}_*(X/Y; R)$$

where X/Y denotes the quotient space of X with Y identified to a point. There is a long exact sequence

$$\begin{aligned} \cdots \longrightarrow H_n(Y; R) &\longrightarrow H_n(X; R) \longrightarrow H_n(X, Y; R) \\ &\xrightarrow{\partial} H_{n-1}(Y; R) \longrightarrow \cdots \longrightarrow H_0(X, Y; R) \longrightarrow 0 \end{aligned}$$

This and the Mayer–Vietoris sequence give two ways of breaking up the problem of computing the homology of a space into computing the homology of related spaces. An iteration of this process leads to the powerful tool of spectral sequences (see Spectral Sequences).

Relation to Homotopy Groups

Let $\pi_1(X, x_0)$ denote the fundamental group of X relative to the base point x_0 . These are the based homotopy classes of based maps from a circle to X .

If X is connected, then $H_1(X; \mathbb{Z})$ is the abelianization of $\pi_1(X, x_0)$ [9]

Indeed, every map from a (triangulated) sphere to X defines a cycle and hence gives rise to a homology class. This defines the Hurewicz map $h: \pi_*(X; x_0) \rightarrow H_*(X; \mathbb{Z})$. In general there is no good description of its image. However, if X is k -connected with $k \geq 1$, then h induces an isomorphism in dimension $k + 1$ and an epimorphism in dimension $k + 2$.

Though [9] indicates that homology cannot distinguish between all homotopy types, the fundamental group is in a sense the only obstruction to this. A simple form of the “Whitehead theorem” states:

Theorem *If a map $f : X \rightarrow Y$ between two simplicial complexes with trivial fundamental groups induces an isomorphism on all homology groups, then it is a homotopy equivalence.*

Warning: This does not imply that two simply connected spaces with isomorphic homology groups are homotopic! The existence of the map f inducing this isomorphism is crucial and counterexamples can easily be constructed.

Dual Chain Complexes and Cohomology

The process of dualizing itself cannot be expected to yield any new information. Nevertheless, the cohomology of a space, which is obtained by dualizing its simplicial chain complex, carries important additional structure: it possesses a product, and moreover, when the coefficients are a primary field, it is an algebra over the rich Steenrod algebra. As with homology we start with the algebraic setup.

Every chain complex (C_*, ∂_*) gives rise to a dual chain complex (C^*, ∂^*) where $C^i = \text{hom}_R(C_i, R)$ is the dual R -module of C_i ; because of [2], the composition of two dual boundary morphisms $\partial^{i+1} : C^i \rightarrow C^{i+1}$ is trivial. Hence we may define the i th dimensional cohomology group as

$$H^i(C^*, \partial^*) := \frac{\ker \partial^{i+1}}{\text{im } \partial^i} \quad [10]$$

Evaluation $(\sigma, \phi) \mapsto \phi(\sigma)$ descends to a dual pairing

$$H_n(C_*, \partial_*) \otimes_R H^n(C^*, \partial^*) \rightarrow R$$

and when R is a field, this identifies the cohomology groups as the duals of the homology groups. More generally, the universal-coefficient theorem relates the two. A simple version states: let (C_*, ∂_*) be a chain complex of free abelian groups (such as the simplicial or singular chain complexes) with finitely generated homology groups. Then,

$$H^i(C^*, \partial^*) \simeq H_i^{\text{free}}(C_*, \partial_*) \oplus H_{i-1}^{\text{tor}}(C_*, \partial) \quad [11]$$

where H_*^{tor} denotes the torsion subgroup of H_* and H_*^{free} denotes the quotient group H_*/H_*^{tor} .

Singular Cohomology

The dual $S^*(X)$ of the singular chain complex of a space X carries a natural pairing, the cup product, $\cup : S^p(X) \otimes S^q(X) \rightarrow S^{p+q}(X)$ defined by

$$\begin{aligned} &(\phi_1 \cup \phi_2)(\sigma) \\ &:= \phi_1(\sigma|_{[v_0, \dots, v_p]}) \phi_2(\sigma|_{[v_p, \dots, v_{p+q}]} \end{aligned}$$

This descends to a multiplication on cohomology groups and makes $H^*(X; R) := \bigoplus_{n \geq 0} H^n(X; R)$ into

an associative, graded commutative ring: $u \cup v = (-1)^{\deg u \deg v} v \cup u$.

The ‘‘K unneth theorem’’ gives some geometric intuition for the cup product. A simple version states: for spaces X and Y with $H^*(Y; R)$ a finitely generated free R -module, the cup product defines an isomorphism of graded rings

$$H^*(X; R) \otimes_R H^*(Y; R) \rightarrow H^*(X \times Y; R)$$

For example, for a sphere, all products are trivial for dimension reasons. Hence,

$$H^*(S^n; \mathbb{Z}) = \bigwedge^*(x) \quad [12]$$

is an exterior algebra on one generator x of degree n . On the other hand, the cohomology of the n -dimensional torus T^n is an exterior algebra on n degree-1 generators,

$$H^*(T^n; \mathbb{Z}) = \bigwedge^*(x_1, \dots, x_n) \quad [13]$$

The dual pairing can be generalized to the slant or cap product

$$\cap : H_n(X; R) \otimes_R H^i(X; R) \rightarrow H_{n-i}(X; R)$$

defined on the chain level by the formula $(\sigma, \phi) \mapsto \phi(\sigma|_{[v_0, \dots, v_i]}) \sigma|_{[v_i, \dots, v_n]}$.

Steenrod Algebra

The cup product on the chain level is homotopy commutative, but not commutative. Steenrod used this defect to define operations

$$Sq^i : H^n(X; \mathbb{Z}_2) \rightarrow H^{n+i}(X; \mathbb{Z}_2)$$

for all $i \geq 0$ which refine the cup-squaring operation: when $n = i$, then $Sq^n(x) = x \cup x$. These are natural group homomorphisms which commute with suspension. Furthermore, they satisfy the Cartan and Adem Relations

$$Sq^n(x \cup y) = \sum_{i+j=n} Sq^i(x) \cup Sq^j(y)$$

$$Sq^i Sq^j = \sum_{k=0}^{\lfloor i/2 \rfloor} \binom{j-k-1}{i-2k} Sq^{i+j-k} Sq^k$$

for $i \leq 2j$

The mod-2 Steenrod algebra \mathcal{A} is then the free \mathbb{Z}_2 -algebra generated by the Steenrod squares $Sq^i, i \geq 0$, subject only to the Adem relations. With the help of Adem’s relations, Serre and Cartan found a \mathbb{Z}_2 -basis for \mathcal{A} :

$$\{Sq^I := Sq^{i_1} \cdots Sq^{i_n} \mid i_j \geq 2i_{j+1} \text{ for all } j\}$$

The Steenrod algebra is also a Hopf algebra with a commutative comultiplication $\Delta: \mathcal{A} \rightarrow \mathcal{A} \otimes \mathcal{A}$ induced by

$$\Delta(Sq^n) := \sum_{i+j=n} Sq^i \otimes Sq^j$$

The Cartan relation implies that the mod-2 cohomology of a space is compatible with the comultiplication, that is, $H^*(X; \mathbb{Z}_2)$ is an algebra over the Hopf algebra \mathcal{A} . There are odd primary analogs of the Steenrod algebra based on the reduced p th power operations

$$P^i: H^n(X; \mathbb{Z}_p) \rightarrow H^{n+2i(p-1)}(X; \mathbb{Z}_p)$$

with similar properties to \mathcal{A} .

One of the most striking applications of the Steenrod algebra can be found in the work of Adams on the “vector fields on spheres problem”: for each n , find the greatest number k , denoted $K(n)$, such that there is a k -field on the $(n - 1)$ -sphere S^{n-1} . Recall that a k -field is an ordered set of k pointwise linear independent tangent vector fields. If we write n in the form $n = 2^{4a+b}(2s + 1)$ with $0 \leq b < 4$, Adams proved that $K(n) = 2^b + 8a - 1$. In particular, when n is odd, $K(n) = 0$. We give an outline of the proof for this special case in the next section.

- The failure of associativity of the cup product at the chain level gives rise to secondary operations, the so-called “Massey products.”

Cohomology of Smooth Manifolds

A smooth manifold M of dimension n can be triangulated by smooth simplices $\sigma: \Delta^n \rightarrow M$. If M is compact, oriented, without boundary, the sum of these simplices define a homology cycle $[M]$, the fundamental class of M . The most remarkable property of the cohomology of manifolds is that they satisfy “Poincaré duality”: taking cap product with $[M]$ defines an isomorphism:

$$D := [M] \cap : H^k(M; \mathbb{Z}) \xrightarrow{\cong} H_{n-k}(M; \mathbb{Z}) \quad \text{for all } k \quad [14]$$

In particular, for connected manifolds, $H^n(M; \mathbb{Z}) \cong \mathbb{Z}$; and every map $f: M' \rightarrow M$ between oriented, compact closed manifolds of the same dimension has a degree: $f^*: H^*(M; \mathbb{Z}) \rightarrow H^*(M'; \mathbb{Z})$ is multiplication by an integer $\text{deg}(f)$, the degree of f . For smooth maps, the degree is the number of points in the inverse image of a generic point $p \in M$ counted with signs:

$$\text{deg}(f) = \sum_{p' \in f^{-1}(p)} \text{sign}(p')$$

where $\text{sign}(p')$ is $+1$ or -1 depending on whether f is orientation preserving or reversing in a neighborhood of p' . For example, a complex polynomial of degree d defines a map of the two-dimensional sphere to itself of degree d : a generic point has n points in its inverse image and the map is locally orientation preserving. On the other hand, a map of S^{n-1} induced by a reflection of \mathbb{R}^n reverses orientation and has degree -1 . Thus, as degrees multiply on composing maps, the antipodal map $x \mapsto -x$ has degree $(-1)^n$. As an application we prove:

Every tangent vector field on an even-dimensional sphere S^{n-1} has a zero.

Proof Assume $v(x)$ is a vector field which is nonzero for all $x \in S^{n-1}$. Then x is perpendicular to $v(x)$, and after rescaling, we may assume that $v(x)$ has length 1. The function $F(x, t) = \cos(t)x + \sin(t)v(x)$ is a well-defined homotopy from the identity map ($t=0$) to the antipodal map ($t=\pi$). But this is impossible as homotopic maps induce the same map in (co)homology and we have already seen that the degree of the identity map is 1 while the degree of the antipodal map is $(-1)^n = -1$ when n is odd.

- It is well known that two self-maps of a sphere of any dimension are homotopic if and only if they have the same degree, that is, $\pi_n(S^n) \cong \mathbb{Z}$ for $n \geq 1$.
- When M is not orientable, $[M]$ still defines a cycle in homology with \mathbb{Z}_2 -coefficients, and $[M] \cap$ defines an isomorphism between the cohomology and homology with \mathbb{Z}_2 coefficients.
- As $[M]$ represents a homology class, so does every other closed (orientable) submanifold of M . It is however not the case that every homology class can be represented by a submanifold or linear combinations of such.

Cohomology is a contravariant functor. Poincaré duality however allows us to define, for any $f: M' \rightarrow M$ between oriented, compact, closed manifolds of arbitrary dimensions, a “transfer” or “Umkehr map,”

$$f^! := D^{-1} f_* D' : H^*(M'; \mathbb{Z}) \rightarrow H^{*-c}(M; \mathbb{Z})$$

which lowers the degree by $c = \dim M' - \dim M$. It satisfies the formula

$$f^!(f^*(x) \cup y) = x \cup f^!(y)$$

for all $x \in H^*(M; \mathbb{Z})$ and $y \in H^*(M'; \mathbb{Z})$. When f is a covering map then $f^!$ can be defined on the chain level by

$$f^!(x)(\sigma) := x \left(\sum_{f(\tilde{\sigma})=\sigma} \tilde{\sigma} \right)$$

where $x \in C^*(M')$ and $\sigma \in C_*(M)$.

de Rham Cohomology

If x_1, \dots, x_n are the local coordinates of \mathbb{R}^n , define an algebra Ω^* to be the algebra generated by symbols dx_1, \dots, dx_n subject to the relations $dx_i dx_j = -dx_j dx_i$ for all i, j . We say $dx_{i_1} \cdots dx_{i_q}$ has degree q . The differential forms on \mathbb{R}^n are the algebra

$$\Omega^*(\mathbb{R}^n) := \{C^\infty \text{ functions on } \mathbb{R}^n\} \otimes_{\mathbb{R}} \Omega^*$$

The algebra $\Omega^*(\mathbb{R}^n) = \bigoplus_{q=0}^n \Omega^q(\mathbb{R}^n)$ is naturally graded by degree. There is a differential operator $d: \Omega^q(\mathbb{R}^n) \rightarrow \Omega^{q+1}(\mathbb{R}^n)$ defined by

1. if $f \in \Omega^0(\mathbb{R}^n)$, then $df = \sum (\partial f / \partial x_i) dx_i$
2. if $\omega = \sum f_I dx_I$, then $d\omega = \sum df_I dx_I$

I stands here for a multi-index. For example, in \mathbb{R}^3 the differential assigns to 0-forms (= functions) the gradient, to 1-forms the curl, and to 2-forms the divergence. An easy exercise shows that $d^2 = 0$ and the q th de Rham cohomology of \mathbb{R}^n is the vector space

$$H_{\text{de R}}^q(\mathbb{R}^n) = \frac{\ker d : \Omega^q(\mathbb{R}^n) \rightarrow \Omega^{q+1}(\mathbb{R}^n)}{\text{im } d : \Omega^{q-1}(\mathbb{R}^n) \rightarrow \Omega^q(\mathbb{R}^n)}$$

More generally, the de Rham complex $\Omega^*(M)$ and its cohomology $H_{\text{de R}}^*(M)$ can be defined for any smooth manifold M .

Let σ be a smooth, singular, real $(q + 1)$ -chain on M , and let $\omega \in \Omega^q(M)$. Stokes theorem then says

$$\int_{\partial\sigma} \omega = \int_{\sigma} d\omega$$

and therefore integration defines a pairing between the q th singular homology and the q th de Rham cohomology of M . This pairing is exact and thus de Rham cohomology is isomorphic to singular cohomology with real coefficients:

$$H_{\text{de R}}^*(M) \simeq (H_*(M; \mathbb{R}))^* \simeq H^*(M; \mathbb{R})$$

Let $\Omega_c^*(M)$ denote the subcomplex of compactly supported forms and $H_c^*(M)$ its cohomology. Integration with respect to the first i coordinates defines a map

$$\Omega_c^*(\mathbb{R}^n) \rightarrow \Omega_c^{*-i}(\mathbb{R}^{n-i})$$

which induces an isomorphism in cohomology; note in particular $H_c^n(\mathbb{R}^n) = \mathbb{R}$. More generally, when $E \rightarrow M$ is an i -dimensional orientable, real vector bundle over a compact, orientable manifold M , integration over the fiber gives the ‘‘Thom isomorphism’’:

$$H_c^*(E) \simeq H_c^{*-i}(M) \simeq H_{\text{de R}}^{*-i}(M)$$

For orientable fiber bundles $F \rightarrow M' \xrightarrow{f} M$ with compact, orientable fiber F , integration over the fiber provides another definition of the transfer map

$$f^! : H_{\text{de R}}^*(M') \rightarrow H_{\text{de R}}^{*-i}(M)$$

Hodge Decomposition

Let M be a compact oriented Riemannian manifold of dimension n . The Hodge star operator, $*$, associates to every q -form an $(n - q)$ -form. For \mathbb{R}^n and any orthonormal basis $\{e_1, \dots, e_n\}$, it is defined by setting

$$*(e_1 \wedge \cdots \wedge e_q) := \pm e_{p+1} \wedge \cdots \wedge e_n$$

where one takes $+$ if the orientation defined by $\{e_1, \dots, e_n\}$ is the same as the given one, and $-$ otherwise. Using local coordinate charts this definition can be extended to M . Clearly, $*$ depends on the chosen metric and orientation of M . If M is compact, we may define an inner product on the q -forms by

$$(\omega, \omega') := \int_M \omega \wedge * \omega'$$

With respect to this inner product $*$ is an isometry. Define the codifferential via

$$\delta := (-1)^{np+n+1} * d * : \Omega^q(M) \rightarrow \Omega^{q-1}(M)$$

and the Laplace–Beltrami operator via

$$\Delta := \delta d + d \delta$$

The codifferential satisfies $\delta^2 = 0$ and is the adjoint of the differential. Indeed, for q -forms ω and $(q + 1)$ -forms ω' :

$$(d\omega, \omega') = (\omega, \delta\omega') \tag{15}$$

It follows easily that Δ is self-adjoint, and furthermore,

$$\Delta\omega = 0 \text{ if and only if } d\omega = 0 \text{ and } \delta\omega = 0 \tag{16}$$

A form ω satisfying $\Delta\omega = 0$ is called ‘‘harmonic.’’ Let \mathcal{H}^q denote the subspace of all harmonic q -forms. It is not hard to prove the ‘‘Hodge decomposition theorem’’:

$$\Omega^q = \mathcal{H}^q \oplus \text{im } d \oplus \text{im } \delta$$

Furthermore, by adjointness [15], a form ω is closed only if it is orthogonal to $\text{im } \delta$. On calculating the de Rham cohomology we can also ignore the summand $\text{im } d$ and find that:

Each de Rham cohomology class on a compact oriented Riemannian manifold M contains a unique harmonic representative, that is, $H_{\text{de R}}^q(M) \simeq \mathcal{H}^q$.

Warning: This is an isomorphism of vector spaces and in general does not extend to an isomorphism of algebras.

Examples

We list the cohomology of some important examples.

Projective Spaces

Let $\mathbb{R}P^n$ be real projective space of dimension n . Then,

$$H^*(\mathbb{R}P^n; \mathbb{Z}_2) = \mathbb{Z}_2[x]/(x^{n+1})$$

is a stunted polynomial ring on a generator x of degree 1.

Similarly, let $\mathbb{C}P^n$ and $\mathbb{H}P^n$ denote complex and quaternionic projective space of real dimensions $2n$ and $4n$, respectively. Then,

$$H^*(\mathbb{C}P^n; \mathbb{Z}) = \mathbb{Z}[y]/(y^{n+1})$$

$$H^*(\mathbb{H}P^n; \mathbb{Z}) = \mathbb{Z}[z]/(z^{n+1})$$

are stunted polynomial rings with $\deg(y) = 2$ and $\deg(z) = 4$.

Lie Groups

Let G be a compact, connected Lie group of rank l , that is, the dimension of the maximal torus of G is l . Then,

$$H^*(G, \mathbb{Q}) \simeq \bigwedge_{\mathbb{Q}}^* [a_{2d_1-1}, a_{2d_2-1}, \dots, a_{2d_l-1}]$$

where $|a_i| = i$ and d_1, \dots, d_l are the fundamental degrees of G which are known for all G . Often this structure lifts to the integral cohomology. In particular we have:

$$H_{\text{free}}^*(SO(2k+1); \mathbb{Z}) \simeq \bigwedge_{\mathbb{Z}}^* [a_3, a_7, \dots, a_{4k-1}]$$

$$H_{\text{free}}^*(SO(2k); \mathbb{Z}) \simeq \bigwedge_{\mathbb{Z}}^* [a_1, a_7, \dots, a_{4k-5}, a_{2k-1}]$$

$$H^*(U(k); \mathbb{Z}) \simeq \bigwedge_{\mathbb{Z}}^* [a_1, a_3, \dots, a_{2k-1}]$$

Classifying Spaces

For any group G there exists a classifying space BG , well defined up to homotopy. Classifying spaces are of central interest to geometers and topologists for the set of isomorphism classes of principal G -bundles over a space X is in one-to-one correspondence with the set of homotopy classes of maps from X to BG . In particular, every cohomology class $c \in H^*(BG; R)$ defines a characteristic class of principle G -bundles E over X : if E corresponds to the map $f_E: X \rightarrow BG$, then $c(E) := f_E^*(c)$.

BG can be constructed as the space of G -orbits of a contractible space EG on which G acts freely. Thus, for example,

$$B\mathbb{Z} = \mathbb{R}/\mathbb{Z} \simeq S^1$$

$$B\mathbb{Z}_2 = (\lim_{n \rightarrow \infty} S^n)/\mathbb{Z}_2 \simeq \mathbb{R}P^\infty$$

$$BS^1 = (\lim_{n \rightarrow \infty} S^{2n+1})/S^1 \simeq \mathbb{C}P^\infty$$

and more generally, infinite Grassmannian manifolds are classifying spaces for linear groups. When G is a compact connected Lie group,

$$H^*(BG; \mathbb{Q}) \simeq \mathbb{Q}[x_{2d_1}, \dots, x_{2d_l}]$$

with d_i as above and $|x_i| = i$. In particular,

$$H^*(BSO(2k+1); \mathbb{Z}[1/2])$$

$$\simeq \mathbb{Z}[1/2][p_1, p_2, \dots, p_k]$$

$$H^*(BSO(2k); \mathbb{Z}[1/2])$$

$$\simeq \mathbb{Z}[1/2][p_1, p_2, \dots, p_{k-1}, e_k]$$

$$H^*(BU(k); \mathbb{Z}) \simeq \mathbb{Z}[c_1, c_2, \dots, c_k]$$

where the Pontryagin, Euler, and Chern classes have degree $|p_i| = 4i$, $|e_k| = 2k$, and $|c_i| = 2i$, respectively.

Moduli Spaces

Let \mathcal{M}_g^n be the space of Riemann surfaces of genus g with n ordered, marked points. There are naturally defined classes κ_i and e_1, \dots, e_n of degree $2i$ and 2 , respectively. By Harer–Ivanov stability and the recent proof of the Mumford conjecture (Madsen–Weiss, preprint 2004), there is an isomorphism up to degree $* < 3g/2$ of the rational cohomology of \mathcal{M}_g^n with

$$\mathbb{Q}[\kappa_1, \kappa_2, \dots] \otimes \mathbb{Q}[e_1, \dots, e_n]$$

The rational cohomology vanishes in degrees $* > 4g - 5$ if $n = 0$, and $* > 4g - 4 + n$ if $n > 0$. Though the stable part of the cohomology is now well understood, the structure of the unstable part, as proposed by Faber (Viehweg 1999), remains conjectural.

Generalized Cohomology Theories

The three basic properties of singular homology appropriately dualized, hold of course also for cohomology. Furthermore, they (essentially) determine (co)homology uniquely as a functor from the category of simplicial spaces and continuous functions to the category of abelian groups. If we drop the dimension axiom (2), we are left with homotopy invariance (1), and the Mayer–Vietoris sequence (3). Abelian group valued functors satisfying (1) and (3) are so called “generalized (co)homology theories.”

K -theory and cobordism theory are two well-known examples but there are many more.

K-Theory

The geometric objects representing elements in complex K -theory $K^0(X)$ are isomorphism classes of finite dimensional complex vector bundles E over X . Vector bundles E, E' can be added to form a new bundle $E \oplus E'$ over X , and $K^0(X)$ is just the group completion of the arising monoid. Thus, for example, for the point space we have $K^0(\text{pt}) = \mathbb{Z}$. Tensor product of vector bundles $E \otimes E'$ induces a multiplication on K -theory making $K^*(X)$ into a graded commutative ring.

In many ways K -theory is easier than cohomology. In particular, the groups are 2-periodic: all even degree groups are isomorphic to the reduced K -theory group $\tilde{K}^0(X) := \text{coker}(K^0(\text{pt}) = \mathbb{Z} \rightarrow K^0(X))$, and all odd degree groups are isomorphic to $K^{-1}(X) := \tilde{K}^0(\Sigma X)$.

The theory of characteristic classes gives a close relation between the two cohomology theories. The Chern character map, a rational polynomial in the Chern classes, defines

$$\begin{aligned} \text{ch} : K^0(X) \otimes_{\mathbb{Z}} \mathbb{Q} &\rightarrow H^{\text{even}}(X; \mathbb{Q}) \\ &:= \bigoplus_{k \geq 0} H^{2k}(X; \mathbb{Q}) \end{aligned}$$

an isomorphism of rings. Thus, the K -theory and cohomology of a space carry the same rational information. But they may have different torsion parts. This became an issue in string theory when D-brane charges which had formerly been thought of as differential forms (and hence cohomology classes) were later reinterpreted more naturally as K -theory classes by [Witten 1998](#))

- There are real and quaternionic K -theory groups which are 8-periodic.

Cobordism Theory

The geometric objects representing an element in the oriented cobordism group $\Omega_{SO}^n(X)$ are pairs (M, f) where M is a smooth, orientable n -dimensional manifold and $f: M \rightarrow X$ is a continuous map. Two pairs (M, f) and (M', f') represent the same cobordism class if there exists a pair (W, F) where W is an $(n+1)$ -dimensional, smooth, oriented manifold with boundary $\partial W = M \cup -M'$ such that $F: W \rightarrow X$ restricts to f and f' on the boundary ∂W . Disjoint union and Cartesian product of manifolds define an addition and multiplication so that $\Omega_{SO}^*(X)$ is a graded, commutative ring.

- Similarly, unoriented, complex, or spin cobordism groups can be defined.

Elliptic Cohomology

Quillen proved that complex cobordism theory is universal for all complex oriented cohomology theories, that is, those cohomology theories that allow a theory of Chern classes. In a complex oriented theory, the first Chern class of the tensor product of two line bundles can be expressed in terms of the first Chern class of each of them via a two-variable power series: $c_1(E \otimes E') = F(c_1(E), c_1(E'))$. F defines a formal group law and Quillen's theorem asserts that the one arising from complex cobordism theory is the universal one.

Vice versa, given a formal group law, one may try to construct a complex oriented cohomology theory from it. In particular, an elliptic curve gives rise to a formal group law and an elliptic cohomology theory. Hopkins *et al.* have described and studied an inverse limit of these elliptic theories, which they call the theory of topological modular forms, tmf , as the theory is closely related to modular forms. In particular, there is a natural map from the groups $\text{tmf}_{2n}(\text{pt})$ to the group of modular forms of weight n over \mathbb{Z} . After inverting a certain element (related to the discriminant), the theory becomes periodic with period $24^2 = 576$.

[Witten \(1998\)](#) showed that the purely theoretically constructed elliptic cohomology theories should play an important role in string theory: the index of the Dirac operator on the free loop space of certain manifolds should be interpreted as an element of it. But unlike for ordinary cohomology, K -theory, and cobordism theory we do not (yet) know a good geometric object representing elements in this theory without which its use for geometry and analysis remains limited. Segal speculated some 20 years ago that conformal field theories should define such geometric objects. Though progress has been made, the search for a good geometric interpretation of elliptic cohomology (and tmf) remains an active and important research area.

Infinite Loop Spaces

Brown's representability theorem implies that for each (reduced) generalized cohomology theory h^* we can find a sequence of spaces E_* such that $h^n(X)$ is the set of homotopy classes $[X, E_n]$ from the space X to E_n for all n . Recall that the Mayer-Vietoris sequence implies that $h^n(X) \simeq h^{n+1}(\Sigma X)$. The suspension functor Σ is adjoint to the based loop space functor Ω which takes a space X to the space of based maps from the circle to X . Hence,

$$\begin{aligned} h^n(X) &= [X, E_n] = [\Sigma X, E_{n+1}] \\ &= [X, \Omega E_{n+1}] \end{aligned}$$

and it follows that every generalized cohomology theory is represented by an infinite loop space

$$E_0 \simeq \Omega E_1 \simeq \cdots \simeq \Omega^n E_n \simeq \cdots$$

Vice versa, any such infinite loop space gives rise to a generalized cohomology theory.

One may think of infinite loop spaces as the abelian groups up to homotopy in the strongest sense. Indeed, ordinary cohomology with integer coefficients is represented by

$$\mathbb{Z} \simeq \Omega S^1 \simeq \Omega^2 CP^\infty \simeq \cdots \simeq \Omega^n K(n, \mathbb{Z}) \simeq \cdots$$

where by definition the Eilenberg–MacLane space $K(n, \mathbb{Z})$ has trivial homotopy groups for all dimensions not equal to n and $\pi_n K(n, \mathbb{Z}) = \mathbb{Z}$. Complex K -theory is represented by

$$\mathbb{Z} \times BU \simeq \Omega(U) \simeq \Omega^2(BU) \simeq \Omega^3(U) \simeq \cdots$$

This is Bott’s celebrated “periodicity theorem.” Finally, oriented cobordism theory is represented by

$$\Omega^\infty MSO := \lim_{n \rightarrow \infty} \Omega^n \text{Th}(\gamma_n)$$

where $\gamma_n \rightarrow BSO_n$ is the universal n -dimensional vector space over the Grassmannian manifold of oriented n -planes in \mathbb{R}^∞ , and $\text{Th}(\gamma_n)$ denotes its Thom space.

A good source of infinite loop spaces are symmetric monoidal categories. Indeed every infinite loop space can be constructed from such a category: the symmetric monoidal structure gives the corresponding homotopy abelian group structure. For

example, the category of finite-dimensional, complex vector spaces and their isomorphisms gives rise to $\mathbb{Z} \times BU$. To give another example, in quantum field theory, one considers the $(d+1)$ -dimensional cobordism category with objects the compact, oriented d -dimensional manifolds, and their $(d+1)$ -dimensional cobordisms as morphisms. Disjoint union of manifolds makes this category into a symmetric monoidal category. The associated infinite loop space and hence generalized cohomology theory has recently been identified as a $(d+1)$ -dimensional slice of oriented cobordism theory (Galatius *et al.* preprint 2005).

See also: Characteristic Classes; Equivariant Cohomology and the Cartan Model; Functional Equations and Integrable Systems; Index Theorems; Intersection Theory; K -Theory; Moduli Spaces: An Introduction; Riemann Surfaces; Spectral Sequences.

Further Reading

- Adams F (1978) Infinite Loop spaces. *Annals of Mathematical Studies* 90: PUP.
- Bott R and Tu L (1982) *Differential Forms in Algebraic Topology*. Springer.
- Galatius, Madsen, Tillmann, Weiss (2005).
- Hatcher A (2002) *Algebraic Topology*, (<http://www.math.cornell.edu>). Cambridge: Cambridge University Press.
- Madsen, Weiss (2004).
- Mosher R and Tangora M (1968) *Cohomology Operations and Applications in Homotopy Theory*. Harper and Row.
- Viehweg (1999) Aspects of Mathematics E33.
- Witten (1998) *Journal of Higher Energy Physics* 12.
- Witten (1998) Springer Lecture Notes in Mathematics, vol. 1326.

Combinatorics: Overview

C Krattenthaler, Universität Wien, Vienna, Austria

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Combinatorics is a vast field which enters particularly in a crucial way in statistical physics. There, it is particularly the enumerative problems that are of importance. Therefore, in this article, we shall mainly concentrate on the enumerative aspects of combinatorics. We first recall the basic terminology, in particular the basic combinatorial objects and numbers, together with the simplest facts about them. We then provide introductions into the most important techniques of enumeration: the generating function

technique, Redfield–Pólya theory, methods of solving functional equations of combinatorial origin, methods of asymptotic enumeration, the theory of heaps, and the transfer matrix method. The subsequent sections then discuss specific problem circles with relation to statistical physics more closely. We discuss lattice path problems, explain Kasteleyn’s method of enumerating perfect matchings and tilings, present the fundamental theorems on nonintersecting paths, and provide an introduction into the research field involving vicious walkers, plane partitions, rhombus tilings, alternating sign matrices, six-vertex configurations, and fully packed loop configurations. Finally, we explain how one should treat binomial and hypergeometric series, which frequently arise in enumeration problems.

Basic Combinatorial Terminology

In this section we review basic combinatorial notions and facts. The reader can find a more detailed treatment and further results, for example, in chapter 1 of [Stanley \(1986\)](#).

The basic combinatorial choice problems and their solutions are: there are 2^n subsets of an n -element set. There are $\binom{n}{k}$ k -element subsets of an n -element set. Given an alphabet $\mathcal{A} = \{a_1, a_2, \dots\}$, a word is a (finite or infinite) sequence of elements of \mathcal{A} . Usually, a finite word is written in the form $w_1 w_2 \dots w_n$ (with $w_i \in \mathcal{A}$). Out of the letters $\{1, 2, \dots, k\}$, one can build k^n words of length n . Out of the letters $\{1, 2, \dots, k\}$, one can build $\binom{n+k-1}{n}$ increasing sequences of length n . The number of permutations of an n -element set is $n!$. The set of permutations of $\{1, 2, \dots, n\}$ is denoted by \mathfrak{S}_n . The number of permutations of an n -element set with exactly k cycles is the Stirling number of the first kind, $s(n, k)$. These numbers are given as the expansion coefficients of falling factorials,

$$x(x-1) \cdots (x-n+1) = \sum_{k=0}^n (-1)^{n-k} s(n, k) x^k$$

or in form of the double (formal) power series

$$\sum_{n, k \geq 0} s(n, k) x^k \frac{y^n}{n!} = (1+y)^x$$

A partition of a set is a collection of pairwise disjoint subsets the union of which is the complete set. The subsets in the collection are called the blocks of the partition. The total number of partitions of an n -element set is the Bell number B_n . These numbers are given by

$$\sum_{n \geq 0} B_n \frac{x^n}{n!} = e^{e^x - 1}$$

The number of partitions of an n -element set into exactly k blocks is the Stirling number of the second kind, $S(n, k)$. These numbers are given by

$$\sum_{n, k \geq 0} S(n, k) x^k \frac{y^n}{n!} = e^{x(e^y - 1)}$$

or, explicitly, by

$$S(n, k) = \frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^n$$

A composition of a positive integer n is a representation of n as a sum $n = s_1 + s_2 + \dots + s_k$ of other positive integers s_i , where the order of the summands matters. The total number of compositions of

n is 2^{n-1} . The number of compositions of n with exactly k summands is $\binom{n-1}{k-1}$. A partition of a positive integer n is a representation of n as a sum $n = \lambda_1 + \lambda_2 + \dots + \lambda_k$ of other positive integers λ_i , where the order of the summands does not matter. Thus, we may assume that the summands are ordered, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$. This is the motivation to write partitions most often in the form of tuples $(\lambda_1, \lambda_2, \dots, \lambda_k)$ the entries of which are weakly decreasing. The summands of a partition are called the parts of the partition. Let $p(n)$ denote the number of partitions of n . These numbers are given by

$$\sum_{n=0}^{\infty} p(n) x^n = \frac{1}{\prod_{i=1}^{\infty} (1-x^i)}$$

If $p(n, k)$ denotes the number of partitions of n into at most k parts, then we have

$$\sum_{n=0}^{\infty} p(n, k) x^n = \frac{1}{\prod_{i=1}^k (1-x^i)}$$

Finally, if $p(n, k, m)$ denotes the number of partitions of n into at most k parts, all of which are at most m , then

$$\begin{aligned} \sum_{n \geq 0} p(n, k, m) x^n &= \frac{(1-x^{k+m})(1-x^{k+m-1}) \cdots (1-x^{m+1})}{(1-x^k)(1-x^{k-1}) \cdots (1-x)} \end{aligned}$$

The expression on the right-hand side is called q -binomial coefficient, and is denoted by $\begin{bmatrix} k+m \\ k \end{bmatrix}_x$.

Partitions are frequently encoded in terms of their Ferrers diagrams. The Ferrers diagram of a partition $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_\ell)$ is an array of cells with ℓ left-justified rows and λ_i cells in row i . For example, the diagram in [Figure 1](#) is the Ferrers diagram of the partition $(3, 3, 2)$.

A lattice path P in \mathbb{Z}^d (where \mathbb{Z} denotes the set of integers) is a path in the d -dimensional integer lattice \mathbb{Z}^d which uses only points of the lattice, that is, it is a sequence (P_0, P_1, \dots, P_l) , where $P_i \in \mathbb{Z}^d$ for all i . The vectors $\overrightarrow{P_0 P_1}, \overrightarrow{P_1 P_2}, \dots, \overrightarrow{P_{l-1} P_l}$ are called the steps of P . The number of steps, l , is called the length of P . [Figure 2](#) shows a lattice path in \mathbb{Z}^2 of length 11.

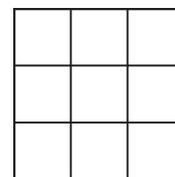


Figure 1 A Ferrers diagram.

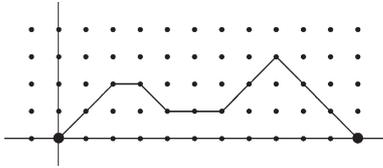


Figure 2 A Motzkin path.

A Dyck path is a lattice path in the integer plane \mathbb{Z}^2 consisting of up-steps $(1, 1)$ and down-steps $(1, -1)$, which starts at the origin, never passes below the x -axis, and ends on the x -axis. See Figure 3 for an example.

The number of Dyck paths of length $2n$ is the Catalan number

$$C_n = \frac{1}{n+1} \binom{2n}{n}$$

The generating function (see the next section for an introduction to the theory of **generating functions**) for these numbers is

$$\sum_{n=0}^{\infty} C_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x} \quad [1]$$

The reader is referred to exercise 6.19 in Stanley (1999) for countless occurrences of the Catalan numbers.

A Motzkin path is a lattice path in the integer plane \mathbb{Z}^2 consisting of up-steps $(1, 1)$, level steps $(1, 0)$, and down-steps $(1, -1)$, which starts at the origin, never passes below the x -axis, and ends on the x -axis. The path in Figure 2 is in fact a Motzkin path. The number of Motzkin paths of length n is the Motzkin number

$$M_n = \sum_{k \geq 0} \frac{1}{k+1} \binom{2k}{k} \binom{n}{2k}$$

The generating function for these numbers is

$$\sum_{n=0}^{\infty} M_n x^n = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2} \quad [2]$$

The reader is referred to exercise 6.38 in Stanley (1999) for numerous occurrences of the Motzkin numbers.

A Schröder path is a lattice path in the integer plane \mathbb{Z}^2 consisting of horizontal steps $(1, 0)$ and

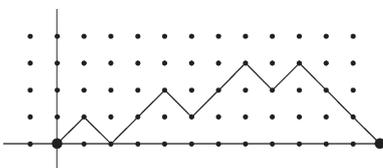


Figure 3 A Dyck path.

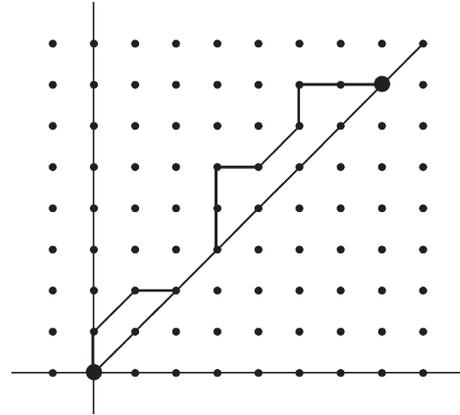


Figure 4 A Schröder path.

vertical steps $(0, 1)$, which starts at the origin, never passes below the diagonal $x = y$, and ends on the diagonal $x = y$. See Figure 4 for an example.

The number of Schröder paths of length n is the (large) Schröder number

$$S_n = \sum_{k \geq 0} \frac{1}{k+1} \binom{2k}{k} \binom{n+k}{2k}$$

The generating function for these numbers is

$$\sum_{n=0}^{\infty} S_n x^n = \frac{1 - x - \sqrt{1 - 6x + x^2}}{2x} \quad [3]$$

The reader is referred to exercise 6.39 in Stanley (1999) for numerous occurrences of the Schröder numbers.

There is another famous sequence of numbers which we did not touch yet, the Fibonacci numbers F_n . They are given by

$$F_n = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{n+1}$$

with generating function

$$\sum_{n=0}^{\infty} F_n x^n = \frac{1}{1 - x - x^2} \quad [4]$$

They also occur in numerous places. For example, the number F_n counts all paths on the integers \mathbb{Z} from 0 to n with steps $(1, 0)$ and $(2, 0)$.

An undirected graph G consists of vertices and edges. An edge is a two-element subset of the vertices, which, however, is thought of as a line or curve connecting the two vertices. See Figure 5a for an example. The usual notation for a graph G is $G = (V, E)$, where V is the set of vertices and E is the set of edges of G . A graph is planar if it is

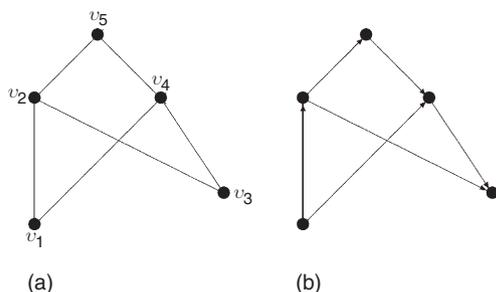


Figure 5 (a) An undirected graph. (b) A directed graph.

embedded in the plane (sphere) in such a way that the curves which mark the edges do not intersect in their interiors. There can be several different ways to embed the same graph in the plane (or in another surface). When we speak of a planar graph then we assume the graph already to be embedded in a given way. For example, the graph in [Figure 5](#) is not a planar graph, by its drawing. However, there is a different embedding which is planar (namely, all embeddings which put the vertex v_3 above the vertex v_5 and leave the other vertices as they are). A tree is a graph without any cycles.

A directed graph (or digraph) G consists of vertices and arcs (which are sometimes also called directed edges). An arc is a pair of vertices, which, however, is thought of an arrow pointing from the first vertex of the pair to the second. See [Figure 5b](#) for an example. The usual notation for a directed graph G is again $G = (V, E)$, where V is the set of vertices and E is the set of arcs of G . All other notions explained for undirected graphs have analogous meanings for directed graphs.

Graphs can be labeled, in which case each vertex is assigned a label, or unlabeled. The (undirected) graph in [Figure 5a](#) is labeled, whereas the (directed) graph in [Figure 5b](#) is unlabeled.

Generating Functions

Generating functions are the very basic tools of enumeration. For introductions to this technique, from different points of view, the reader is referred to [Bergeron et al. \(1998\)](#), [Flajolet and Sedgewick](#) (chapter 1 in the reference listed in “[Further reading](#)” section), and [Stanley \(1998, chapter 1; 1999, chapter 4\)](#).

Let \mathcal{A} be a set of (unlabeled) objects. Each object a in \mathcal{A} has a certain size, $|a|$, which is a non-negative integer. Let us also assume that there is only a finite number of objects from \mathcal{A} of a given size. Let a_n be the number of objects from \mathcal{A} of size n . The

(ordinary) generating function for \mathcal{A} is the formal power series

$$F_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} x^{|a|} = \sum_{n=0}^{\infty} a_n x^n$$

(“formal” means that x is just an indeterminate, not a real or complex number. One can compute with formal power series in the same way as with analytic series, only that convergence issues do not arise, respectively that “convergence” has a different meaning; cf. [Stanley \(1998, section 1.1\)](#)) Typical examples are **Sets** (the collection containing all “unlabeled sets,” that is, all objects of the form $\{\bullet, \bullet, \dots, \bullet\}$, including the empty set, where the size of $\{\bullet, \bullet, \dots, \bullet\}$ is the number of \bullet ’s), **Sequences** (the collection containing all “unlabeled sequences,” that is, all objects of the form $(\bullet, \bullet, \dots, \bullet)$, including the empty sequence), **Cycles** (unlabeled cycles), with respective generating function

$$F_{\text{Sets}}(x) = F_{\text{Sequences}}(x) = \frac{1}{1-x} \quad [5]$$

$$F_{\text{Cycles}}(x) = \frac{x}{1-x}$$

or **Trees** (unlabeled trees).

If \mathcal{A} and \mathcal{B} are two sets of objects, one can define several other sets of objects using them. The union of \mathcal{A} and \mathcal{B} , written $\mathcal{A} \cup \mathcal{B}$, has as a groundset the disjoint union of \mathcal{A} and \mathcal{B} , and the size of an element from \mathcal{A} is its size in \mathcal{A} , while the size of an element from \mathcal{B} is its size in \mathcal{B} . We have

$$F_{\mathcal{A} \cup \mathcal{B}}(x) = F_{\mathcal{A}}(x) + F_{\mathcal{B}}(x) \quad [6]$$

The product of \mathcal{A} and \mathcal{B} , written $\mathcal{A} \times \mathcal{B}$, has as a groundset the set of pairs $\mathcal{A} \times \mathcal{B}$, and the size of an element (a, b) from $\mathcal{A} \times \mathcal{B}$ is the sum of the sizes of a (in \mathcal{A}) and of b (in \mathcal{B}). We have

$$F_{\mathcal{A} \times \mathcal{B}}(x) = F_{\mathcal{A}}(x) \cdot F_{\mathcal{B}}(x) \quad [7]$$

The substitution of two sets \mathcal{A} and \mathcal{B} of objects can only be defined in certain circumstances, and only in certain more restrictive circumstances the generating function for the substitution can be computed by substituting the generating functions for \mathcal{A} and \mathcal{B} . Let us assume that any object a from \mathcal{A} of size n , by its structure, has n atoms (nodes). For example, if \mathcal{A} is a certain set of trees, where the size of a tree is the number of leaves in the tree, then we may take, as the atoms, the leaves of the tree. In this situation, the substitution of \mathcal{B} in \mathcal{A} , denoted by $\mathcal{A}(\mathcal{B})$, is the set of objects which arises by replacing the atoms of objects from \mathcal{A} by objects from \mathcal{B} in all possible ways. The size of an object from $\mathcal{A}(\mathcal{B})$ is the sum of the sizes of the objects from \mathcal{B} that it

contains. In order that $\mathcal{A}(\mathcal{B})$ contains only a finite number of objects of a given size, we must assume that \mathcal{B} contains no elements of size 0. If, in addition, the atoms of any element a from \mathcal{A} inherit an order (e.g., if \mathcal{A} is a set of binary trees, then the leaves of a binary tree are ordered in a natural way from “left” to “right”), then we have

$$F_{\mathcal{A}(\mathcal{B})}(x) = F_{\mathcal{A}}(F_{\mathcal{B}}(x)) \quad [8]$$

However, this equation is not true in general. The general formula comes out of Redfield–Pólya theory (see [21] and [24]) and requires the notion of cycle index series. For example, if \mathcal{B} is the set of connected (unlabeled) graphs, \mathcal{A} is **Sets**, so that $\mathcal{A}(\mathcal{B})$ is the set of all (connected and disconnected) graphs, then [8] is not true, but what is true is

$$F_{\mathbf{Sets}(\mathcal{B})} = \exp(F_{\mathcal{B}}(x) + \frac{1}{2}F_{\mathcal{B}}(x^2) + \frac{1}{3}F_{\mathcal{B}}(x^3) + \dots) \quad [9]$$

This holds, in fact, for any set \mathcal{B} of unlabeled objects. (This is seen by combining [24], [17], and [21].)

Next we deal with the enumeration of labeled objects. Let \mathcal{A} be a set of labeled objects, again, each object a with a certain size $|a|$ which is a non-negative integer. “Labeled” means that each object of size n , by its structure, comes with n atoms (nodes) which are labeled $1, 2, \dots, n$. For example, \mathcal{A} may be the set of all labeled graphs, where the size of a graph is the number of its vertices, and where the vertices are labeled $1, 2, \dots, n$. Again, we assume that there is only a finite number of objects from \mathcal{A} of a given size. Let a_n be the number of objects from \mathcal{A} of size n . The exponential generating function for \mathcal{A} is the formal power series

$$E_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} \frac{x^{|a|}}{|a|!} = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$$

Typical examples are **Sets** (the collection containing all “labeled sets,” that is all objects of the form $\{1, 2, \dots, n\}$, including the empty set), **Permutations**, **Cycles** (labeled cycles), with respective generating functions

$$E_{\mathbf{Sets}}(x) = \exp(x) \quad [10]$$

$$E_{\mathbf{Permutations}}(x) = \frac{1}{1-x} \quad [11]$$

$$E_{\mathbf{Cycles}}(x) = \log \frac{1}{1-x} \quad [12]$$

or **Trees** (labeled trees). The explicit form of the generating function for **Trees** is discussed in the section “Solving equations for generating functions: the Lagrange inversion formula and the kernel method.”

If \mathcal{A} and \mathcal{B} are two sets of objects, one defines again several other sets of objects using them. The union of \mathcal{A} and \mathcal{B} , written $\mathcal{A} \cup \mathcal{B}$, has as a groundset the disjoint union of \mathcal{A} and \mathcal{B} , and the size of an element from \mathcal{A} is its size in \mathcal{A} , while the size of an element from \mathcal{B} is its size in \mathcal{B} . We have

$$E_{\mathcal{A} \cup \mathcal{B}}(x) = E_{\mathcal{A}}(x) + E_{\mathcal{B}}(x) \quad [13]$$

To define the product of \mathcal{A} and \mathcal{B} , written $\mathcal{A} \times \mathcal{B}$, we cannot simply take $\mathcal{A} \times \mathcal{B}$ as a groundset, we must also say something about the labeling of the objects. So, as a groundset we take all pairs (a, b) with $a \in \mathcal{A}$ and $b \in \mathcal{B}$, but labeled in all possible ways by $1, 2, \dots, |a| + |b|$ such that the order of labels assigned to a respects the original order of labels of a , and the same for b . The size of such an element (a, b) is again the sum of the sizes of a (in \mathcal{A}) and of b (in \mathcal{B}). We have

$$E_{\mathcal{A} \times \mathcal{B}}(x) = E_{\mathcal{A}}(x) \cdot E_{\mathcal{B}}(x) \quad [14]$$

Since, in the labeled world, objects come automatically with atoms, the substitution of two sets \mathcal{A} and \mathcal{B} of objects can now always be defined. The substitution of \mathcal{B} in \mathcal{A} , denoted by $\mathcal{A}(\mathcal{B})$, is the set of objects which arises by replacing the atoms of objects from \mathcal{A} by objects from \mathcal{B} in all possible ways, and labeling the substituted objects in all possible ways by $1, 2, \dots, \sum_b |b|$ (the sum being over the objects from \mathcal{B} which were put in the places of the atoms) that are consistent with the original labelings of the objects from \mathcal{B} . The size of an object from $\mathcal{A}(\mathcal{B})$ is the sum of the sizes of the objects from \mathcal{B} that it contains. In order that $\mathcal{A}(\mathcal{B})$ contains only a finite number of objects of a given size, we must assume that \mathcal{B} contains no elements of size 0. Then we have

$$E_{\mathcal{A}(\mathcal{B})}(x) = E_{\mathcal{A}}(E_{\mathcal{B}}(x)) \quad [15]$$

An example of a composition is

$$\mathbf{Permutations} = \mathbf{Sets}(\mathbf{Cycles})$$

Thus, from [15] we have

$$E_{\mathbf{Permutations}}(x) = E_{\mathbf{Sets}}(E_{\mathbf{Cycles}}(x))$$

corresponding to the identity

$$\frac{1}{1-x} = \exp(\log 1/(1-x))$$

Another manifestation of the composition rule is, for example, the fact (which is sometimes called the “exponential principle”) that, if one takes the log of the partition function for some maps, the result is the partition function for the connected maps among them.

All of the above can be generalized to a weighted setting. Namely, if \mathcal{A} is a set of objects (labeled or unlabeled), and if $w: \mathcal{A} \rightarrow R$ is a weight function from \mathcal{A} into some ring R , then all of the above remains true, if we replace the definitions of $F_{\mathcal{A}}(x)$ and $E_{\mathcal{A}}(x)$ above by the weighted sums

$$F_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} w(a)x^{|a|}$$

and

$$E_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} w(a) \frac{x^{|a|}}{|a|!}$$

respectively, if in the definition of the union of \mathcal{A} and \mathcal{B} we define the weight of an object to be its weight in \mathcal{A} , respectively \mathcal{B} , if in the definition of the product of \mathcal{A} and \mathcal{B} we define the weight of an object (a, b) to be the product of the weights of a and b , and if in the definition of the substitution we define the weight of an object in $\mathcal{A}(\mathcal{B})$ as the product of the weights of the objects from \mathcal{B} that were put in place of the atoms.

Redfield–Pólya Theory of Colored Enumeration

The natural and uniform environment for the separate treatment of generating functions for unlabeled and labeled objects in the last section is the theory for counting colored objects founded by Redfield and Pólya, in the modern treatment through cycle index series due to Joyal. We refer the reader to Bergeron *et al.* (1998, appendix 1), de Bruijn (1981), and Stanley (1999, chapter 7) for further reading.

Let \mathcal{A} be a set of labeled objects with the constraint that there is only a finite number of objects of a given size. The cycle index series for \mathcal{A} is the formal multivariable series

$$Z_{\mathcal{A}}(x_1, x_2, \dots) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \text{fix}_{\sigma}(\mathcal{A}) x_1^{c_1(\sigma)} x_2^{c_2(\sigma)} x_3^{c_3(\sigma)} \dots \quad [16]$$

where $\text{fix}_{\sigma}(\mathcal{A})$ is the number of objects a from \mathcal{A} that remain invariant when the labels are permuted according to the permutation σ (in particular, if $\sigma \in \mathfrak{S}_n$, the size of a must be n in order that σ can be applied to the labels), and where $c_i(\sigma)$ denotes the number of cycles of length i of σ .

In most cases, it is difficult to obtain compact expressions for the cycle index series. However, for

our familiar families of objects, compact expressions are available:

$$Z_{\text{Sets}}(x_1, x_2, \dots) = \exp\left(x_1 + \frac{x_2}{2} + \frac{x_3}{3} + \dots\right) \quad [17]$$

$$Z_{\text{Permutations}}(x_1, x_2, \dots) = \prod_{i=1}^{\infty} \frac{1}{1 - x_i} \quad [18]$$

$$Z_{\text{Cycles}}(x_1, x_2, \dots) = \sum_{i=1}^{\infty} \frac{\Phi(i)}{i} \log \frac{1}{1 - x_i} \quad [19]$$

where $\Phi(i)$ is the Euler totient function (the number of positive integers $j \leq i$ relatively prime to i).

What makes the cycle index series so fundamental is the fact that the generating functions from the last section are specializations of it. Namely, the exponential generating function for \mathcal{A} is equal to

$$E_{\mathcal{A}}(x) = Z_{\mathcal{A}}(x, 0, 0, \dots) \quad [20]$$

If, given the set of labeled objects \mathcal{A} , we produce a set of unlabeled objects $\tilde{\mathcal{A}}$ by taking all the objects from \mathcal{A} but forgetting the labels, then the ordinary generating function for $\tilde{\mathcal{A}}$ is another specialization of the cycle index series,

$$F_{\tilde{\mathcal{A}}}(x) = Z_{\mathcal{A}}(x, x^2, x^3, \dots) \quad [21]$$

The cycle index series satisfies the following properties with respect to union, product and composition of sets of objects:

$$Z_{\mathcal{A} \cup \mathcal{B}}(x_1, x_2, \dots) = Z_{\mathcal{A}}(x_1, x_2, \dots) + Z_{\mathcal{B}}(x_1, x_2, \dots) \quad [22]$$

$$Z_{\mathcal{A} \times \mathcal{B}}(x_1, x_2, \dots) = Z_{\mathcal{A}}(x_1, x_2, \dots) \times Z_{\mathcal{B}}(x_1, x_2, \dots) \quad [23]$$

$$Z_{\mathcal{A}(\mathcal{B})}(x_1, x_2, \dots) = Z_{\mathcal{A}}(Z_{\mathcal{B}}(x_1, x_2, x_3, \dots), Z_{\mathcal{B}}(x_2, x_4, x_6, \dots), Z_{\mathcal{B}}(x_3, x_6, x_9, \dots), \dots) \quad [24]$$

Similar to the theory of generating functions surveyed in the last section, one can also develop a weighted version of the cycle index series. Given a set of labeled objects \mathcal{A} , where each object a is assigned a weight $w(a)$, one changes the definition [16] insofar as $\text{fix}_{\sigma}(\mathcal{A})$ gets replaced by the weighted sum $\sum_{\sigma(a)=a} w(a)$, where $\sigma(a)$ means the object arising from a by permuting the labels according to σ . Then all the above formulas remain true in this weighted setting.

Cycle index series are instrumental in the enumeration of colored objects. The basic situation is that we have given a set $\tilde{\mathcal{A}}$ of unlabeled objects so that every object of size n comes with n atoms (nodes). For example, we may think of $\tilde{\mathcal{A}}$ as the set of cycles. We are now going to color each atom by a

color from the set of colors C . The question that we pose is: *how many different colored objects of a given size are there?* In our example, if C consists of the two colors “black” and “white,” then we are asking the question of how many necklaces one can make out of n pearls that can be black or white. In terms of generating functions, we want to compute

$$\Gamma_{\tilde{\mathcal{A}}}(x) = \sum_c x^{|c|}$$

where the sum is over all colored objects c that one can obtain by coloring the objects from $\tilde{\mathcal{A}}$.

The central result of Redfield–Pólya theory is that, if \mathcal{A} is the set of labeled objects that one obtains from $\tilde{\mathcal{A}}$ by labeling the objects of $\tilde{\mathcal{A}}$ in all possible ways, then

$$\Gamma_{\tilde{\mathcal{A}}}(x) = Z_{\mathcal{A}}(|C|x, |C|x^2, |C|x^3, \dots)$$

There is again a weighted version. One allows the objects a from $\tilde{\mathcal{A}}$ to have weight $w(a) \in R$. Moreover, one assumes a weight function $f : C \rightarrow R$ on the colors with values in the ring R . One defines the weight of a colored object obtained by coloring the atoms of a to be $w(a)$ multiplied by the product of all $f(\gamma)$, where γ ranges over all the colors of the atoms (including repetitions of colors). Let $\Gamma_{\tilde{\mathcal{A}}}(w, f)$ denote the sum of all the weights of all colored objects obtained from $\tilde{\mathcal{A}}$. Then

$$\Gamma_{\tilde{\mathcal{A}}}(w, f) = Z_{\mathcal{A}}\left(\sum_{c \in C} f(c), \sum_{c \in C} f(c)^2, \sum_{c \in C} f(c)^3, \dots\right)$$

We remark that these results cover also the case of enumeration of objects under a group action. This includes the enumeration of objects on which we impose certain symmetries. See Bergeron *et al.* (1998, appendix 1), de Bruijn (1981), and Stanley (1999, chapter 7) for more details. The enumeration of asymmetric objects is the subject of an ongoing research program (cf. Labelle and Lamathe (2004)).

Solving Equations for Generating Functions: The Lagrange Inversion Formula and the Kernel Method

In this section, we describe two methods to solve functional equations for generating functions. The Lagrange inversion makes it possible (in some situations) to find explicit expressions for the coefficients of an implicitly given series. The kernel method (and its extensions), on the other hand, is a powerful method to obtain an explicit expression for an implicitly given function. We refer the reader to Flajolet and

Sedgewick, (section VII.5 of the reference in “Further reading” section) for further reading.

In many situations it will happen that, when we apply the methods from the last section, we end up with a functional equation for the generating function $f(x) = \sum_{n=0}^{\infty} f_n x^n$ that we wanted to compute. For example, if t_n denotes the number of labeled rooted trees with n nodes, and if we write $T(x) = \sum_{n=1}^{\infty} t_n x^n / n!$, then, by applying a straightforward decomposition of a tree into its root and its set of subtrees attached to the root, we obtain the equation

$$T(z) = z \exp(T(z)) \tag{25}$$

How does one solve such an equation? As a matter of fact, for $T(z)$, there is no expression in terms of known functions. However, the Lagrange inversion formula enables one to find the coefficients $t_n/n!$ of $T(z)$ explicitly. The theorem reads as follows.

Theorem *Let $g(x)$ be a formal Laurent series containing only a finite number of negative powers of x , and let $f(x)$ be a formal power series without constant term. If we expand $g(x)$ in powers of $f(x)$,*

$$g(x) = \sum_k c_k f^k(x) \tag{26}$$

then the coefficients c_n are given by

$$c_n = \frac{1}{n} [x^{-1}] g'(x) f^{-n}(x) \quad \text{for } n \neq 0 \tag{27}$$

or, alternatively, by

$$c_n = [x^{-1}] g(x) f'(x) f^{-n-1}(x) \tag{28}$$

Here, $[x^n]h(x)$ denotes the coefficient of x^n in the power series $h(x)$.

With this theorem in hand, eqn [25] is easy to solve. We write it in the form

$$T(x) \exp(-T(x)) = x \tag{29}$$

We want to know the coefficients in the expansion $T(x) = \sum_{n=0}^{\infty} t_n x^n / n!$. Since, by [29], $T(x)$ is the compositional inverse of $x \exp(-x)$, substitution of $x \exp(-x)$ instead of x gives

$$x = \sum_{n=0}^{\infty} \frac{t_n}{n!} (x \exp(-x))^n$$

This equation is in the form [26] with $f(x) = x \exp(-x)$ and $g(x) = x$. Hence, by [27], we obtain

$$\begin{aligned} \frac{t_n}{n!} &= \frac{1}{n} [x^{-1}] (x \exp(-x))^{-n} \\ &= \frac{1}{n} [x^{n-1}] \exp(nx) = \frac{n^{n-1}}{n!} \end{aligned}$$

and, thus, $t_n = n^{n-1}$.

The second method to solve functional equations which we explain in this section is the kernel method. We illustrate the method by an example. Let us consider the problem of counting Dyck paths of length $2n$ (see the section “[Basic combinatorial terminology](#)”). Rather than attempting to arrive at a solution of the problem directly, we consider the more general problem of counting the number $a_{n,k}$ of paths consisting of steps $(1, 1)$ and $(1, -1)$, which start at the origin, never drop below $y=0$, have length n , and end at height k . We then form the bivariate generating function $F(u, x) = \sum_{n,k \geq 0} a_{n,k} x^n u^k$. We then have the functional equation

$$F(u, x) = 1 + xuF(u, x) + \frac{x}{u}(F(u, x) - F(0, x)) \quad [30]$$

since a path can be empty (this explains the term 1), it can end by a step $(1,1)$ (this explains the term $xuF(u)$), or it can end by a step $(1,-1)$. The latter can only happen if the path before that last step did not end at height 0. The generating function for these paths is $F(u, x) - F(0, x)$, and this explains the third term in the eqn [30]. In fact, we may replace [30] by

$$F(u, x) = 1 + xuF(u, x) + \frac{x}{u}(F(u, x) - F_1(x)) \quad [31]$$

because [31] implies that $F_1(x) = F(0, x)$.

The idea of the kernel method is to get rid of the unknown series $F(u, x)$. This is possible because $F(u, x)$ occurs linearly in [31], which can be rewritten as

$$F(u, x) \left(1 - xu - \frac{x}{u}\right) = 1 - \frac{x}{u} F_1(x) \quad [32]$$

We simply equate the coefficient of $F(u, x)$ in this equation to zero,

$$1 - xu - \frac{x}{u} = 0$$

solve this for u ,

$$u = \frac{1 - \sqrt{1 - 4x^2}}{2x}$$

(the other solution for u makes no sense in [31]), and substitute this back in [32], to obtain

$$F_1(x) = \frac{1 - \sqrt{1 - 4x^2}}{2x^2}$$

the familiar generating function [2] for the Catalan numbers. Now, by substituting this result in [31], we can even compute the full series $F(u, x)$.

While this was certainly a complicated, and unusual, way to compute the Catalan numbers, this approach generalizes when one considers paths with different step sets (see section VII.5 of the Flajolet and Sedgewick reference in “[Further](#)

[reading](#)” section). In a more general situation, one has a functional equation

$$P(F(u, x), F_1(x), \dots, F_k(x), x, u) = 0 \quad [33]$$

where $F(u, x)$ appears linearly, as well as the unknown series $F_1(x), \dots, F_k(x)$, whereas x and u appear rationally. It is clear that one can apply the same technique, namely collecting all the terms involving $F(u, x)$, equating the coefficient of $F(u, x)$ to zero, solving for u and substituting back in [33]. If there is more than one function $F_i(x)$, then this will only give one equation for $F_i(x)$. However, when equating the coefficient of $F(u, x)$, which was a polynomial equation, there can be more solutions. (That was actually also the case in our example, although only one solution could be used.) All these solutions can be substituted in [33] to give many more equations for $F_i(x)$. The kernel method will work if we have enough equations to determine the unknown functions $F_i(x)$ (see the Flajolet and Sedgewick reference, section VII.5 for further details). In the variant of the “obstinate kernel method,” more equations are produced in more sophisticated ways. The method has been largely extended by Bousquet-Mélou and co-workers to cover equations of the form [33], where P is a polynomial such that eqn [33] determines all involved series uniquely. This extension covers in particular the so-called quadratic method due to Brown, which is of great significance in the work of Tutte on the enumeration of maps. We refer the reader to [Bousquet-Mélou and Jehanne \(2005\)](#) and the references given there for these extensions.

Extracting Asymptotic Information from Generating Functions

There is powerful machinery available to extract the asymptotic behavior of the coefficients of a power series out of analytic properties of the power series. We describe the corresponding methods, singularity analysis and the saddle point method in this section. The survey by [Odlyzko \(1995\)](#) and the Flajolet and Sedgewick reference in “[Further reading](#)” are excellent sources for [further reading](#), which, in particular, contain several other methods which we cannot cover here for reasons of limited space.

Let us suppose that we are interested in the asymptotic behavior of the sequence $(f_n)_{n \geq 0}$ of real (or complex) numbers as n tends to infinity. Let us suppose that the power series $f(z) = \sum_{n=0}^{\infty} f_n z^n$ converges in some neighborhood of the origin. (If this series converges only at $z=0$, then either one has to try to scale, that is, for example, look at the

power series $f(z) = \sum_{n=0}^{\infty} f_n z^n / n!$ instead, or one must apply methods other than singularity analysis or the saddle point method. In the latter case, depending on the nature of the coefficients f_n , this may be the Euler–Maclaurin or the Poisson summation formulas, the Mellin transform technique, or other direct methods. The reader is referred to Odlyzko (1995) and the Flajolet and Sedgewick reference.) The idea is then to consider $f(z)$ as a complex function in z (and extend the range of f beyond the disk of convergence about the origin), and to study the singularities of $f(z)$. (The point at infinity can also be a singularity.) The upshot is that the singularities of $f(z)$ with smallest modulus dictate the asymptotic behavior of the coefficients f_n . These singularities of smallest modulus are called the dominating singularities.

If there is an infinite number of dominant singularities, then one has to try the circle method. We refer the reader to Andrews (1976) and Ayoub (1963) for details of this method.

If there is a finite number of dominant singularities, then there can be again two different situations, depending on whether these are “small” or “large” singularities. Roughly speaking, a singularity is small if the function $f(z)$ grows at most polynomially when z approaches the singularity, otherwise it is “large.” A typical example of a small singularity is $z=1/4$ in $(1-4z)^{-1/2}$, whereas a typical example of a large singularity is $z=\infty$ in $\exp(x)$ or $z=1$ in $\exp(1/(1-z))$.

The method to apply for small singularities is the method of singularity analysis as developed by Flajolet and Odlyzko. (Singularity analysis implies Darboux’s method, which occurs frequently in the literature, and, thus, supersedes it.) For the sake of simplicity, we consider first only the case of a unique dominant singularity. We shall address the issue of several dominant singularities shortly. Furthermore, we assume the singularity to be $z=1$, again for the sake of simplicity of presentation. The general result can then be obtained by rescaling z .

The basic idea is the transfer principle:

$$\begin{aligned} \text{If } f(z) &= \sigma(z) + O(\tau(z)) \quad \text{then} \\ f_n &= \sigma_n + O(\tau_n) \end{aligned} \quad [34]$$

where $\sigma(z) = \sum_{n=0}^{\infty} \sigma_n z^n$ is a linear combination of standard functions of the form $(1-z)^{-\alpha}$, or logarithmic variants, and $\tau(z) = \sum_{n=0}^{\infty} \tau_n z^n$ also lies in the scale (see sections VI.3,4 of the Flajolet and Sedgewick reference for the exact statement). The expansion for $f(z)$ in [34] is called the singular

expansion of $f(z)$. For the above-mentioned standard functions, we have

$$\begin{aligned} [z^n](1-z)^{-\alpha} \left(\frac{1}{z} \log \frac{1}{1-z} \right)^\beta \\ \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^\beta \left(1 + \frac{C_1}{1!} \frac{\beta}{\log n} \right. \\ \left. + \frac{C_2}{2!} \frac{\beta(\beta-1)}{(\log n)^2} + \dots \right) \end{aligned} \quad [35]$$

where $[z^n]g(z)$ denotes the coefficient of z^n in $g(z)$, and where

$$C_k = \Gamma(\alpha) \frac{d^k}{ds^k} \frac{1}{\Gamma(s)} \Big|_{s=\alpha}$$

If α is a nonpositive integer, then this expansion has to be taken with care (cf. section VI.2 of the Flajolet and Sedgewick reference).

To see how this works, consider the example $f_n = \sum_{k=0}^n \binom{2k}{k}$. We have

$$\sum_{n=0}^{\infty} f_n z^n = \frac{1}{(1-z)\sqrt{1-4z}}$$

The function on the right-hand side is meromorphic in all of \mathbb{C} (where \mathbb{C} denotes the complex numbers), with singularities at $z=1$ and $z=1/4$. The dominant singularity is $z=1/4$. We determine the singular expansion of $f(z)$ about $z=1/4$,

$$\begin{aligned} f(z) &= \frac{4}{3}(1-4z)^{-1/2} - \frac{4}{9}(1-4z)^{1/2} \\ &\quad + \frac{4}{27}(1-4z)^{3/2} + O((1-4z)^{5/2}) \end{aligned}$$

(We stopped the expansion after three terms. The farther we go, the more terms can we compute of the asymptotic expansion for f_n .) Hence, we obtain

$$\begin{aligned} f_n &= 4^n \left(\frac{4}{3} \frac{n^{-1/2}}{\Gamma(1/2)} \left(1 - \frac{1}{8n} + \frac{1}{128n^2} \right) \right. \\ &\quad - \frac{4}{9} \frac{n^{-3/2}}{\Gamma(-1/2)} \left(1 + \frac{3}{8n} \right) \\ &\quad \left. + \frac{4}{27} \frac{n^{-5/2}}{\Gamma(-3/2)} + O(n^{-7/2}) \right) \\ &= \frac{4^n}{\sqrt{\pi n}} \left(\frac{4}{3} + \frac{1}{18n} + \frac{11}{288n^2} + O\left(\frac{1}{n^3}\right) \right) \end{aligned}$$

If there are several small dominant singularities (but only a finite number of them), then one simply applies the above procedure for all of them and, to obtain the desired asymptotic expansion, one adds up the corresponding contributions.

The method to apply for large singularities is the saddle point method. For the following considerations, we assume that $f(z)$ is analytic in $|z| < R \leq \infty$. At the heart of the saddle point method lies Cauchy's formula

$$f_n = [z^n]f(z) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f(z)}{z^{n+1}} dz \quad [36]$$

for writing the n th coefficient in the power series expansion of $f(z)$. Here, \mathcal{C} is some simple closed contour around the origin that stays in the range $|z| < R$. The idea is to exploit the fact that we are free to deform the contour. The aim is to choose a contour such that the main contribution to the integral in [36] comes from a very tiny part of the contour, whereas the contribution of the rest is negligible. This will be possible if we put the contour through a saddle point of the integrand $f(z)/z^{n+1}$. Under suitable conditions, the main contribution will then come from the small passage of the path through the saddle point, and the contribution of the rest will be negligible.

In practice, the saddle point method is not always straightforward to apply, but has to be adapted to the specific properties of the function $f(z)$ that we are encountering. We refer the reader to the corresponding chapters in the Flajolet and Sedgewick reference and Odlyzko (1995) for more details. There is one important exception though, namely the Hayman admissible functions. We will not reproduce the definition of Hayman admissibility because it is cumbersome (cf. section VIII.5 in the Flajolet and Sedgewick reference and definition 12.4 of Odlyzko (1995)). However, in many applications, it is not even necessary to go back to it because of the closure properties of Hayman admissible functions. Namely, it is known (cf. Odlyzko (1995), theorem 12.8) that $\exp(p(z))$ is Hayman admissible in $|z| < \infty$ for any polynomial $p(z)$ with real coefficients as long as the coefficients a_n of the Taylor series of $\exp(p(z))$ are positive for all sufficiently large n (thus, e.g., $\exp(z)$ is Hayman admissible), and it is known that, if $f(z)$ and $g(z)$ are Hayman admissible in $|z| < R \leq \infty$, then $\exp(f(z))$ and $f(z)g(z)$ are also (thus, e.g., $\exp(\exp(z) - 1)$ is Hayman admissible).

The central result of Hayman's theory is the following: if $f(z) = \sum_{n \geq 0} f_n z^n$ is Hayman admissible in $|z| < R$, then

$$f_n \sim \frac{f(r_n)}{r_n^n \sqrt{2\pi b(r_n)}} \quad \text{as } n \rightarrow \infty \quad [37]$$

where r_n is the unique solution for large n of the equation $a(r) = n$ in (R_0, R) , with $a(r) = rf'(r)/f(r)$, $b(r) = ra'(r)$, and a suitably chosen constant $R_0 > 0$.

This result covers only the first term in the asymptotic expansion. There is an even more sophisticated theory due to Harris and Schoenfeld, which allows one to also find a complete asymptotic expansion. We refer the reader to section VIII.5 of the Flajolet and Sedgewick reference and Odlyzko (1995) for more details.

Methods for the asymptotic analysis of multi-variable generating functions are also available (see the corresponding chapters in Flajolet and Sedgewick, Odlyzko (1995) and the recent important development surveyed in the Pemantle and Wilson reference listed in "Further reading"). We add that both the method of singularity analysis and Hayman's theory of admissible functions have been made largely automatic, and that this has been implemented in the Maple program `gdev` (see "Further reading").

The Theory of Heaps

The theory of heaps, developed by Viennot, is a geometric rendering of the theory of the partial commutation monoid of Cartier and Foata, which is now most often called the Cartier–Foata monoid. Its importance stems from the fact that several objects which appear in statistical physics, such as Motzkin paths, animals, respectively polyominoes, or Lorentzian triangulations (see the Viennot and James reference in "Further reading" and the references therein), are in bijection with heaps.

Informally, a heap is what we would imagine. We take a collection of "pieces," say B_1, B_2, \dots , and put them one upon the other, sometimes also sideways, to form a "heap," see Figure 6.

There, we imagine that pieces can only move vertically, so that the heap in Figure 6 would indeed form a stable arrangement. Note that we allow several copies of a piece to appear in a heap. (This means that they differ only by a vertical translation.) For example, in Figure 6 there appear two copies of B_2 . Under these assumptions, there are pieces which can move past each other, and others which cannot. For example, in Figure 6, we can move the piece B_6 higher up, thus moving it higher than B_1 if we wish. However, we cannot move B_7 higher than B_6 ,

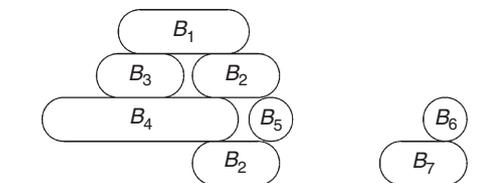


Figure 6 A heap of pieces.

because B_6 blocks the way. On the other hand, we can move B_7 past B_1 (thus taking B_6 with us). Thus, a rigorous way to introduce heaps is by beginning with a set \mathcal{B} of pieces (in our example, $\mathcal{B} = \{B_1, B_2, \dots, B_7\}$), and we declare which pieces can be moved past another and which cannot. We indicate this by a symmetric relation \mathcal{R} : we write $a\mathcal{R}b$ to indicate that a cannot move past b (and vice versa). When we consider a word $a_1a_2\dots a_n$ of pieces, $a_i \in \mathcal{B}$, we think of it as putting first a_1 , then putting a_2 on top of it (and, possibly, moving it past a_1), then putting a_3 on top of what we already have, etc. We declare two words to be equivalent if one arises from the other by commuting adjacent letters which are not in relation. A heap is then an equivalence class of words under this equivalence relation. What we have described just now is indeed the original definition of Cartier and Foata.

The class of heaps which occurs most frequently in applications is the class of heaps of monomers and dimers, which we now introduce. Let $\mathcal{B} = M \cup D$, where $\mathcal{M} = \{m_0, m_1, \dots\}$ is the set of monomers and $\mathcal{D} = \{d_1, d_2, \dots\}$ is the set of dimers. We think of a monomer m_i as a point, symbolized by a circle, with x -coordinate i , see Figure 7. We think of a dimer d_i as two points, symbolized by circles, with x -coordinates $i - 1$ and i which are connected by an edge, see Figure 7. We impose the relations $m_i\mathcal{R}m_i, m_i\mathcal{R}d_i, m_i\mathcal{R}d_{i+1}, i = 0, 1, \dots, d_i\mathcal{R}d_i, i - 1 \leq j \leq i$, and extend \mathcal{R} to a symmetric relation. Figure 8 shows two heaps of monomers and dimers.

For example, Motzkin paths are in bijection with heaps of monomers and dimers. To see this, given a Motzkin path, we read the steps of the path from

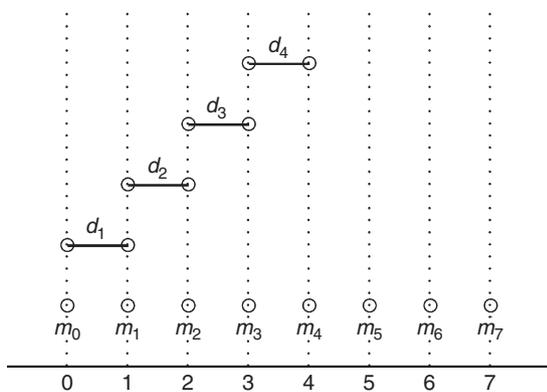


Figure 7 Monomers and dimers.

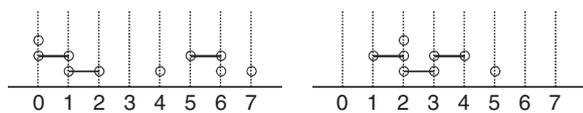


Figure 8 Two heaps of monomers and dimers.

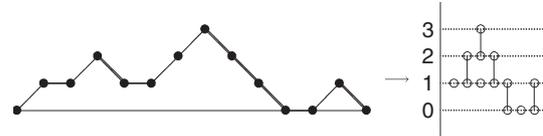


Figure 9 Bijection between Motzkin paths and heaps of monomers and dimers.

the beginning to the end. Whenever we read a level-step at height h , we make it into a monomer with x -coordinate h , whenever we read a down-step from height h to height $h - 1$, we make it into a dimer whose endpoints have x -coordinates $h - 1$ and h . Up-steps are ignored. Figure 9 shows an example. In the figure, the heap is not in “standard” fashion, in the sense that the x -axis is not shown as a horizontal line but as a vertical line (cf. Figure 7). But it could be easily transformed into “standard” fashion by a simple reflection with respect to a line of slope 1.

Lattice animals on the triangular lattice and on the quadratic lattice are also in bijection with heaps, this time with heaps consisting entirely out of dimers. Given an animal, one simply replaces each vertex of the animal by a dimer, see Figures 10 and 11. While in the case of animals on the triangular lattice this gives a constraintless bijection (see Figure 10), in the case of the quadratic lattice this sets up a bijection with heaps of dimers in which two dimers of the same type can never be placed directly one over the other (see Figure 11). For example, two dimers d_5 , one placed directly over the other (as they occur in Figure 10), are forbidden under this rule.

Next we make heaps into a monoid by introducing a composition of heaps. (A monoid is a set with a binary operation which is associative.) Intuitively, given two heaps H_1 and H_2 , the composition of H_1 and H_2 , the heap $H_1 \circ H_2$, is the heap which results

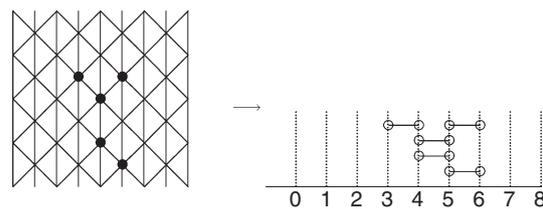


Figure 10 Bijection between animals and heaps of dimers.

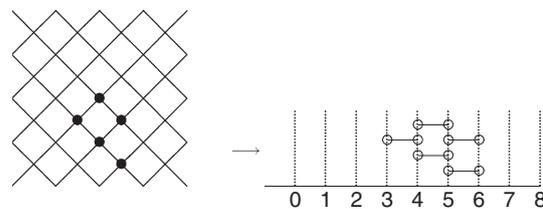


Figure 11 Bijection between animals and heaps of dimers.

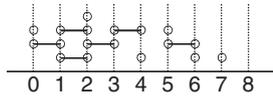


Figure 12 The composition of the heaps in **Figure 8**.

by putting H_2 on top of H_1 . In terms of words, the composition of two heaps is the equivalence class of the concatenation uw , where u is a word from the equivalence class of H_1 , and w is a word from the equivalence class of H_2 .

The composition of the two heaps in **Figure 8** is shown in **Figure 12**.

Given pieces \mathcal{B} with relation \mathcal{R} , let $\mathcal{H}(\mathcal{B}, \mathcal{R})$ be the set of all heaps consisting of pieces from \mathcal{B} , including the empty heap, the latter denoted by \emptyset . It is easy to see that the composition makes $(\mathcal{H}(\mathcal{B}, \mathcal{R}), \circ)$ into a monoid with unit \emptyset .

For the statement of the main theorem in the theory of heaps, we need two more terms. A trivial heap is a heap consisting of pieces all of which are pairwise unrelated. **Figure 13a** shows a trivial heap consisting of monomers and dimers. A pyramid is a heap with exactly one maximal (=topmost) element. **Figure 13a** shows a pyramid consisting of monomers and dimers. Finally, if H is a heap, then we write $|H|$ for the number of pieces in H .

In applications, heaps will have weights, which are defined by introducing a weight $w(B)$ for each piece B in \mathcal{B} , and by extending the weight w to all heaps H by letting $w(H)$ denote the product of all weights of the pieces in H (multiplicities of pieces included).

Let \mathcal{M} be a subset of the pieces \mathcal{B} . Then, the generating function for all heaps with maximal pieces contained in \mathcal{M} is given by

$$\sum_{\substack{H \in \mathcal{H}(\mathcal{B}, \mathcal{R}) \\ \text{maximal pieces} \subseteq \mathcal{M}}} w(H) = \frac{\sum_{T \in \mathcal{T}(\mathcal{B} \setminus \mathcal{M}, \mathcal{R})} (-1)^{|T|} w(T)}{\sum_{T \in \mathcal{T}(\mathcal{B}, \mathcal{R})} (-1)^{|T|} w(T)} \quad [38]$$

where $\mathcal{T}(\mathcal{B}, \mathcal{R})$ denotes the set of all trivial heaps with pieces from \mathcal{B} . In particular, the generating function for all heaps is given by

$$\sum_{H \in \mathcal{H}(\mathcal{B}, \mathcal{R})} w(H) = \frac{1}{\sum_{T \in \mathcal{T}(\mathcal{B}, \mathcal{R})} (-1)^{|T|} w(T)} \quad [39]$$

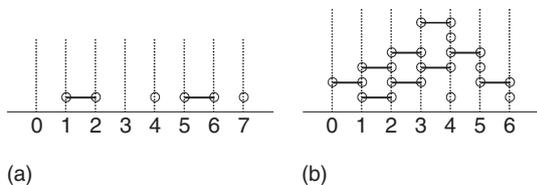


Figure 13 (a) A trivial heap. (b) A pyramid.

Furthermore, if $\mathcal{P}(\mathcal{B}, \mathcal{R})$ denotes the set of all pyramids with pieces from \mathcal{B} , then

$$\sum_{P \in \mathcal{P}(\mathcal{B}, \mathcal{R})} \frac{w(P)}{|P|} = \log \left(\sum_{H \in \mathcal{H}(\mathcal{B}, \mathcal{R})} w(H) \right) \quad [40]$$

where $|P|$ is the number of pieces of P . (As the reader will have guessed, this is a consequence of the “exponential principle” mentioned in the section “generating functions.”)

The Transfer Matrix Method

The transfer matrix method (cf. Stanley (1986), chapter 4 for further reading) applies whenever we are able to build the combinatorial objects that we are interested in by moving on a finite number of states in a step-by-step fashion, where the current step does not depend on the previous ones. (In statistical language, we are considering a finite-state Markov chain.) For example, Motzkin paths which are constrained to stay between two parallel lines, say between $y=0$ and $y=K$, can be described in such a way: the states are the heights $0, 1, \dots, K$, and, if we are in state h , then in the next step we are allowed to move to states $h+1, h$, or $h-1$, except that from state 0 we cannot move to -1 (there is no state -1), and when we are in state K we cannot move to $K+1$ (there is no state $K+1$).

For describing the general situation, let $G = (V, E)$ be a directed graph with vertex set V and edge set E . Let $w_n(u, v)$ denote the number of walks from vertex u to vertex v along edges of G . To compute these numbers, we consider the adjacency matrix of $G, A(G)$. By definition, using our notation, $A(G) = (w_1(u, v))_{u, v \in V}$. Obviously, $(w_n(u, v))_{u, v \in V} = (A(G))^n$. Thus,

$$\left(\sum_{n=0}^{\infty} w_n(u, v) x^n \right)_{u, v \in V} = \sum_{n=0}^{\infty} (A(G))^n x^n = (I_n - A(G)x)^{-1}$$

where I_n is the $n \times n$ identity matrix. In other words, the generating functions $\sum_{n=0}^{\infty} w_n(u, v) x^n$ for the walk numbers between u and v form the entries of a matrix which is the inverse matrix of $I_n - A(G)x$. By elementary linear algebra,

$$\sum_{n=0}^{\infty} w_n(u, v) x^n = \frac{(-1)^{\#u+\#v} \det(I_n - A(G)x)_{v,u}}{\det(I_n - A(G)x)} \quad [41]$$

where $\det(I_n - A(G)x)_{v,u}$ is the minor of $I_n - A(G)x$ with the row indexed by v and the column indexed

by u omitted, and where $\#u$ denotes the row number of u and similarly for $\#v$. A weighted version could also be developed in the same way, where we put a weight $w(e)$ on each edge, and the weight of a walk is the product of the weights of all its edges.

In particular, the expression [41] is a rational function in x . Then, by the basic theorem on rational generating functions (cf. Stanley (1986), section 4.1), the number $w_n(u, v)$ can be expressed as a sum $\sum_{i=1}^d P_i(n)\gamma_i^n$, where the γ_i 's are the different roots of the polynomial $\det(xI_n - A(G))$, and $P_i(n)$ is a polynomial of degree less than the multiplicity of the root γ_i . (The $P_i(n)$'s depend on u and v , whereas the γ_i 's do not.) If there exists a unique root γ_j with maximal modulus, then this implies that, asymptotically as $n \rightarrow \infty$, $w_n(u, v) \sim P_j(n)\gamma_j^n$.

Lattice Paths

Recall from the section on basic combinatorial terminology that a lattice path P in \mathbb{Z}^d is a path in the d -dimensional integer lattice \mathbb{Z}^d which uses only points of the lattice, that is, it is a sequence (P_0, P_1, \dots, P_l) , where $P_i \in \mathbb{Z}^d$ for all i . The vectors $\overrightarrow{P_0P_1}, \overrightarrow{P_1P_2}, \dots, \overrightarrow{P_{l-1}P_l}$ are called the steps of P . The number of steps, l , is called the length of P .

The enumeration of lattice paths has always been an intensively studied topic in statistics, because of their importance in the study of random walks, of rank order statistics for non-parametric testing, and of queueing processes. The reader is referred to Feller (1957) and particularly Mohanty's (1979) book, which is a rich source for enumerative results on lattice paths, albeit in a statistical language. We review the most important results in this section. Most of these concern two-dimensional lattice paths, that is, the case $d=2$.

To begin with, we consider paths in the integer plane \mathbb{Z}^2 consisting of horizontal and vertical unit steps in the positive direction. Clearly, the number of all (unrestricted) paths from the origin to (n, m) is the binomial coefficient $\binom{n+m}{n}$. By the reflection principle, which is commonly attributed to D André (see, e.g., Comtet (1974) p. 22), it follows that the number of paths from the origin to (n, m) which do not pass above the line $y = x + t$, where $m \leq n + t$, is given by

$$\binom{n+m}{n} - \binom{n+m}{n+t+1} \tag{42}$$

Roughly, the reflection principle sets up a bijection between the paths from the origin to (n, m) which do pass above the line $y = x + t$ and all paths

from $(-t-1, t+1)$ to (n, m) , by reflecting the path portion between the origin and the last touching point on $y = x + t + 1$ in this latter line. Thus, the result of the enumeration problem is the number of all paths from $(0, 0)$ to (n, m) , which is given by the binomial coefficient $\binom{n+m}{n}$, minus the number of all paths from $(-t-1, t+1)$ to (n, m) , which is given by the binomial coefficient $\binom{n+m}{n+t+1}$, whence the formula [42].

If one considers more generally paths bounded by the line $my = nx + t$, no compact formula is known. It seems that the most conceptual way to approach this problem is through the so-called kernel method (see the section on solving equations for generating functions), which, in combination with the saddle point method, allows one also to obtain strong asymptotic results. There is one special instance, however, which has a "nice" formula. The number of all lattice paths from the origin to (n, m) which never pass above $x = \mu y$, where μ is a positive integer, is given by

$$\frac{n - \mu m + 1}{n + m + 1} \binom{n + m + 1}{m} \tag{43}$$

The most elegant way to prove this formula is by means of the cycle lemma of Dvoretzky and Motzkin (see Mohanty (1979), p. 9 where the cycle lemma occurs under the name of "penetrating analysis").

Iteration of the reflection principle shows that the number of paths from the origin to (n, m) which stay between the lines $y = x + t$ and $y = x + s$ (being allowed to touch them), where $t \geq 0 \geq s$ and $n + t \geq m \geq n + s$, is given by the finite (!) sum (see, e.g., Mohanty (1979), p. 6)

$$\sum_{k \in \mathbb{Z}} \left(\binom{n+m}{n-k(t-s+2)} - \binom{n+m}{n-k(t-s+2)+t+1} \right) \tag{44}$$

The enumeration of lattice paths restricted to regions bounded by hyperplanes has also been considered for other regions, such as quadrants, octants, and rectangles, as well as in higher dimensions. A general result due to Gessel and Zeilberger, and Biane, independently, on the number of lattice paths in a chamber (alcove) of an (affine) reflection group (see Krattenthaler (2003) for the corresponding references and pointers to further results) shows how far one can go when one uses the reflection principle. In particular, this result covers [42] and [44], the enumeration of lattice paths in quadrants, octants, rectangles, and many other results that have

appeared (before and after) in the literature. We present a particularly elegant (and frequently occurring) special case. (In reflection group language, it corresponds to the reflection group of “type A_{n-1} .” See [Humphreys \(1990\)](#) for terminology and information on reflection groups.)

Let $A = (a_1, a_2, \dots, a_d)$ and $E = (e_1, e_2, \dots, e_d)$ be points in \mathbb{Z}^d with $a_1 \geq a_2 \geq \dots \geq a_d$ and $e_1 \geq e_2 \geq \dots \geq e_d$. The number of all paths from A to E in the integer lattice \mathbb{Z}^d , which consist of positive unit steps and which stay in the region $x_1 \geq x_2 \geq \dots \geq x_d$, equals

$$\left(\sum_{i=1}^d (e_i - a_i) \right)! \det_{1 \leq i, j \leq d} \left(\frac{1}{(e_i - a_j - i + j)!} \right) \quad [45]$$

The counting problem of the theorem is equivalent to numerous other counting problems. It has been originally formulated as an n -candidate ballot problem, but it is as well equivalent to counting the number of standard Young tableaux of a given shape. In the case that all a_i 's are equal, the determinant does in fact evaluate into a closed-form product. In Young tableaux theory, a particular way to write the result is known as the hook-length formula (see, e.g., [Stanley \(1999\)](#), corollary 7.21.6).

We return to lattice paths in the plane, mentioning some more closely related results. The first is a result of [Mohanty \(1979, section 4.2\)](#), which expresses the number of all lattice paths from the origin to (n, m) which touch the line $y = x + t$ exactly r times, never crossing it, as the difference

$$\binom{n+m-r}{n+t-1} - \binom{n+m-r}{n+t}, \quad r \geq 1 \quad [46]$$

Not forbidding that the paths cross the bounding line, we arrive at the problem of counting the lattice paths from the origin to (n, m) , which cross the main diagonal $y = x$ exactly r times, the answer being

$$\begin{cases} \frac{m-n+2r+1}{m+n+1} \binom{m+n+1}{n-r} & \text{if } m > n \\ \frac{2r+2}{n} \binom{2n}{n-r-1} & \text{if } m = n \end{cases} \quad [47]$$

Next, we give the number of lattice paths from the origin to (n, n) which have $2r$ steps on one side of the line $y = x$, as

$$\binom{2r}{r} \binom{2n-2r}{n-r} \quad [48]$$

a result due to Sparre Andersen. We refer the reader to [Mohanty \(1979, chapter 3\)](#) for further results in this direction.

Enumerating lattice paths with a fixed number of maximal straight pieces (which correspond to runs), is intimately connected to another basic enumeration problem concerning lattice paths: the enumeration of lattice paths having a fixed number of turns. An effective way to attack the latter problem is by means of two-rowed arrays (see the survey article by [Krattenthaler \(1997\)](#), where in particular analogs of the reflection principle for two-rowed arrays are developed. These imply formulas for the number of lattice paths with fixed starting points and endpoints and a fixed number of north-east (respectively east-north) turns, for unrestricted paths, as well as for paths bounded by lines. (A north-east turn in a lattice path is a point where the direction changes from “north” to “east.” An east-north turn is defined analogously.) In particular, analogs of [42]–[44] are known when the number of north-east (respectively east-north) turns is fixed.

These formulas imply for example (see again [Krattenthaler \(1997, section 3.5\)](#)) that the number of lattice paths from the origin to (n, n) which never pass above the line $y = x + t$ and have exactly $2r$ maximal straight pieces is given by

$$\begin{aligned} & 2 \binom{n-1}{r-1}^2 - \binom{n+t-1}{r-2} \binom{n-t-1}{r} \\ & - \binom{n+t-1}{r-1} \binom{n-t-1}{r-1} \end{aligned} \quad [49]$$

with a similar result for the case of $2r + 1$ maximal straight pieces. (If $t = 0$, the numbers in [49] become

$$\frac{1}{n} \binom{n}{r} \binom{n}{r-1}$$

and they are known as the Narayana numbers.) Furthermore, they imply that the number of lattice paths from the origin to (n, n) which never pass above the line $y = x + t$ and never below the line $y = x - t$ and have exactly $2r$ maximal straight pieces is given by

$$\begin{aligned} & \sum_{k=-\infty}^{\infty} \left\{ 2 \binom{n-2kt-1}{r+k-1} \binom{n+2kt-1}{r-k-1} \right. \\ & - \binom{n-2kt+t-1}{r+k-2} \binom{n+2kt-t-1}{r-k} \\ & \left. - \binom{n-2kt+t-1}{r+k-1} \binom{n+2kt-t-1}{r-k-1} \right\} \end{aligned} \quad [50]$$

with a similar result for the case of $2r + 1$ maximal straight pieces.

The most general boundary for lattice paths that one can imagine is the restriction that it stays

between two given (fixed) paths. Let us assume that the horizontal steps of the upper (fixed) path are at heights $a_1 \leq a_2 \leq \dots \leq a_n$, whereas the horizontal steps of the lower (fixed) path are at heights $b_1 \leq b_2 \leq \dots \leq b_n$, $a_i \geq b_i$, $i = 1, 2, \dots, n$. Then the number of all paths from $(0, b_1)$ to (n, a_n) satisfying the property that for all $i = 1, 2, \dots, n$ the height of the i th horizontal step is between b_i and a_i is given by the determinant

$$\det_{1 \leq i, j \leq n} \left(\binom{a_i - b_j + 1}{j - i + 1} \right) \quad [51]$$

In the statistical literature, this formula is often known as ‘‘Steck’s formula,’’ but it is actually a special case of a much more general theorem due to Kreweras. A generalization of [51] to higher-dimensional paths was given by Handa and Mohanty (see Mohanty (1979, section 2.4)).

Next, we consider three-step lattice paths in the integer plane \mathbb{Z}^2 , that is, paths consisting of up-steps $(1, 1)$, level steps $(1, 0)$, and down-steps $(1, -1)$. The particular problem that we are interested in is to count such three-step paths starting at $(0, r)$ and ending at (ℓ, s) , which do not pass below the x -axis and do not pass above the horizontal line $y = K$. Furthermore, we assign the weight 1 to an up-step, the weight b_b to a level-step at height b , and the weight λ_b to a down-step from height b to $b - 1$. The weight $w(P)$ of a path P is defined as the product of the weights of all its steps. Then we have the following result, which can be obtained by the transfer matrix method described in the last section.

Define the sequence $(p_n(x))_{n \geq 0}$ of polynomials by

$$xp_n(x) = p_{n+1}(x) + b_n p_n(x) + \lambda_n p_{n-1}(x) \quad [52]$$

for $n \geq 1$

with initial conditions $p_0(x) = 1$ and $p_1(x) = x - b_0$. Furthermore, define $(Sp_n(x))_{n \geq 0}$ to be the sequence of polynomials which arises from the sequence $(p_n(x))$ by replacing λ_i by λ_{i+1} and b_i by b_{i+1} , $i = 0, 1, 2, \dots$, everywhere in the three-term recurrence [52] and in the initial conditions. Finally, given a polynomial $p(x)$ of degree n , we denote the corresponding reciprocal polynomial $x^n p(1/x)$ by $p^*(x)$.

With the weight w defined as before, the generating function $\sum_P w(P)x^{\ell(P)}$, where the sum is over all three-step paths which start at $(0, r)$, terminate at height s , do not pass below the x -axis, and do not pass above the line $y = K$, is given by

$$\begin{cases} \frac{x^{s-r} p_r^*(x) S^{s+1} p_{K-s}^*(x)}{p_{K+1}^*(x)}, & r \leq s \\ \lambda_r \cdots \lambda_{s+1} \frac{x^{r-s} p_s^*(x) S^{r+1} p_{K-r}^*(x)}{p_{K+1}^*(x)}, & r \geq s \end{cases} \quad [53]$$

The sequence of polynomials $(p_n(x))_{n \geq 0}$ is in fact a sequence of orthogonal polynomials (cf. Koekoek and Swarttouw (1998) and Szegö (1959)).

We remark that in the case that $r = s = 0$ there is also an elegant expression for the generating function due to Flajolet (see section V.2 of the Flajolet and Sedgewick reference in ‘‘Further reading’’) in terms of a continued fraction.

In order to solve our problem, we just have to extract the coefficient of x^ℓ in [53]. By a partial fraction expansion, a formula of the type

$$\sum_m c_m \xi_m^\ell \quad [54]$$

results, where the ξ_m ’s are the zeroes of $p_{K+1}(x)$, and the c_m ’s are some coefficients, only a finite number of them being nonzero.

It should be noted that, because of the many available parameters (the b_n ’s and λ_n ’s), by appropriate specializations one can also obtain numerous results about enumerating three-step paths according to various statistics, such as the number of touchings on the bounding lines, etc.

There are two important special cases in which a completely explicit solution in terms of elementary functions can be given.

The first case occurs for $b_i = 0$ and $\lambda_i = 1$ for all i . In this case, the polynomials $p_n(x)$ defined by the three-term recurrence [52] are Chebyshev polynomials of the second kind, $p_n(x) = U_n(x/2)$. (The Chebyshev polynomial of the second kind $U_n(x)$ is defined by $U_n(\cos t) = \sin((n+1)t)/\sin t$ (see Koekoek and Swarttouw (1998) for almost exhaustive information on these polynomials and, more generally, on hypergeometric orthogonal polynomials)). The result which is then obtained from the general theorem (clearly, the zeros of $U_n(x)$ are $x = \cos(2k\pi/(n+1))$, $k = 1, 2, \dots, n$, and therefore the partial fraction expansion of [53] is easily determined) is that the number of lattice paths from $(0, r)$ to (ℓ, s) with only up- and down-steps, which always stay between the x -axis and the line $y = K$, is given by (see also Feller (1957, chapter XIV, eqn [5.7])

$$\frac{2}{K+2} \sum_{k=1}^{K+1} \left(2 \cos \frac{\pi k}{K+2} \right)^\ell \times \sin \frac{\pi k(r+1)}{K+2} \sin \frac{\pi k(s+1)}{K+2} \quad [55]$$

a formula which goes back to Lagrange.

The second case occurs for $b_i = 1$ and $\lambda_i = 1$ for all i . In this case, the polynomials $p_n(x)$ defined by the three-term recurrence [52] are again Chebyshev polynomials of the second kind, $p_n(x) = U_n((x-1)/2)$. The result which is then

obtained from the general theorem is that the number of three-step lattice paths from $(0, r)$ to (ℓ, s) , which always stay between the x -axis and the line $y = K$, is given by

$$\frac{2}{K+2} \sum_{k=1}^{K+1} \left(2 \cos \frac{\pi k}{K+2} + 1 \right)^\ell \times \sin \frac{\pi k(r+1)}{K+2} \sin \frac{\pi k(s+1)}{K+2} \quad [56]$$

Perfect Matchings and Tilings

In this section we consider the problem of counting the perfect matchings of a graph. For an introduction into the problem, and into methods to solve it, as well as for a report on recent developments, we refer the reader to [Propp \(1999\)](#).

Let $G = (V, E)$ be a finite loopless graph with vertex set V and edge set E . A matching (also called 1-factor in graph theory) is a subset of the edges with the property that no two edges share a vertex. A matching is perfect if it covers all the edges. Let $M(G)$ denote the number of perfect matchings of the graph G . More generally, we could assign a weight $w(e)$ to each edge e of the graph and define the weight of a matching to be the product of the weights of all its edges. Let $M_w(G)$ denote the sum of all weights of all matchings of the graph G .

Kasteleyn’s method for determining $M(G)$, respectively $M_w(G)$, makes use of determinants and Pfaffians. Recall that the Pfaffian $\text{Pf}(A)$ of a triangular array $A = (a_{i,j})_{1 \leq i < j \leq 2n}$ is defined by

$$\text{Pf}(A) = \sum_m (\text{sgn } m) \prod_{\{i,j\} \in m} \alpha_{i,j} \quad [57]$$

where the sum is over all perfect matchings of the complete graph on vertices $\{1, 2, \dots, 2n\}$, and where the product is over all edges $\{i, j\}, i < j$, of m . The sign $\text{sgn } m$ of m is $(-1)^{\#\text{crossings of } m}$, where a crossing is a pair $(\{i, j\}, \{k, l\})$ of edges such that $i < k < j < l$. Usually, one extends the triangular array A to a matrix by setting $a_{j,i} = -a_{i,j}, i < j$, and $a_{i,i} = 0$ for all i . Then, abusing notation, we identify the triangular array with the skew-symmetric matrix $A = (a_{i,j})_{1 \leq i, j \leq 2n}$. The Pfaffian satisfies the following useful properties:

$$\text{Pf}(B^t A B) = \det(B) \text{Pf}(A)$$

and

$$\text{Pf}(A)^2 = \det(A) \quad [58]$$

The latter equality shows in particular that Pfaffians are very close to determinants. They do, in fact, generalize determinants since

$$\text{Pf} \begin{pmatrix} 0 & B \\ -B & 0 \end{pmatrix} = \det B \quad [59]$$

for any square matrix B .

Thus, given a graph with vertices v_1, v_2, \dots, v_{2n} , specializing $a_{i,j}$ to the weight of the edge between v_i and v_j , if it exists, and setting $a_{i,j} = 0$ otherwise in the definition of the Pfaffian, we obtain almost $M_w(G)$, the only difference is that there could be signs in front of the individual terms of the sum, whereas in $M_w(G)$ the sign in front of each term must be $+$. (The object obtained by omitting the sign in [57] is called Hafnian. Unfortunately, in contrast to the Pfaffian, it does not have any nice properties and it is therefore extremely difficult to compute.) Kasteleyn’s idea is to circumvent this problem by orienting the edges of the graph, defining signed weights of the edges, in such a way that the Pfaffian of the array with signed weights produces exactly $M_w(G)$.

More precisely, given a (weighted) graph G with vertices v_1, v_2, \dots, v_{2n} , we make it into an oriented (weighted) graph \vec{G} . That is, if there is an edge between v_i and $v_j, e_{i,j}$ say, we orient it either from v_i to v_j or the other way. Now we define the signed adjacency matrix $A(\vec{G})$ of \vec{G} by letting its (i, j) -entry to be $+w(e_{i,j})$ if there is an edge from v_i to v_j oriented that way, $-w(e_{i,j})$ if there is an edge from v_j to v_i oriented that way, and 0 if there is no edge between v_i and v_j . Such an orientation is called Pfaffian if

$$\text{Pf}(A(\vec{G})) = \pm M_w(G)$$

Clearly, the question remains whether a Pfaffian orientation can be found for a given graph. In general, this is an open question. However, Kasteleyn shows that for planar graphs such a Pfaffian orientation can always be found. Moreover, he shows that any orientation of a planar graph which has the property that around any face bounded by $4k$ edges an odd number of edges is oriented in either direction and that around any face bounded by $4k + 2$ edges an even number of edges is oriented in either direction is Pfaffian.

For bipartite graphs (i.e., for graphs in which the set of vertices can be split into two disjoint sets such that all the edges connect the vertex of one of these sets to a vertex of the other), the situation is even nicer. This is because for a bipartite graph G in which both parts of the bipartition of the vertices are of the same size (otherwise, there is no perfect matching), any signed

adjacency matrix $A(\vec{G})$ has the block form of the matrix on the left-hand side of [59] and, hence, the Pfaffian reduces to a determinant. More precisely, let G be a bipartite graph with vertex set $V = U \cup W$, $U = \{u_1, u_2, \dots, u_n\}$ and $W = \{w_1, w_2, \dots, w_n\}$, with edges connecting some u_i to some w_j . Given a Pfaffian orientation \vec{G} , we build the signed bipartite adjacency matrix $B(\vec{G}) = (b_{i,j})_{1 \leq i, j \leq n}$ of \vec{G} by setting $b_{i,j} = +w(e_{i,j})$ if there is an edge from u_i to w_j oriented that way, $-w(e_{i,j})$ if there is an edge from w_j to u_i oriented that way, and 0 if there is no edge between u_i and w_j . Then we have

$$\det(B(\vec{G})) = \pm M_w(G)$$

In particular, this holds for any bipartite planar graph. See Robertson *et al.* (1999) for a structural description about which (not necessarily planar) bipartite graphs admit a Pfaffian orientation.

Kasteleyn’s construction in the planar case has been generalized to graphs on surfaces of any genus g in Dolbilen *et al.* (1996), Galluccio and Loebli (1999), and Tesler (2000), independently. As predicted by Kasteleyn, the solution is in terms of a linear combination of 4^g Pfaffians.

With the help of his method, Kasteleyn computed the number of dimer coverings of an $m \times n$ rectangle. (A dimer is a 2×1 rectangle. Thus, this is equivalent to counting the number of perfect matchings on the $m \times n$ grid graph. The formula was independently found by Temperley and Fisher.) The result is

$$\prod_{i=1}^m \prod_{j=1}^n \left(2 \cos \frac{\pi i}{m+1} + 2\sqrt{-1} \cos \frac{\pi j}{n+1} \right)$$

For even m and n , the formula can be rewritten as

$$\prod_{i=1}^{m/2} \prod_{j=1}^{n/2} \left(4 \cos^2 \frac{\pi i}{m+1} + 4 \cos^2 \frac{\pi j}{n+1} \right)$$

There is a similar rewriting if one of m or n is odd. (If both m and n are odd, there is no dimer covering.)

For further reading and references see Dimer Problems and Kuperberg (1998).

Nonintersecting Paths

Let $G = (V, E)$ be a directed acyclic graph with vertices V and directed edges E . Furthermore, we are given a function w which assigns a weight $w(x)$ to every vertex or edge x . Let us define the weight $w(P)$ of a walk P in the graph by $\prod_e w(e) \prod_v w(v)$, where the first product is over all edges e of the walk P and the second product is over all vertices v of P . We

denote the set of all walks in G from u to v by $\mathcal{P}(u \rightarrow v)$, and the set of all families (P_1, P_2, \dots, P_n) of walks, where P_i runs from u_i to $v_i, i = 1, 2, \dots, n$, by $\mathcal{P}(\mathbf{u} \rightarrow \mathbf{v})$, with $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$. The symbol $\mathcal{P}^+(\mathbf{u} \rightarrow \mathbf{v})$ stands for the set of all families (P_1, P_2, \dots, P_n) in $\mathcal{P}(\mathbf{u} \rightarrow \mathbf{v})$ with the additional property that no two walks share a vertex. We call such families of walk(er)s “vicious walkers” or, alternatively, “nonintersecting paths.” The weight $w(P)$ of a family $P = (P_1, P_2, \dots, P_n)$ of walks is defined as the product $\prod_{i=1}^n w(P_i)$ of all the weights of the walks in the family. Finally, given a set \mathcal{M} with weight function w , we write $\text{GF}(\mathcal{M}; w)$ for the generating function $\sum_{x \in \mathcal{M}} w(x)$.

We need two further notations before we are able to state the Lindström–Gessel–Viennot theorem. (For references and historical remarks, we refer the reader to footnote 5 in Krattenthaler (2005a).) As earlier, the symbol \mathfrak{S}_n denotes the symmetric group of order n . Given a permutation $\sigma \in \mathfrak{S}_n$, we write \mathbf{u}_σ for $(u_{\sigma(1)}, u_{\sigma(2)}, \dots, u_{\sigma(n)})$. Then

$$\begin{aligned} \sum_{\sigma \in \mathfrak{S}_n} (\text{sgn } \sigma) \cdot \text{GF}(\mathcal{P}^+(\mathbf{u}_\sigma \rightarrow \mathbf{v}); w) \\ = \det_{1 \leq i, j \leq n} (\text{GF}(\mathcal{P}(u_j \rightarrow v_i); w)) \end{aligned} \quad [60]$$

Most often, this theorem is applied in the case where the only permutation σ for which vicious walks exist is the identity permutation, so that the sum on the left-hand side reduces to a single term that counts all families (P_1, P_2, \dots, P_n) of vicious walks, the i th walk P_i running from A_i to $E_i, i = 1, 2, \dots, n$. This case occurs, for example, if for any pair of walks (P, Q) with P running from u_a to v_d and Q running from u_b to $v_c, a < b$ and $c < d$, it is true that P and Q must have a common vertex. Explicitly, in that case we have

$$\text{GF}(\mathcal{P}^+(\mathbf{u} \rightarrow \mathbf{v}); w) = \det_{1 \leq i, j \leq n} (\text{GF}(\mathcal{P}(u_j \rightarrow v_i); w)) \quad [61]$$

If the starting points or/and the endpoints are not fixed, then the corresponding number is given by a Pfaffian, a result obtained by Okada and Stembridge (see Bressoud (1999) for references). For a set \mathcal{A} of starting points, let $\mathcal{P}^+(\mathcal{A} \rightarrow \mathbf{v})$ denote the set of all families $(P_1, P_2, \dots, P_{2n})$ of nonintersecting lattice paths, where P_i runs from some point of \mathcal{A} to $v_i, i = 1, 2, \dots, 2n$. Furthermore, let us suppose that the elements of $\mathcal{A} = \{u_1, u_2, \dots\}$ are ordered in such a way that for any pair of walks (P, Q) with P running from u_a to v_d and Q running from u_b to $v_c, a < b$ and $c < d$, it is true that P and Q must have a common vertex. (This is the same condition as the one which makes [61] valid, with the only difference that, here,

the number of u_i 's could be larger than the number of v_i 's.) Then,

$$\begin{aligned}
 & \text{GF}(\mathcal{P}^+(\mathcal{A} \rightarrow \mathbf{v}); w) \\
 &= \text{Pf}_{1 \leq i, j \leq 2n} \left(\sum_{a < b} (\text{GF}(\mathcal{P}(u_a \rightarrow v_i); w) \text{GF}(\mathcal{P}(u_b \rightarrow v_j); w) \right. \\
 & \quad \left. - \text{GF}(\mathcal{P}(u_b \rightarrow v_i); w) \text{GF}(\mathcal{P}(u_a \rightarrow v_j); w)) \right) \quad [62]
 \end{aligned}$$

If the number of paths is odd, then one can use the same formula by adding an artificial point to the endpoints and to the set of starting points \mathcal{A} . There is also a theorem by Okada and Stembridge which covers the case that starting points and endpoints vary. Refinements when the number of turns is fixed can be found in Krattenthaler (1997).

Vicious Walkers, Plane Partitions, Rhombus Tilings, and Fully Packed Loop Configurations

In this section we describe the interrelations between four frequently appearing objects in statistical mechanics and combinatorics: vicious walkers, plane partitions, rhombus tilings, and fully packed loop configurations.

Given a lattice, vicious walkers, as introduced by Fisher (1984), are particles which move on lattice sites in such a way that two particles never occupy the same lattice site. Models of vicious walkers have been the object of numerous studies from various points of view. Rather than accomplishing the impossible task of providing a complete overview of references, the reader is referred to the basic reference Fisher (1984) and to Krattenthaler (2005a) for further pointers to the literature.

Most of the known results apply for vicious walkers on the line. There are in fact two different models: in the random turns vicious walker model, n walkers move on the integral points of the real line in such a way that at each tick of the clock exactly one walker moves to the right or to the left, whereas in the lock step vicious walker model n walkers move on the integral points of the real line in such a way that at each tick of the clock each walker moves to the right or to the left.

The first model is equivalent to a model of one walker in \mathbb{Z}^n (\mathbb{Z} denoting the set of integers) which at each tick of the clock moves a positive or negative unit step in the direction of one of the coordinate axes, always staying in the wedge $x_1 > x_2 > \dots > x_n$. This point of view was already put forward by Fisher (1984). However, this problem belongs to the problem of counting paths in chambers of reflection groups discussed in the section “Lattice paths.”

The second model could also be realized as a single walker model (cf. Krattenthaler (2003)). However, most often it is realized as a model of n paths in the plane consisting of steps $(1, 1)$ and $(1, -1)$ with the property that no two paths have a point in common. In this picture, the x -axis becomes the time line, the k th path doing an up-step $(1, 1)$ from $(t - 1, y)$ to $(t, y + 1)$ meaning that the k th particle moves to the left at time t , whereas the k th path doing a down-step $(1, -1)$ from $(t - 1, y)$ to $(t, y - 1)$ meaning that the k th particle moves to the right at time t .

The reader should consult Figure 14a for an example. (The labelings should be ignored at this point.) Clearly, what we encounter here is a particular instance of the nonintersecting paths of the last section. Therefore, for fixed starting points and endpoints, formula [61] applies, whereas if the starting points vary and the endpoints are fixed, it is formula [62] that applies.

At this point, the links to the other objects, semistandard tableaux and plane partitions (cf. Bressoud (1999)), emerge. A filling of the cells of the Ferrers diagram of λ with elements of the set $\{1, 2, \dots\}$, which is weakly increasing along rows and strictly increasing along columns is called a (semistandard) tableau of shape λ . Figure 14b shows such a semistandard tableau of shape $(4, 3, 2)$. In fact, vicious walkers and semistandard tableaux are equivalent objects. To see this, first label down-steps by the x -coordinate of their endpoint, so that a step from $(a - 1, b)$ to $(a, b - 1)$ is labeled by a , see Figure 14a. Then, out of the labels of the j th path, form the j th column of the corresponding tableau,

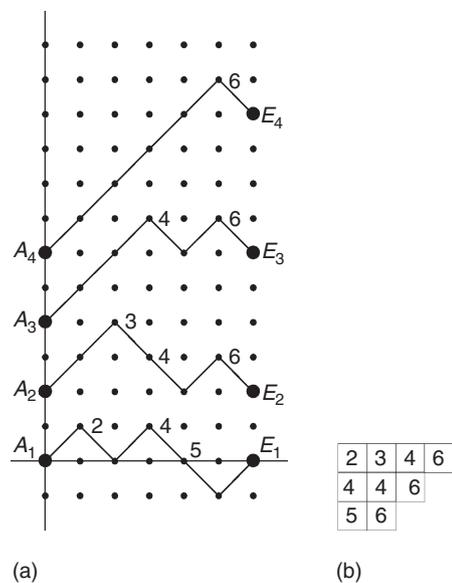


Figure 14 (a) Vicious walkers. (b) A tableau.

see [Figure 14b](#). The resulting array of numbers is indeed a semistandard tableau. This can be readily seen, since the entries are trivially strictly increasing along columns, and they are weakly increasing along rows because the paths do not touch each other. Thus, problems of enumerating vicious walkers can be translated into tableau enumeration problems, and vice versa.

The significance of semistandard tableaux lies particularly in the representation theory for classical groups, see *Classical Groups and Homogenous Spaces and Compact Groups and Their Representations*. Namely, the irreducible characters for $GL(n, \mathbb{C})$ and $SL(n, \mathbb{C})$, the Schur functions, are generating functions for semistandard tableaux of a given shape. If the entries of the i th row of a semistandard tableau are required to be at least $2i - 1$, then one speaks of symplectic tableaux, and the irreducible characters for $Sp(2n, \mathbb{C})$ are generating functions for symplectic tableaux of a given shape. We refer the reader to [Krattenthaler et al. \(2000\)](#) for more information on these topics.

Objects which are very close to semistandard tableaux are plane partitions. According to MacMahon, a plane partition of shape λ is a filling of the Ferrers diagram of λ with non-negative integers which is weakly decreasing along rows and columns. See [Figure 15b](#) for an example of a plane partition of shape $(3, 3, 3)$. In particular, semistandard tableaux and plane partitions of rectangular shape are actually equivalent. For, let T be a semistandard tableau of rectangular shape. Then, from each element of the i th row we subtract i . Finally, the obtained array is rotated by 180° . As a result, we obtain a plane partition. See [Figure 15](#) for a semistandard tableau and a plane partition which correspond to each other under these transformations.

On the other hand, plane partitions can also be realized as three-dimensional objects, by interpreting each entry in the array as a pile of unit cubes of the size of the entry. For example, the plane partition in [Figure 15](#) corresponds to the pile of cubes in [Figure 16a](#). But then, forgetting the three-dimensional view, by embedding the picture in a minimally bounding hexagon, and by filling the emerging empty regions by rhombi of unit length in the unique way this is possible, we obtain a rhombus tiling of a hexagon in

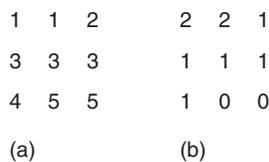


Figure 15 (a) A semistandard tableau. (b) A plane partition.

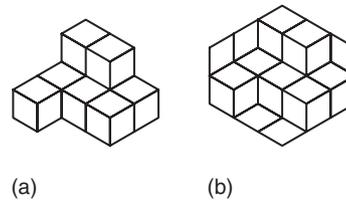


Figure 16 (a) A plane partition; three-dimensional view. (b) A rhombus tiling.

which opposite sides have the same length, see [Figure 16b](#).

From the rhombus tiling, there is then again an elegant way to go to nonintersecting paths: we mark the mid-points of the edges along two opposite sides, see [Figure 17a](#). Now we draw lattice paths which connect points on different sides, by “following” along the other lozenges, as indicated in [Figure 17a](#) by the dashed lines. Clearly, the resulting paths are nonintersecting, that is, no two paths have a common vertex. If we slightly distort the underlying lattice, we get orthogonal paths with horizontal and vertical steps in the positive direction, see [Figure 17b](#).

Rhombus tilings, on their part, are equivalent to perfect matchings of hexagonal graphs. To see this, one places the tiling on the underlying triangular grid, see [Figure 18a](#). Then one places a bond into each rhombus, so that it connects the mid-points of the two triangles out of which the rhombus is composed, see [Figure 18b](#). Finally, one forgets the contour of the tiling, but instead one introduces all the other edges which connect mid-points of adjacent triangles of the triangular grid, see [Figure 18c](#). Thus, one arrives at a perfect matching of the hexagonal graph consisting of the edges connecting mid-points of triangles.

Because of these various connections, enumeration problems for vicious walkers, plane partitions, tableaux, rhombus tilings can be approached by the different methods which are available for the various objects: the determinant theorem from the section “[Nonintersecting paths](#),” together with determinant evaluation techniques (cf. the survey [Krattenthaler \(2005b\)](#)), apply, as well as the “Kasteleyn method” from the section “[Perfect](#)

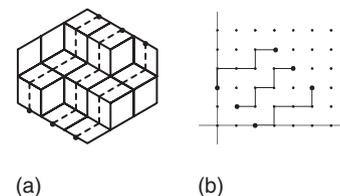


Figure 17 (a) A rhombus tiling. (b) A family of nonintersecting paths.

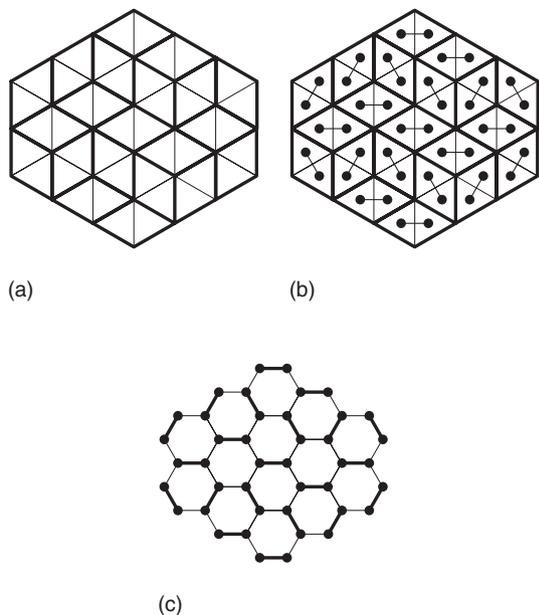


Figure 18 (a) A rhombus tiling. (b) Bonds in rhombi. (c) A perfect matching of a hexagonal graph.

matchings and tilings,” and also methods from character theory for the classical groups. All of these methods have been applied extensively (see the surveys by Kenyon (2003), Propp (1999), and Krattenthaler *et al.* (2000)), the first and third more frequently for exact enumeration, while the second particularly for asymptotic studies. It should be noted that methods from random matrix theory also apply in certain situations, see Johansson (2002). See Growth Processes in Random Matrix Theory and Random Matrix Theory in Physics.

In fact, we missed mentioning a further object, from statistical physics, which in some cases is equivalent to vicious walkers, etc.: fully packed loop configurations. (Fully packed loop configurations are in bijection with six-vertex configurations, see the next section.) If one imposes certain “connectivity constraints” on fully packed loop configurations, then one can construct bijections with rhombus tilings and, hence, with nonintersecting paths and with the other objects discussed in this section. The reader is referred to Di Francesco *et al.* (2004) and references therein.

Having explained the various connections, we cite some fundamental results in the area. (We refer the reader to Bressoud (1999) and Stanley (1999, chapter 7).) MacMahon proved that the number of all plane partitions contained in an $a \times b \times c$ box (when viewed in three dimensions) is equal to

$$\prod_{i=1}^a \prod_{j=1}^b \prod_{k=1}^c \frac{i+j+k-1}{i+j+k-2} \tag{63}$$

Thus, the number of rhombus tilings of a hexagon with side lengths a, b, c, a, b, c is given by the same number, as well as the number of all vicious walkers (P_1, P_2, \dots, P_a) , where P_i runs from $(0, 2i)$ to $(b+c, b-c+2i), i=1, 2, \dots, a$. More generally, the number of semistandard tableaux of shape λ with entries at most m is given by the hook-content formula

$$\prod_{u \in \lambda} \frac{c(u) + m}{h(u)} \tag{64}$$

where u ranges over all the cells of the Ferrers diagram of λ , with $c(u)$ being the content of u , defined as the difference of the column number and the row number of u , and with $h(u)$ being the hook length of u , the latter consisting of the cells to the right of u in the same row and below u in the same column, including u . Thus, this also gives a formula for the number of all vicious walkers (P_1, P_2, \dots, P_a) , where P_i runs from $(0, 2i)$ to (N, b_i) . See Krattenthaler *et al.* (2000, section 2) for details. There it is also explained that a Schur function summation formula, together with an analog of the hook-content formula for special orthogonal characters, proves that the number of all vicious walkers (P_1, P_2, \dots, P_a) , where P_i runs from $(0, 2i)$ for N steps is given by

$$\prod_{1 \leq i < j \leq N} \frac{a+i+j-1}{i+j-1} \tag{65}$$

The reader is referred to the references given in this section for many more results, in particular, on the enumeration of plane partitions with symmetry, the enumeration of rhombus tilings of regions other than hexagons, and the enumeration of vicious walkers with various starting points and endpoints, under various constraints.

Six-Vertex Model and Alternating-Sign Matrices

An alternating-sign matrix is a square matrix of 0’s, 1’s and -1 ’s for which the sum of entries in each row and in each column is 1 and the nonzero entries of each row and of each column alternate in sign. For instance,

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

is a 5×5 alternating-sign matrix. Zeilberger proved that the number of $n \times n$ alternating-sign matrices is given by

$$\prod_{i=0}^{n-1} \frac{(3i+1)!}{(n+i)!} \tag{66}$$

and he went on to prove the finer version that the number of $n \times n$ alternating-sign matrices with the (unique) 1 in the first row in position j is given by

$$\frac{\binom{n+j-2}{n-1} \binom{2n-j-1}{n-1}}{\binom{3n-2}{n-1}} \prod_{i=0}^{n-1} \frac{(3i+1)!}{(n+i)!} \tag{67}$$

The first number is also equal to the number of totally symmetric self-complementary plane partitions contained in the $(2n) \times (2n) \times (2n)$ box, but there is no intrinsic explanation why this is so. We refer the reader to Bressoud (1999) for an exposition of these results, and for pointers to the literature containing further unexplained connections between alternating-sign matrices and plane partitions.

While the first result was achieved by a brute-force constant-term approach, the second result is based on the observation that alternating-sign matrices are in bijection with configurations in the six-vertex model on the square grid under domain-wall boundary conditions. This then allowed one to use a formula due to Izergin for the partition function for these six-vertex configurations. Similar formulas for variations of the model have been found by Kuperberg, and by Razumov and Stroganov (see Razumov and Stroganov (2005) and references therein).

A configuration in the six-vertex model is an orientation of edges of a 4-regular graph (i.e., at each vertex there meet exactly four edges) such that at each vertex two edges are oriented towards the vertex and two are oriented away from the vertex. Thus, there are six possible vertex configurations, giving the name of the model, see Figure 19. To go from one object to the other, one uses the translation between local configurations at a vertex and entries in alternating-sign matrices indicated in the figure. An example of the correspondence can be found in Figure 20.

Another manifestation of alternating-sign matrices and six-vertex configurations are fully packed loop configurations. A fully packed loop configuration on a graph is a collection of edges such that each vertex is

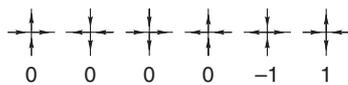


Figure 19 The six vertex configurations.

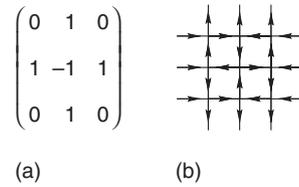


Figure 20 (a) An alternating-sign matrix. (b) A six-vertex configuration.

incident to exactly two edges. One obtains a fully packed loop configuration out of a six-vertex configuration by dividing the square lattice into its even and odd sublattice denoted by A and B , respectively. Instead of arrows, only those edges are drawn that, on sublattice A , point inward and, on sublattice B , point outward. The reader is referred to de Gier (2005) and Di Francesco *et al.* (2004) for further reading.

The story of alternating-sign matrices and their connection to the six-vertex model is given a vivid account in Bressoud (1999), with further important results by Kuperberg, Okada, Razumov and Stroganov, referenced in Razumov and Stroganov (2005).

Fully packed loop configurations seem to play an important role in the explicit form of the ground-state vectors of certain Hamiltonians in the dense $O(1)$ loop model. The corresponding conjectures are surveyed in de Gier (2005). There is important progress on these conjectures by Di Francesco and Zinn-Justin (2005, and references therein).

Binomial Sums and Hypergeometric Series

When dealing with enumerative problems, it is inevitable to deal with binomial sums, that is, sums in which the summands are products/quotients of binomial coefficients and factorials, such as, for example,

$$\sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k}$$

In most cases, the right environment in which one should work is the theory of (generalized) hypergeometric series. These are defined as follows:

$${}_rF_s \left[\begin{matrix} a_1, \dots, a_r \\ b_1, \dots, b_s \end{matrix} ; z \right] = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_r)_k}{(b_1)_k \cdots (b_s)_k} \frac{z^k}{k!}$$

where $(\alpha)_k = \alpha(\alpha+1)(\alpha+2) \cdots (\alpha+k-1)$ for $k > 0$, and $(\alpha)_0 = 1$. The symbol $(\alpha)_k$ is called the Pochhammer symbol or shifted factorial. For in-depth treatments of the subject, we refer the reader

to Andrews *et al.* (1999), Gasper and Rahman (2004), and Slater (1966).

Hypergeometric series can be characterized as series in which the quotient of the $(k + 1)$ st by the k th summand is a rational function in k . This is also the way to convert binomial sums into their hypergeometric form (respectively to see if this is possible; in most cases it is): form the quotient of the $(k + 1)$ st by the k th summand and read off the parameters $a_1, \dots, a_r, b_1, \dots, b_s$, and the argument z from the factorization of the numerator and the denominator polynomials of the rational function, out of these form the corresponding hypergeometric series, and multiply the series by the summand for $k = 0$. This is, in fact, a completely routine task, and, indeed, computer algebra programs such as Maple and Mathematica do this automatically.

The reason why hypergeometric series are much more fundamental than the binomial sums themselves is that there are hundreds of ways to write the same sum using binomial coefficients and factorials, whereas there is just one hypergeometric form, that is, hypergeometric series are a kind of normal form for binomial sums. In particular, given a specific binomial sum, it is a hopeless enterprise to scan through all the identities available in the literature for this sum. There may be an identity for it, but perhaps written differently. On the contrary, given a specific hypergeometric series, the list of available identities which apply to this series is usually not large, and tables of such identities can be set up in a systematic way. This has been done (cf. Slater (1966); the most comprehensive table available to this date is contained in the manual of the Mathematica package HYP – see “Further reading”), and scanning through these tables is largely facilitated by the use of the Mathematica package HYP.

We give here some of the most important identities for hypergeometric series. Aside from the binomial theorem, the most important summation formulas are: the Gauß ${}_2F_1$ -summation formula

$${}_2F_1 \left[\begin{matrix} a, b \\ c \end{matrix} ; 1 \right] = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$$

provided $\Re(c-a-b) > 0$,
the Pfaff–Saalschütz summation formula

$${}_3F_2 \left[\begin{matrix} a, b, -n \\ c, 1+a+b-c-n \end{matrix} ; 1 \right] = \frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n}$$

provided n is a non-negative integer, and
the Dougall summation formula

$${}_7F_6 \left[\begin{matrix} a, a/2+1, b, c, d, 1+2a-b-c-d+n, -n \\ a/2, 1+a-b, 1+a-c, 1+a-d, \\ -a+b+c+d-n, a+1+n \end{matrix} ; 1 \right] = \frac{(1+a)_n(1+a-b-c)_n(1+a-b-d)_n(1+a-c-d)_n}{(1+a-b)_n(1+a-c)_n(1+a-d)_n(1+a-b-c-d)_n}$$

provided n is a non-negative integer.

Some of the most important transformation formulas are
the Euler transformation formula

$${}_2F_1 \left[\begin{matrix} a, b \\ c \end{matrix} ; z \right] = (1-z)^{c-a-b} {}_2F_1 \left[\begin{matrix} c-a, c-b \\ c \end{matrix} ; z \right]$$

provided $|z| < 1$,

the Kummer transformation formula

$${}_3F_2 \left[\begin{matrix} a, b, c \\ d, e \end{matrix} ; 1 \right] = \frac{\Gamma(e)\Gamma(d+e-a-b-c)}{\Gamma(e-a)\Gamma(d+e-b-c)} \times {}_3F_2 \left[\begin{matrix} a, d-b, d-c \\ d, d+e-b-c \end{matrix} ; 1 \right]$$

provided both series converge,

and the Whipple transformation formulas

$${}_4F_3 \left[\begin{matrix} a, b, c, -n \\ e, f, 1+a+b+c-e-f-n \end{matrix} ; 1 \right] = \frac{(e-a)_n(f-a)_n}{(e)_n(f)_n} \times {}_4F_3 \left[\begin{matrix} -n, a, 1+a+c-e-f-n, 1+a+b-e-f-n \\ 1+a+b+c-e-f-n, 1+a-e-n, 1+a-f-n \end{matrix} ; 1 \right] \tag{68}$$

where n is a non-negative integer, and

$${}_7F_6 \left[\begin{matrix} a, 1+\frac{a}{2}, b, c, d, e, -n \\ \frac{a}{2}, 1+a-b, 1+a-c, 1+a-d, 1+a-e, 1+a+n \end{matrix} ; 1 \right] = \frac{(1+a)_n(1+a-d-e)_n}{(1+a-d)_n(1+a-e)_n} \times {}_4F_3 \left[\begin{matrix} 1+a-b-c, d, e, -n \\ 1+a-b, 1+a-c, -a+d+e-n \end{matrix} ; 1 \right] \tag{69}$$

provided n is a non-negative integer.

Since about 1990, for the verification of binomial and hypergeometric series, there are automatic tools available. The book by Petkovšek *et al.* (1996) is an excellent introduction into these aspects. The philosophy is as follows. Suppose we are given a binomial or hypergeometric series $S(n) = \sum_k F(n, k)$. The Gosper–Zeilberger algorithm (see “Further reading”) (cf. Petkovšek *et al.* (1996); a simplified version was presented in the reference Zeilberger in “Further reading”) will find a linear recurrence

$$A_0(n)S(n) + A_1(n)S(n+1) + \dots + A_d(n)S(n+d) = C(n) \quad [70]$$

for some d , where the coefficients $A_i(n)$ are polynomials in n , and where $C(n)$ is a certain function in n , with proof!

If, for example, we suspected that $S(n) = \text{RHS}(n)$, where $\text{RHS}(n)$ is some closed-form expression, then we just have to verify that $\text{RHS}(n)$ satisfies the recurrence [70] and check $S(n) = \text{RHS}(n)$ for sufficiently many initial values of n to have a proof for the identity $S(n) = \text{RHS}(n)$ for all n . On the other hand, if $\text{RHS}(n)$ was a different sum, then we would apply the algorithm to find a recurrence for $\text{RHS}(n)$. If it turns out to be the same recurrence then, again, a check of $S(n) = \text{RHS}(n)$ for a few initial values will provide a full proof of $S(n) = \text{RHS}(n)$ for all n .

Even in the case that we do not have a conjectured expression $\text{RHS}(n)$, this is not the end of the story. Given a recurrence of the type [70], the Petkovšek algorithm (see “Further reading”) (cf. Petkovšek *et al.* (1996)) is able to find a closed-form solution (where “closed form” has a precise meaning), respectively tell that there is no closed-form solution.

The fascinating point about both algorithms is that neither do we have to know what the algorithm does internally nor do we have to check that. For the Petkovšek algorithm, this is obvious anyway because, once the computer says that a certain expression is a solution of [70], it is a routine matter to check that. This is less obvious for the Gosper–Zeilberger algorithm. However, what the Gosper–Zeilberger algorithm does is, for a given sum $S(n) = \sum_k F(n, k)$, it finds polynomials $A_0(n), A_1(n), \dots, A_d(n)$ and an expression $G(n, k)$ (which is, in fact, a rational multiple of $F(n, k)$), such that

$$A_0(n)F(n, k) + A_1(n)F(n+1, k) + \dots + A_d(n)F(n+d, k) = G(n, k+1) - G(n, k) \quad [71]$$

for some d . Because of the properties of $F(n, k)$ and $G(n, k)$, which are part of the theory, this is an identity which can be directly verified by clearing all common factors and checking the remaining identity between rational functions in n and k . However, we

may now sum both sides of [71] over k to obtain a recurrence of the form [70].

Algorithms for multiple sums are also available (see “Further reading”). They follow ideas by Wilf and Zeilberger (1992) (of which a simplified version is presented in a Mohammed and Zeilberger preprint (see “Further reading”)); however, they run more quickly in capacity problems. Schneider (2005) is currently developing a very promising new algorithmic approach to the automatic treatment of multisums. See q-Special Functions and Statistical Mechanics and Combinatorial Problems.

See also: Classical Groups and Homogeneous Spaces; Compact Groups and Their Representations; Dimer Problems; Growth Processes in Random Matrix Theory; Ordinary Special Functions; q -Special Functions; Saddle Point Problems; Statistical Mechanics and Combinatorial Problems.

Further Reading

- <http://algo.inria.fr> This site includes, among its libraries, the Maple program `gdev`.
- Andrews GE (1976) The Theory of Partitions, *Encyclopedia of Mathematics and Its Applications*, vol. 2. (reprinted by Cambridge University Press, Cambridge, 1998). Reading: Addison–Wesley.
- Andrews GE, Askey RA, and Roy R (1999) In: Rota GC (ed.) *Special Functions*, *Encyclopedia of Mathematics and Its Applications*, vol. 71. Cambridge: Cambridge University Press.
- Ayoub R (1963) An Introduction to the Analytic Theory of Numbers. *Mathematical Surveys*, vol. 10, Providence, RI: American Mathematical Society.
- Bergeron F, Labelle G, and Leroux P (1998) *Combinatorial Species and Tree-Like Structures*. Cambridge: Cambridge University Press.
- Bousquet-Mélou M and Jehanne A (2005), Polynomial equations with one catalytic variable, algebraic series, and map enumeration. *Preprint*, arXiv:math.CO/0504018.
- Bressoud DM (1999) *Proofs and Confirmations – The Story of the Alternating Sign Matrix Conjecture*. Cambridge: Cambridge University Press.
- de Bruijn NG (1964) Pólya’s theory of counting. In: Beckenbach EF (ed.) *Applied Combinatorial Mathematics*, New York: Wiley, (reprinted by Krieger, Malabar, Florida, 1981).
- Comtet L (1974) *Advanced Combinatorics*. Dordrecht: Reidel.
- Dolbilin NP, Mishchenko AS, Shtan’ko MA, Shtogrin MI, and Zinoviev YuM (1996) Homological properties of dimer configurations for lattices on surfaces. *Functional Analysis and its Application* 30: 163–173.
- Feller W (1957) *An Introduction to Probability Theory and Its Applications*, vol. 1, 2nd edn. New York: Wiley.
- Fisher ME (1984) Walks, walls, wetting and melting. *Journal of Statistical Physics* 34: 667–729.
- Flajolet P and Sedgewick R, *Analytic Combinatorics*, book project, available at <http://algo.inria.fr>.
- Di Francesco P, Zinn-Justin P and Zuber J.-B. (2004), Determinant formulae for some tiling problems and application to fully packed loops, *Preprint*, arXiv:math-ph/0410002.
- Di Francesco P and Zinn-Justin P (2005), Quantum Knizhnik–Zamolodchikov equation, generalized Razumov–Stroganov sum rules and extended Joseph polynomials. *Preprint*, arXiv:math-ph/0508059.

- Galluccio A and Loeb M (1999) On the theory of Pfaffian orientations I. Perfect matchings and permanents. *Electronic Journal of Combinatorics* 6: Article #R6, 18 pp. <http://www.fmf.uni-lj.si> – website of Faculty of Mathematics of University of Ljubljana. A Mathematica implementation by Marko Petkovšek is available here.
- Gasper G and Rahman M (2004) *Basic Hypergeometric Series*, 2nd edn. Encyclopedia of Mathematics and Its Applications, vol. 96. Cambridge: Cambridge University Press.
- de Gier J (2005) Loops matchings and alternating-sign matrices. *Discrete Mathematics* 365–388.
- Humphreys JE (1990) *Reflection Groups and Coxeter Groups*. Cambridge: Cambridge University Press.
- Johansson K (2002) Non-intersecting paths, random tilings and random matrices. *Probability Theory and Related Fields* 123: 225–280.
- Kenyon R (2003) *An Introduction to the Dimer Model*, Lecture Notes for a Short Course at the ICTP, 2002; [arxiv:math.CO/0310326](http://arxiv.org/abs/math.CO/0310326).
- Koekoek R and Swarttouw RF, *The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue*, TU Delft, The Netherlands, 1998; on the www: <http://aw.twi.tudelft.nl>.
- Krattenthaler C (1997) The enumeration of lattice paths with respect to their number of turns. In: Balakrishnan N (ed.) *Advances in Combinatorial Methods and Applications to Probability and Statistics*, pp. 29–58. Boston: Birkhäuser.
- Krattenthaler C (2003), Asymptotics for random walks in alcoves of affine Weyl groups. *Preprint*, [arxiv:math.CO/0301203](http://arxiv.org/abs/math.CO/0301203).
- Krattenthaler C (2005a), Watermelon configurations with wall interaction: exact and asymptotic results. *Preprint*, [arxiv:math.CO/0506323](http://arxiv.org/abs/math.CO/0506323).
- Krattenthaler C (2005b) Advanced determinant calculus: a complement. *Linear Algebra Applications* 411: 68–166.
- Krattenthaler C, Guttman AJ, and Viennot XG (2000) Vicious walkers, friendly walkers and Young tableaux II: with a wall. *Journal of Physics A: Mathematical and General* 33: 8835–8866.
- Kuperberg G (1998) An exploration of the permanent-determinant method. *Electronic Journal of Combinatorics* 5: Article #R46, 34 pp.
- Labelle G and Lamathe C (2004) A shifted asymmetry index series. *Advances in Applied Mathematics* 32: 576–608.
- Mohammed M and Zeilberger D (2005) Multi-variable Zeilberger and Almkvist–Zeilberger algorithms and the sharpening of Wilf–Zeilberger theory. *Advanced Applications in Mathematics* (to appear).
- Mohanty SG (1979) *Lattice Path Counting and Applications*. New York: Academic Press.
- Odlitzko AM (1995) Asymptotic enumeration methods. In: Graham RL, Grötschel M, and Lovász L (eds.) *Handbook of Combinatorics*, pp. 1063–1229. Amsterdam: Elsevier.
- Pemantle R and Wilson MC, Twenty combinatorial examples of asymptotics derived from multivariate generating functions. *Preprint*, available at <http://www.cs.auckland.ac.nz>.
- Petkovšek M, Wilf H, and Zeilberger D (1996) *A = B* Wellesley: Peters AK.
- <http://www.mat.univie.ac.at> – Website of Faculty of Mathematics, University of Vienna. It provides the manual of the Mathematica package HYP.
- Propp J (1999) Enumeration of matchings: problems and progress. In: Billera L, Björner A, Greene C, Simion R, and Stanley RP (eds.) *New Perspectives in Algebraic Combinatorics*, Mathematical Sciences Research Institute Publications, vol. 38, pp. 255–291. Cambridge: Cambridge University Press.
- Razumov AV and Stroganov YG (2005) Enumeration of quarter-turn symmetric alternating-sign matrices of odd order. *Preprint*, [arxiv:math-ph/0507003](http://arxiv.org/abs/math-ph/0507003).
- Robertson N, Seymour PD, and Thomas R (1999) Permanents, Pfaffian orientations, and even directed circuits. *Annals of Mathematics* 150(2): 929–975.
- Schneider C (2005) A new Sigma approach to multi-summation. *Advances in Applied Mathematics* 34(4): 740–767.
- Slater LJ (1966) *Generalized Hypergeometric Functions*. Cambridge: Cambridge University Press.
- Stanley RP (1986) *Enumerative Combinatorics*, Pacific Grove, CA: Wadsworth & Brooks/Cole, (reprinted by Cambridge University Press, Cambridge, 1998).
- Stanley RP (1999) *Enumerative Combinatorics*, vol. 2. Cambridge: Cambridge University Press.
- Szegő G (1959) *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23. New York. Providence RI: American Mathematical Society.
- Tesler G (2000) Matchings in graphs on non-oriented surfaces. *Journal of Combinatorial Theory Series B* 78: 198–231. <http://www.risc.uni.linz.ac.at> – website of RISC (Research Institute for Symbolic Computation). Mathematica implementations written by Peter Paule and Markus Schorn, and Axel Riese and Kurt Wegschaider are available here.
- <http://www.math.rutgers.edu> – website of Department of Mathematics, Rutgers University. Computer implementations written by D Zeilberger are available here.
- Viennot X and James W Heaps of segments, q-Bessel functions in square lattice enumeration and applications in quantum gravity. *Preprint*.
- Wilf HS and Zeilberger D (1992) An algorithmic proof theory for hypergeometric (ordinary and “q”) multisum/integral identities. *Inventiones Mathematicae* 108: 575–633.
- Zeilberger D (2005) Deconstructing the Zeilberger algorithm. *Journal of Difference Equations and Applications* 11: 851–856.

Compact Groups and Their Representations

A Kirillov, University of Pennsylvania, Philadelphia, PA, USA

A Kirillov, Jr., Stony Brook University, Stony Brook, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

In this article, we describe the structure and representation theory of compact Lie groups. Throughout the article, G is a compact real Lie

group with Lie algebra \mathfrak{g} . Unless otherwise stated, G is assumed to be connected. The word “group” will always mean a “Lie group” and the word “subgroup” will mean a closed Lie subgroup. The notation $\text{Lie}(H)$ stands for the Lie algebra of a Lie group H . We assume that the reader is familiar with the basic facts of the theory of Lie groups and Lie algebras, which can be found in Lie Groups: General Theory, or in the books listed in the bibliography.

Examples of Compact Lie Groups

Examples of compact groups include

- finite groups,
- quotient groups $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$, or more generally, V/L , where V is a finite-dimensional real vector space and L is a lattice in V , that is, a discrete subgroup generated by some basis in V – groups of this type are called “tori”; it is known that every commutative connected compact group is a torus;
- unitary groups $U(n)$ and special unitary groups $SU(n), n \geq 2$;
- orthogonal groups $O(n)$ and $SO(n), n \geq 3$; and
- the groups $U(n, \mathbb{H}), n \geq 1$, of unitary quaternionic transformations, which are isomorphic to $Sp(n) := Sp(n, \mathbb{C}) \cap SU(2n)$.

The groups $O(n)$ have two connected components, one of which is $SO(n)$. The groups $SU(n)$ and $Sp(n)$ are connected and simply connected.

The groups $SO(n)$ are connected but not simply connected: for $n \geq 3$, the fundamental group of $SO(n)$ is \mathbb{Z}_2 . The universal cover of $SO(n)$ is a simply connected compact Lie group denoted by $Spin(n)$. For small n , we have isomorphisms: $Spin(3) \simeq SU(2)$, $Spin(4) \simeq SU(2) \times SU(2)$, $Spin(5) \simeq Sp(4)$, and $Spin(6) \simeq SU(4)$.

Relation to Semisimple Lie Algebras and Lie Groups

Reductive Groups

A Lie algebra \mathfrak{g} is called

- “simple” if it is nonabelian and has no ideals different from $\{0\}$ and \mathfrak{g} itself;
- “semisimple” if it is a direct sum of simple ideals; and
- “reductive” if it is a direct sum of semisimple and commutative ideals.

We call a connected Lie group G “simple” or “semisimple” if $\text{Lie}(G)$ has this property.

Theorem 1 *Let G be a connected compact Lie group and $\mathfrak{g} = \text{Lie}(G)$. Then*

- (i) *The Lie algebra $\mathfrak{g} = \text{Lie}(G)$ is reductive: $\mathfrak{g} = \alpha \oplus \mathfrak{g}'$, where α is abelian and $\mathfrak{g}' = [\mathfrak{g}, \mathfrak{g}]$ is semisimple.*
- (ii) *The group G can be written in the form $G = (A \times K)/Z$, where A is a torus, K is a connected, simply connected compact semisimple Lie group, and Z is a finite central subgroup in $A \times K$.*
- (iii) *If G is simply connected, it is a product of simple compact Lie groups.*

The proof of these results is based on the fact that the Killing form of \mathfrak{g} is negative semidefinite.

Example 1 The group $U(n)$ contains as the center the subgroup C of scalar matrices. The quotient group $U(n)/C$ is simple and isomorphic to $SU(n)/\mathbb{Z}_n$. The presentation of Theorem 1 in this case is

$$\begin{aligned} U(n) &= (\mathbb{T}^1 \times SU(n))/\mathbb{Z}_n \\ &= (C \times SU(n))/(C \cap SU(n)) \end{aligned}$$

For the group $SO(4)$ the presentation is $(SU(2) \times SU(2))/\{\pm(1 \times 1)\}$.

This theorem effectively reduces the study of the structure of connected compact groups to the study of simply connected compact simple Lie groups.

Complexification of a Compact Lie Group

Recall that for a real Lie algebra \mathfrak{g} , its complexification is $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes \mathbb{C}$ with obvious commutator. It is also well known that $\mathfrak{g}_{\mathbb{C}}$ is semisimple or reductive iff \mathfrak{g} is semisimple or reductive, respectively. There is a subtlety in the case of simple algebras: it is possible that a real Lie algebra is simple, but its complexification $\mathfrak{g}_{\mathbb{C}}$ is only semisimple. However, this problem never arises for Lie algebras of compact groups: if \mathfrak{g} is a Lie algebra of a real compact Lie group, then \mathfrak{g} is simple if and only if $\mathfrak{g}_{\mathbb{C}}$ is simple.

The notion of complexification for Lie groups is more delicate.

Definition 1 Let G be a connected real Lie group with Lie algebra \mathfrak{g} . A complexification of G is a connected complex Lie group $G_{\mathbb{C}}$ (i.e., a complex manifold with a structure of a Lie group such that group multiplication is given by a complex analytic map $G_{\mathbb{C}} \times G_{\mathbb{C}} \rightarrow G_{\mathbb{C}}$), which contains G as a closed subgroup, and such that $\text{Lie}(G_{\mathbb{C}}) = \mathfrak{g}_{\mathbb{C}}$. In this case, we will also say that G is a real form of $G_{\mathbb{C}}$.

It is not obvious why such a complexification exists at all; in fact, for arbitrary real group it may not exist. However, for compact groups we do have the following theorem.

Theorem 2 *Let G be a connected compact Lie group. Then it has a unique complexification $G_{\mathbb{C}} \supset G$. Moreover, the following properties hold:*

- (i) *The inclusion $G \subset G_{\mathbb{C}}$ is a homotopy equivalence. In particular, $\pi_1(G) = \pi_1(G_{\mathbb{C}})$ and the quotient space $G_{\mathbb{C}}/G$ is contractible.*
- (ii) *Every complex finite-dimensional representation of G can be uniquely extended to a complex analytic representation of $G_{\mathbb{C}}$.*

Since the Lie algebra of a compact Lie group G is reductive, we see that $G_{\mathbb{C}}$ must be reductive; if G is semisimple or simple, then so is $G_{\mathbb{C}}$. The natural question is whether every complex reductive group can be obtained in this way. The following theorem gives a partial answer.

Theorem 3 *Every connected complex semisimple Lie group H has a compact real form: there is a compact real subgroup $K \subset H$ such that $H = K_{\mathbb{C}}$. Moreover, such a compact real form is unique up to conjugation.*

Example 2

- (i) The unitary group $U(n)$ is a compact real form of the group $GL(n, \mathbb{C})$.
- (ii) The orthogonal group $SO(n)$ is a compact real form of the group $SO(n, \mathbb{C})$.
- (iii) The group $Sp(n)$ is a compact real form of the group $Sp(n, \mathbb{C})$.
- (iv) The universal cover of $GL(n, \mathbb{C})$ has no compact real form.

These results have a number of important applications. For example, they show that study of representations of a semisimple complex group H can be replaced by the study of representations of its compact form; in particular, every representation is completely reducible (this argument is known as Weyl’s unitary trick).

Classification of Simple Compact Lie Groups

Theorem 1 essentially reduces such classification to classification of simply connected simple compact groups, and **Theorems 2 and 3** reduce it to the classification of simple complex Lie algebras. Since the latter is well known, we get the following result.

Theorem 4 *Let G be a connected, simply connected simple compact Lie group. Then $\mathfrak{g}_{\mathbb{C}}$ must be a simple complex Lie algebra and thus can be described by a Dynkin diagram of one the following types: $A_n, B_n, C_n, D_n, E_6, E_7, E_8, F_4, G_2$.*

Conversely, for each Dynkin diagram in the above list, there exists a unique, up to isomorphism, simply connected simple compact Lie group whose Lie algebra is described by this Dynkin diagram.

For types A_n, \dots, D_n , the corresponding compact Lie groups are well-known classical groups shown in the table below:

$A_n, n \geq 1$	$B_n, n \geq 2$	$C_n, n \geq 3$	$D_n, n \geq 4$
$SU(n+1)$	$Spin(2n+1)$	$Sp(n)$	$Spin(2n)$

The restrictions on n in this table are made to avoid repetitions which appear for small values of n . Namely, $A_1 = B_1 = C_1$, which gives $SU(2) = Spin(3) = Sp(1)$; $D_2 = A_1 \cup A_1$, which gives $Spin(4) = SU(2) \times SU(2)$; $B_2 = C_2$, which gives $SO(5) = Sp(4)$; and $A_3 = D_3$, which gives $SU(4) = Spin(6)$. Other than that, all entries are distinct.

Exceptional groups E_6, \dots, G_2 also admit explicit geometric and algebraic descriptions which are related to the exceptional nonassociative algebra \mathbb{O} of the so-called octonions (or Cayley numbers). For example, the compact group of type G_2 can be defined as a subgroup of $SO(7)$ which preserves an almost-complex structure on S^6 . It can also be described as the subgroup of $GL(7, \mathbb{R})$ which preserves one quadratic and one cubic form, or, finally, as a group of all automorphisms of \mathbb{O} .

Maximal Tori

Main Properties

In this section, G is a compact connected Lie group.

Definition 2 A “maximal torus” in G is a maximal connected commutative subgroup $T \subset G$.

The following theorem lists the main properties of maximal tori.

Theorem 5

- (i) *For every element $g \in G$, there exists a maximal torus $T \ni g$.*
- (ii) *Any two maximal tori in G are conjugate.*
- (iii) *If $g \in G$ commutes with all elements of a maximal torus T , then $g \in T$.*
- (iv) *A connected subgroup $H \subset G$ is a maximal torus iff the Lie algebra $Lie(H)$ is a maximal abelian subalgebra in $Lie(G)$.*

Example 3 Let $G = U(n)$. Then the set T of diagonal unitary matrices is a maximal torus in G ; moreover, every maximal torus is of this form after a suitable unitary change of basis. In particular, this implies that every element in G is conjugate to a diagonal matrix.

Example 4 Let $G = SO(3)$. Then the set D of diagonal matrices is a maximal commutative subgroup in G , but not a torus. Here D consists of four elements and is not connected.

Maximal Tori and Cartan Subalgebras

The study of maximal tori in compact Lie groups is closely related to the study of Cartan subalgebras in reductive complex Lie algebras. For convenience of readers, we briefly recall the appropriate definitions

here; details can be found in Serre (2001) or in Lie Groups: General Theory.

Definition 3 Let \mathfrak{a} be a complex reductive Lie algebra. A Cartan subalgebra $\mathfrak{h} \subset \mathfrak{a}$ is a maximal commutative subalgebra consisting of semisimple elements.

Note that for general Lie algebras Cartan subalgebra is defined in a different way; however, for reductive algebras the definition given above is equivalent to the standard one.

A choice of a Cartan subalgebra gives rise to the so-called root decomposition: if $\mathfrak{h} \subset \mathfrak{a}$ is a Cartan subalgebra in a complex reductive Lie algebra, then we can write

$$\mathfrak{a} = \mathfrak{h} \oplus \left(\bigoplus_{\alpha \in R} \mathfrak{a}_\alpha \right) \tag{1}$$

where

$$\begin{aligned} \mathfrak{a}_\alpha &= \{x \in \mathfrak{a} \mid \text{ad } h \cdot x = \langle \alpha, h \rangle x \ \forall h \in \mathfrak{h}\} \\ R &= \{\alpha \in \mathfrak{h}^* - \{0\} \mid \mathfrak{a}_\alpha \neq 0\} \subset \mathfrak{h}^* \end{aligned}$$

The set R is called the “root system” of \mathfrak{a} with respect to Cartan subalgebra \mathfrak{h} ; elements $\alpha \in R$ are called “roots.” We will also frequently use elements $\alpha^\vee \in \mathfrak{h}$ defined by $\langle \alpha^\vee, \beta \rangle = 2\langle \alpha, \beta \rangle / \langle \alpha, \alpha \rangle$ where $\langle \cdot, \cdot \rangle$ is a nondegenerate invariant bilinear form on \mathfrak{a}^* and $\langle \cdot, \cdot \rangle$ is the pairing between \mathfrak{a} and \mathfrak{a}^* . It can be shown that so defined α^\vee does not depend on the choice of the form $\langle \cdot, \cdot \rangle$.

Theorem 6 Let G be a connected compact Lie group with Lie algebra \mathfrak{g} , and let $T \subset G$ be a maximal torus in G , $\mathfrak{t} = \text{Lie}(T) \subset \mathfrak{g}$. Let $\mathfrak{g}_\mathbb{C}, G_\mathbb{C}$ be the complexification of \mathfrak{g}, G as in Theorem 2.

Let $\mathfrak{h} = \mathfrak{t}_\mathbb{C} \subset \mathfrak{g}_\mathbb{C}$. Then \mathfrak{h} is a Cartan subalgebra in $\mathfrak{g}_\mathbb{C}$, and the corresponding root system $R \subset \mathfrak{h}^*$. Conversely, every Cartan subalgebra in $\mathfrak{g}_\mathbb{C}$ can be obtained as $\mathfrak{t}_\mathbb{C}$ for some maximal torus $T \subset G$.

Weights and Roots

Let G be semisimple. Recall that the root lattice $Q \subset \mathfrak{t}^*$ is the abelian group generated by roots $\alpha \in R$, and let the coroot lattice $Q^\vee \subset \mathfrak{t}$ be the abelian group generated by coroots $\alpha^\vee, \alpha \in R$. Define also the weight and coweight lattices by

$$\begin{aligned} P &= \{\lambda \mid \langle \alpha^\vee, \lambda \rangle \in \mathbb{Z} \ \forall \alpha \in R\} \subset \mathfrak{t}^* \\ P^\vee &= \{t \mid \langle t, \alpha \rangle \in \mathbb{Z} \ \forall \alpha \in R\} \subset \mathfrak{t}, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the pairing between \mathfrak{t} and the dual vector space \mathfrak{t}^* .

It follows from the definition of root system that we have inclusions

$$\begin{aligned} Q &\subset P \subset \mathfrak{t}^* \\ Q^\vee &\subset P^\vee \subset \mathfrak{t} \end{aligned} \tag{2}$$

Both P, Q are lattices in \mathfrak{t}^* ; thus, the index $(P : Q)$ is finite. It can be computed explicitly: if α_i is a basis of the root system, then the fundamental weights ω_i defined by

$$\langle \alpha_i^\vee, \omega_j \rangle = \delta_{ij}$$

form a basis of P . The simple roots α_i are related to fundamental weights ω_j by the Cartan matrix $A: \alpha_i = \sum A_{ij} \omega_j$. Therefore, $(P : Q) = (P^\vee : Q^\vee) = |\det A|$.

Definitions of P, Q, P^\vee, Q^\vee also make sense when \mathfrak{g} is reductive but not semisimple. However, in this case they are no longer lattices: $\text{rk } Q < \dim \mathfrak{t}^*$, and P is not discrete.

We can now give more precise information about the structure of the maximal torus.

Lemma 1 Let T be a compact connected commutative Lie group, and $\mathfrak{t} = \text{Lie}(T)$ its Lie algebra. Then the exponential map is surjective and preimage of unit is a lattice $L \subset \mathfrak{t}$. There is an isomorphism of Lie groups

$$\exp : \mathfrak{t}/L \rightarrow T$$

In particular, $T \simeq \mathbb{R}^r / \mathbb{Z}^r = \mathbb{T}^r, r = \dim T$.

Let $X(T) \subset \mathfrak{t}^*$ be the lattice dual to $(2\pi i)^{-1}L$:

$$X(T) = \{\lambda \in \mathfrak{t}^* \mid \langle \lambda, l \rangle \in 2\pi i \mathbb{Z} \ \forall l \in L\} \tag{3}$$

It is called the “character lattice” for T (see the subsection “Examples of representations”).

Theorem 7 Let G be a compact connected Lie group, and let $T \subset G$ be a maximal torus in G .

Then $Q \subset X(T) \subset P$. Moreover, the group G is uniquely determined by the Lie algebra \mathfrak{g} and the lattice $X(T) \in \mathfrak{t}^*$ which can be any lattice between Q and P .

Corollary For a given complex semisimple Lie algebra \mathfrak{a} , there are only finitely many (up to isomorphism) compact connected Lie groups G with $\mathfrak{g}_\mathbb{C} = \mathfrak{a}$.

The largest of them is the simply connected group, for which $T = \mathfrak{t}/2\pi i Q^\vee, X(T) = P$; the smallest is the so-called “adjoint group,” for which $T = \mathfrak{t}/2\pi i P^\vee, X(T) = Q$.

Example 5 Let $G = U(n)$. Then $\mathfrak{it} = \{\text{real diagonal matrices}\}$. Choosing the standard basis of matrix

units in it, we identify it $\simeq \mathbb{R}^n$, which also allows us to identify it $\simeq \mathbb{R}^n$. Under this identification,

$$\begin{aligned} Q &= \{(\lambda_1, \dots, \lambda_n) \mid \lambda_i \in \mathbb{Z}, \sum \lambda_i = 0\} \\ P &= \{(\lambda_1, \dots, \lambda_n) \mid \lambda_i \in \mathbb{R}, \lambda_i - \lambda_j \in \mathbb{Z}\} \\ X(T) &= \mathbb{Z}^n \end{aligned}$$

Note that Q, P are not lattices: $Q \simeq \mathbb{Z}^{n-1}$, $P \simeq \mathbb{R} \times \mathbb{Z}^{n-1}$.

Now let $G = \text{SU}(n)$. Then $\text{it}^* = \mathbb{R}^n / \mathbb{R} \cdot (1, \dots, 1)$, and Q, P are the images of Q, P for $G = \text{U}(n)$ in this quotient. In this quotient they are lattices, and $(P:Q) = n$. The character lattice in this case is $X(T) = P$, since $\text{SU}(n)$ is simply connected. The adjoint group is $\text{PSU}(n) = \text{SU}(n)/C$, where $C = \{\lambda \cdot \text{id} \mid \lambda^n = 1\}$ is the center of $\text{SU}(n)$.

Weyl Group

Let us fix a maximal torus $T \subset G$. Let $N(T) \subset G$ be the normalizer of T in G : $N(T) = \{g \in G \mid gTg^{-1} = T\}$. For any $g \in N(T)$ the transformation $A(g): t \mapsto gTg^{-1}$ is an automorphism of T . According to Theorem 5, this automorphism is trivial iff $g \in T$. So in fact, it is the quotient group $N(T)/T$ which acts on T .

Definition 4 The group $W = N(T)/T$ is called the “Weyl group” of G .

Since the Weyl group acts faithfully on \mathfrak{t} and \mathfrak{t}^* , it is common to consider W as a subgroup in $\text{GL}(\mathfrak{t}^*)$. It is known that W is finite.

The Weyl group can also be defined in terms of Lie algebra \mathfrak{g} and its complexification $\mathfrak{g}_{\mathbb{C}}$.

Theorem 8 *The Weyl group coincides with the subgroup in $\text{GL}(\mathfrak{t}^*)$ generated by reflections $s_{\alpha}: x \mapsto x - (2(\alpha, x))/(\alpha, \alpha)\alpha$, $\alpha \in R$, where, as before, $(,)$ is a nondegenerate invariant bilinear form on \mathfrak{g}^* .*

Theorem 9

- (i) Two elements $t_1, t_2 \in T$ are conjugate in G iff $t_2 = w(t_1)$ for some $w \in W$.
- (ii) There exists a natural homeomorphism of quotient spaces $G/\text{Ad}G \simeq T/W$, where $\text{Ad}G$ stands for action of G on itself by conjugation. (Note, however, that these quotient spaces are not manifolds: they have singularities.)
- (iii) Let us call a function f on G central if $f(hgb^{-1}) = f(g)$ for any $g, b \in G$. Then the restriction map gives an isomorphism

$$\begin{aligned} &\{\text{continuous central functions on } G\} \\ &\simeq \{W\text{-invariant continuous functions on } T\} \end{aligned}$$

Example 6 Let $G = \text{U}(n)$. The set of diagonal unitary matrices is a maximal torus, and the Weyl group is the symmetric group S_n acting on diagonal matrices by permutations of entries. In this case, Theorem 9 shows that if $f(U)$ is a central function of a unitary matrix, then $f(U) = \tilde{f}(\lambda_1, \dots, \lambda_n)$, where λ_i are eigenvalues of U and \tilde{f} is a symmetric function in n variables.

Representations of Compact Groups

Basic Notions

By a representation of G we understand a pair (π, V) , where V is a complex vector space and π is a continuous homomorphism $G \rightarrow \text{Aut}(V)$. This notation is often shortened to π or V . In this article, we only consider finite-dimensional (f.d.) representations; in this case, the homomorphism π is automatically smooth and even real-analytic.

We associate to any f.d. representation (π, V) of G the representation (π_*, V) of the Lie algebra $\mathfrak{g} = \text{Lie}(G)$ which is just the derivative of the map $\pi: G \rightarrow \text{Aut}V$ at the unit point $e \in G$. In terms of the exponential map, we have the following commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{\pi} & \text{Aut}V \\ \text{exp} \uparrow & & \uparrow \text{exp} \\ \mathfrak{g} & \xrightarrow{\pi_*} & \text{End}V \end{array}$$

Choosing a basis in V , we can write the operators $\pi(g)$ and $\pi_*(X)$ in matrix form and consider π and π_* as matrix-valued functions on G and \mathfrak{g} . The diagram above means that

$$\pi(\text{exp } X) = e^{\pi_*(X)} \tag{4}$$

Recall that if G is connected, simply connected, then every representation of \mathfrak{g} can be uniquely lifted to a representation of G . Thus, classification of representations of connected simply connected Lie groups is equivalent to the classification of representations of Lie algebras.

Let (π_1, V_1) and (π_2, V_2) be two representations of the same group G . An operator $A \in \text{Hom}(V_1, V_2)$ is called an “intertwining operator,” or simply an “intertwiner,” if $A \circ \pi_1(g) = \pi_2(g) \circ A$ for all $g \in G$. Two representations are called “equivalent” if they admit an invertible intertwiner. In this case, using an appropriate choice of bases, we can write π_1 and π_2 by the same matrix-valued function.

Let (π, V) be a representation of G . If all operators $\pi(g), g \in G$, preserve a subspace $V_1 \subset V$, then the restrictions $\pi_1(g) = \pi(g)|_{V_1}$ define a “subrepresentation” (π_1, V_1) of (π, V) . In this case, the quotient space $V_2 = V/V_1$ also has a canonical structure of a representation, called the “quotient representation.”

A representation (π, V) is called “reducible” if it has a nontrivial (different from V and $\{0\}$) subrepresentation. Otherwise it is called “irreducible.”

We call representation (π, V) “unitary” if V is a Hilbert space and all operators $\pi(g), g \in G$, are unitary, that is, given by unitary matrices in any orthonormal basis. We use a short term “unirrep” for a “unitary irreducible representation.”

Main Theorems

The following simple but important result was one of the first discoveries in representation theory. It holds for representations of any group, not necessarily compact.

Theorem 10 (Schur lemma). *Let $(\pi_i, V_i), i = 1, 2$, be any two irreducible finite-dimensional representations of the same group G . Then any intertwiner $A: V_1 \rightarrow V_2$ is either invertible or zero.*

Corollary 1 *If V is an irreducible f.d. representation, then any intertwiner $A: V \rightarrow V$ is scalar: $A = c \cdot \text{id}, c \in \mathbb{C}$.*

Corollary 2 *Every irreducible representation of a commutative group is one dimensional.*

The following theorem is one of the fundamental results of the representation theory of compact groups. Its proof is based on the technique of invariant integrals on a compact group, which will be discussed in the next section.

Theorem 11

- (i) *Any f.d. representation of a compact group is equivalent to a unitary representation.*
- (ii) *Any f.d. representation is completely reducible: it can be decomposed into direct sum*

$$V = \bigoplus n_i V_i$$

where V_i are pairwise nonequivalent unirreps. Numbers $n_i \in \mathbb{Z}_+$ are called “multiplicities.”

Examples of Representations

The representation theory looks rather different for abelian (i.e., commutative) and nonabelian groups. Here we consider two simplest examples of both kinds.

Our first example is a one-dimensional compact connected Lie group. Topologically, it is a circle which we realize as a set $\mathbb{T} \simeq \text{U}(1)$ of all complex numbers t with absolute value 1.

Every unirrep of \mathbb{T} is one dimensional; thus, it is just a continuous multiplicative map π of \mathbb{T} to itself. It is well known that every such map has the form

$$\pi_k(t) = t^k \quad \text{for some } k \in \mathbb{Z}$$

The collection of all unirreps of \mathbb{T} is itself a group, called “Pontrjagin dual” of \mathbb{T} and denoted by $\widehat{\mathbb{T}}$. This group is isomorphic to \mathbb{Z} .

By **Theorem 11**, any f.d. representation π of \mathbb{T} is equivalent to a direct sum of one-dimensional unirreps. So, an equivalence class of π is defined by the multiplicity function μ on $\widehat{\mathbb{T}} = \mathbb{Z}$ taking non-negative values:

$$\pi \simeq \sum_{k \in \mathbb{Z}} \mu(k) \cdot \pi_k$$

The many-dimensional case of compact connected abelian Lie group can be treated in a similar way. Let T be a torus, that is, an abelian compact group, $\mathfrak{t} = \text{Lie}(T)$. Then every irreducible representation of T is one dimensional and thus is defined by a group homomorphism $\chi: T \rightarrow \mathbb{T}^1 = \text{U}(1)$. Such homomorphisms are called “characters” of T . One easily sees that such characters themselves form a group (Pontrjagin dual of T). If we denote by L the kernel of the exponential map $\mathfrak{t} \rightarrow T$ (see Lemma 1), one easily sees that every character has a form

$$\chi(\exp(t)) = e^{(t, \lambda)}, \quad t \in \mathfrak{t}, \lambda \in X(T)$$

where $X(T) \subset \mathfrak{t}^*$ is the lattice defined by [3]. Thus, we can identify the group of characters \widehat{T} with $X(T)$. In particular, this shows that $\widehat{T} \simeq \mathbb{Z}^{\dim T}$.

The second example is the group $G = \text{SU}(2)$, the simplest connected, simply connected nonabelian compact Lie group. Topologically, G is a three-dimensional sphere since the general element of G is a matrix of the form

$$g = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \quad a, b \in \mathbb{C}, |a|^2 + |b|^2 = 1$$

Let V be two-dimensional complex vector space, realized by column vectors $\begin{pmatrix} u \\ v \end{pmatrix}$. The group G acts naturally on V . This action induces the representation Π of G in the space $S(V)$ of all polynomials in u, v . It is infinite dimensional, but has many f.d. subrepresentations. In particular, let $S^k(V)$, or simply S^k , be the space of all homogeneous polynomials of degree k . Clearly, $\dim S^k = k + 1$.

It turns out that the corresponding f.d. representations $(\Pi_k, S^k), k \geq 0$, are irreducible, pairwise non-equivalent, and exhaust the set \widehat{G} of all unirreps.

Some particular cases are of special interest:

1. $k = 0$. The space V_0 consists of constant functions and Π_0 is the trivial one-dimensional representation: $\Pi_0(g) \equiv 1$.
2. $k = 1$. The space V_1 is identical to V and Π_1 is just the tautological representation $\pi(g) \equiv g$.
3. $k = 2$. The space V_2 is spanned by monomials u^2, uv, v^2 . The remarkable fact is that this

representation is equivalent to a real one. Namely, in the new basis

$$x = \frac{u^2 + v^2}{2}, \quad y = \frac{u^2 - v^2}{2i}, \quad z = iuv$$

we have

$$\Pi_2 \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} = \begin{pmatrix} \operatorname{Re}(a^2 + b^2) & 2\operatorname{Im}(ab) & \operatorname{Im}(b^2 - a^2) \\ 2\operatorname{Im}(a\bar{b}) & |a|^2 - |b|^2 & 2\operatorname{Re}(a\bar{b}) \\ \operatorname{Im}(a^2 + b^2) & 2\operatorname{Re}(ab) & \operatorname{Re}(a^2 - b^2) \end{pmatrix}$$

This formula defines a homomorphism $\Pi_2: \operatorname{SU}(2) \rightarrow \operatorname{SO}(3)$. It can be shown that this homomorphism is surjective, and its kernel is the subgroup $\{\pm 1\} \subset \operatorname{SU}(2)$:

$$1 \rightarrow \{\pm 1\} \hookrightarrow \operatorname{SU}(2) \xrightarrow{\Pi_2} \operatorname{SO}(3) \rightarrow 1$$

The simplest way to see it is to establish the equivalence of Π_2 with the adjoint representation of G in \mathfrak{g} . The corresponding intertwiner is

$$S^2 \ni (\alpha + i\gamma)u^2 + 2i\beta uv + (\alpha - i\gamma)v^2 \longleftrightarrow \begin{pmatrix} i\beta & \alpha + i\gamma \\ -\alpha + i\gamma & -i\beta \end{pmatrix} \in \mathfrak{g}$$

Note that $\operatorname{SU}(2)$ and $\operatorname{SO}(3)$ are the only compact groups associated with the Lie algebra $\mathfrak{sl}(2, \mathbb{C})$.

The group G contains the subgroup H of diagonal matrices, isomorphic to \mathbb{T}^1 . Consider the restriction of Π_n to \mathbb{T}^1 . It splits into the sum of unirreps π_k as follows:

$$\operatorname{Res}_{\mathbb{T}^1}^G \Pi_n = \sum_{s=0}^{s=\lfloor n/2 \rfloor} \pi_{n-2s}$$

The characters π_k which enter this decomposition are called the weights of Π_n . The collection of all weights (together with multiplicities) forms a multiset in $\widehat{\mathbb{T}}$ denoted by $P(\Pi_n)$ or $P(S^n)$.

Note the following features of this multiset:

1. $P(\Pi_n)$ is invariant under reflection $k \mapsto -k$.
2. All weights of Π_n are congruent modulo 2.
3. The nonequivalent unirreps have different multisets of weights.

Below we show how these features are generalized to all compact connected Lie groups.

Fourier Transform

Haar Measure and Invariant Integral

The important feature of compact groups is the existence of the so-called “invariant integral,” or “average.”

Theorem 12 For every compact Lie group G , there exists a unique measure dg on G , called “Haar measure,” which is invariant under left shifts $L_g: h \mapsto gh$ and satisfies $\int_G dg = 1$.

In addition, this measure is also invariant under right shifts $h \mapsto hg$ and under involution $h \mapsto h^{-1}$.

Invariance of the Haar measure implies that for every integrable function $f(g)$, we have

$$\int_G f(g) dg = \int_G f(hg) dg = \int_G f(gh) dg = \int_G f(g^{-1}) dg$$

For a finite group G , the integral with respect to the Haar measure is just averaging over the group:

$$\int_G f(g) dg = \frac{1}{|G|} \sum_{g \in G} f(g)$$

For compact connected Lie groups, the Haar measure is given by a differential form of top degree which is invariant under right and left translations.

For a torus $T^n = \mathbb{R}^n / \mathbb{Z}^n$ with real coordinates $\theta_k \in \mathbb{R}/\mathbb{Z}$ or complex coordinates $t_k = e^{2\pi i \theta_k}$, the Haar measure is $d^n \theta := d\theta_1 d\theta_2 \cdots d\theta_n$ or

$$d^n t := \prod_{k=1}^n \frac{dt_k}{2\pi i t_k}$$

In particular, consider a central function f (see Theorem 9). Since every conjugacy class contains elements of the maximal torus T (see Theorem 5), such a function is determined by its values on T , and the integral of a central function can be reduced to integration over T . The resulting formula is called “Weyl integration formula.” For $G = \operatorname{U}(n)$ it looks as follows:

$$\int_{\operatorname{U}(n)} f(g) dg = \frac{1}{n!} \int_T f(t) \prod_{i < j} |t_i - t_j|^2 d^n t$$

where T is the maximal torus consisting of diagonal matrices

$$t = \operatorname{diag}(t_1, \dots, t_n), \quad t_k = e^{2\pi i \theta_k}$$

and $d^n t$ is defined above.

Weyl integration formula for arbitrary compact group G can be found in Simon (1996) or Bump (2004, section 18).

The main applications of the Haar measure are the proof of complete reducibility theorem (Theorem 11) and orthogonality relations (see below).

Orthogonality Relations and Peter-Weyl Theorem

Let V_1, V_2 be unirreps of a compact group G . Taking any linear operator $A: V_1 \rightarrow V_2$ and averaging the expression $A(g) := \pi_2(g^{-1}) \circ A \circ \pi_1(g)$ over

G , we get an intertwining operator $\langle A \rangle = \int_G A(g)dg$. Comparing this fact with the Schur lemma, one obtains the following fundamental results.

Let (π, V) be any unirrep of a compact group G . Choose any orthonormal basis $\{v_k, 1 \leq k \leq \dim V\}$ in V and denote by t_{kl}^V , or t_{kl}^π , the function on G defined by

$$t_{kl}^V(g) = (\pi(g)v_l, v_k)$$

The functions t_{kl}^V are called “matrix elements” of the unirrep (π, V) .

Theorem 13 (Orthogonality relations)

- (i) *The matrix elements t_{kl}^V are pairwise orthogonal and have norm $(\dim V)^{-1/2}$ in $L^2(G, dg)$.*
- (ii) *The matrix elements corresponding to equivalent unirreps span the same subspace in $L^2(G, dg)$.*
- (iii) *The matrix elements of two nonequivalent unirreps are orthogonal.*
- (iv) *The linear span of all matrix elements of all unirreps is dense in $C(G), C^\infty(G)$, and in $L^2(G, dg)$ (generalized Peter–Weyl theorem).*

In particular, this theorem implies that the set \widehat{G} of equivalence classes of unirreps is countable. For an f.d. representation (π, V) we introduce the character of π as a function

$$\chi_\pi(g) = \text{tr} \pi(g) = \sum_{k=1}^{\dim V} t_{kk}^\pi(g) \quad [5]$$

It is obviously a central function on G .

Remark Traditionally, in representation theory the word “character” has two different meanings: (1) a multiplicative map from a group to $U(1)$, and (2) the trace of a representation operator $\pi(g)$. For one-dimensional representations both notions coincide.

From the orthogonality relations we get the following result.

Corollary *The characters of unirreps of G form an orthonormal basis in the subspace of central functions in $L^2(G, dg)$.*

Noncommutative Fourier Transform

The noncommutative Fourier transform on a compact group G is defined as follows. Let \widehat{G} denote the set of equivalence classes of unirreps of G . Choose for any $\lambda \in \widehat{G}$ a representation (π_λ, V_λ) of class λ and an orthonormal basis in V_λ . Denote by $d(\lambda)$ the dimension of V_λ .

We introduce the Hilbert space $L^2(\widehat{G})$ as the space of matrix-valued functions on \widehat{G} whose value at a point $\lambda \in \widehat{G}$ belongs to $\text{Mat}_{d(\lambda)}(\mathbb{C})$. The norm is defined as

$$\|F\|_{L^2(\widehat{G})}^2 = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \text{tr}(F(\lambda)F(\lambda)^*)$$

For a function f on G define its Fourier transform \widetilde{f} as a matrix-valued function on \widehat{G} :

$$\widetilde{f}(\lambda) = \int_G f(g^{-1})\pi_\lambda(g)dg$$

Note that in the case $G = \mathbb{T}^1$ this transform associates to a function f the set of its Fourier coefficients. In general this transform keeps some important features of Fourier coefficients.

Theorem 14

- (i) *For a function $f \in L^1(G, dg)$ the Fourier transform \widetilde{f} is well defined and bounded (by matrix norm) function on \widehat{G} .*
- (ii) *For a function $f \in L^1(G, dg) \cap L^2(G, dg)$ the following analog of the Plancherel formula holds:*

$$\begin{aligned} \|f\|_{L^2(G, dg)}^2 &:= \int_G |f(g)|^2 dg \\ &= \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \text{tr}(\widetilde{f}(\lambda)\widetilde{f}(\lambda)^*) =: \|\widetilde{f}\|_{L^2(\widehat{G})}^2 \end{aligned}$$

- (iii) *The following inversion formula expresses f in terms of \widetilde{f} :*

$$f(g) = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \text{tr}(\widetilde{f}(\lambda)\pi_\lambda(g))$$

- (iv) *The Fourier transform sends the convolution to the matrix multiplication:*

$$\widetilde{f_1 * f_2} = \widetilde{f_1} \cdot \widetilde{f_2}$$

where the convolution product $*$ is defined by

$$(f_1 * f_2)(h) = \int_G f_1(hg)f_2(g^{-1}) dg$$

Note the special case of the inversion formula for $g = e$:

$$f(e) = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \text{tr}(\widetilde{f}(\lambda)),$$

or

$$\delta(g) = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \chi_\lambda(g)$$

where $\delta(g)$ is Dirac’s delta-function: $\int_G f(g)\delta(g) dg = f(e)$. Thus, we get a presentation of Dirac’s delta-function as a linear combination of characters.

Classification of Finite-Dimensional Representations

In this section, we give a classification of unirreps of a connected compact Lie group G .

Weight Decomposition

Let G be a connected compact group with maximal torus T , and let (π, V) be a f.d. representation of G . Restricting it to T and using complete reducibility, we get the following result.

Theorem 15 *The vector space V can be written in the form*

$$V = \bigoplus_{\lambda \in X(T)} V_\lambda, \tag{6}$$

$$V_\lambda = \{v \in V \mid \pi_*(t)v = \langle \lambda, t \rangle v \ \forall t \in \mathfrak{t}\}$$

where $X(T)$ is the character group of T defined by [3]. The spaces V_λ are called “weight subspaces,” vectors $v \in V_\lambda$ – “weight vectors” of weight λ . The set

$$P(V) = \{\lambda \in X(T) \mid V_\lambda \neq \{0\}\} \tag{7}$$

is called the “set of weights” of π , or the “spectrum” of $\text{Res}_T^G \pi$, and

$$\text{mult}_{(\pi, V)}(\lambda) := \dim V_\lambda$$

is called the “multiplicity” of λ in V .

The next theorem easily follows from the definition of the Weyl group.

Theorem 16 *For any f.d. representation V of G , the set of weights with multiplicities is invariant under the action of the Weyl group:*

$$w(P(V)) = P(V), \quad \text{mult}_{(\pi, V)}(\lambda) = \text{mult}_{(\pi, V)}(w(\lambda))$$

for any $w \in W$.

Classification of Unirreps

Recall that R is the root system of $\mathfrak{g}_\mathbb{C}$. Assume that we have chosen a basis of simple roots $\alpha_1, \dots, \alpha_r \subset R$. Then $R = R_+ \cup R_-$; roots $\alpha \in R_+$ can be written as a linear combination of simple roots with positive coefficients, and $R_- = -R_+$.

A (not necessarily f.d.) representation of $\mathfrak{g}_\mathbb{C}$ is called a “highest-weight representation” if it is generated by a single vector $v \in V_\lambda$ (the highest-weight vector) such that $\mathfrak{g}_\alpha v = 0$ for all positive roots $\alpha \in R_+$.

It can be shown that for every $\lambda \in X(T)$, there is a unique irreducible highest-weight representation of $\mathfrak{g}_\mathbb{C}$ with highest weight λ , which is denoted $L(\lambda)$.

However, this representation can be infinite dimensional; moreover, it may not be possible to lift it to a representation of G .

Definition 5 A weight $\lambda \in X(T)$ is called “dominant” if $\langle \lambda, \alpha_i^\vee \rangle \in \mathbb{Z}_+$ for any simple root α_i . The set of all dominant weights is denoted by $X_+(T)$.

Theorem 17

- (i) *All weights of $L(\lambda)$ are of the form $\mu = \lambda - \sum n_i \alpha_i$, $n_i \in \mathbb{Z}_+$.*
- (ii) *Let $\lambda \in X_+$. Then the irreducible highest-weight representation $L(\lambda)$ is f.d. and lifts to a representation of G .*
- (iii) *Every irreducible f.d. representation of G is of the form $L(\lambda)$ for some $\lambda \in X_+$.*

Thus, we have a bijection $\{\text{unirreps of } G\} \leftrightarrow X_+$.

Example 7 Let $G = \text{SU}(2)$. There is a unique simple root α and the unique fundamental weight ω , related by $\alpha = 2\omega$. Therefore, $X_+ = \mathbb{Z}_+ \cdot \omega$ and unirreps are indexed by non-negative integers. The representation with highest weight $k \cdot \omega$ is precisely the representation Π_k constructed in the subsection “Examples of representations.”

Example 8 Let $G = \text{U}(n)$. Then $X = \mathbb{Z}^n$, and $X_+ = \{(\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n \mid \lambda_1 \geq \dots \geq \lambda_n\}$. Such objects are well known in combinatorics: if we additionally assume that $\lambda_n \geq 0$, then such dominant weights are in bijection with partitions with n parts. They can also be described by “Young diagrams” with n rows (see [Fulton and Harris \(1991\)](#)).

Explicit Construction of Representations

In addition to description of unirreps as highest-weight representations, they can also be constructed in other ways. In particular, they can be defined analytically as follows. Let $B = \text{HN}_+$ be the Borel subgroup in $G_\mathbb{C}$; here $H = \exp \mathfrak{h}$, $N_+ = \exp \sum_{\alpha \in R_+} (\mathfrak{g}_\mathbb{C})_\alpha$. For $\lambda \in \mathfrak{h}^*$, let $\chi_\lambda : B \rightarrow \mathbb{C}^\times$ be a multiplicative map defined by

$$\chi_\lambda(hn) = e^{\langle \lambda, h \rangle} \tag{8}$$

Theorem 18 (Cartan–Borel–Weil). *Let $\lambda \in X(T)$. Denote by $V(\lambda)$ the space of complex-analytic functions on $G_\mathbb{C}$ which satisfy the following transformation property:*

$$f(gb) = \chi_\lambda^{-1}(b)f(g), \quad g \in G_\mathbb{C}, \ b \in B$$

The group $G_\mathbb{C}$ acts on $V(\lambda)$ by left shifts:

$$(\pi(g)f)(h) = f(g^{-1}h) \tag{9}$$

Then

- (i) $V(\lambda) \neq \{0\}$ iff $-\lambda \in X_+$.
- (ii) If $-\lambda \in X_+$, the representation of G in $V(\lambda)$ is equivalent to $L(w_0(\lambda))$, where $w_0 \in W$ is the unique element of the Weyl group which sends R_+ to R_- .

This theorem can also be reformulated in more geometric terms: the spaces $V(\lambda)$ are naturally interpreted as spaces of global sections of appropriate line bundles on the “flag variety” $\mathcal{B} = G_{\mathbb{C}}/B = G/T$.

For classical groups, irreducible representations can also be constructed explicitly as the subspaces in tensor powers $(\mathbb{C}^n)^{\otimes k}$, transforming in a certain way under the action of the symmetric group S_k .

Characters and Multiplicities

Characters

Let (π, V) be a f.d. representation of G and let χ_π be its character as defined by [5]. Since χ_π is central, and every element in G is conjugate to an element of T , χ_π is completely determined by its restriction to T , which can be computed from the weight decomposition [6]:

$$\begin{aligned} \chi_\pi|_T &= \sum_{\lambda \in X(T)} \dim V_\lambda \cdot e_\lambda \\ &= \sum_{\lambda \in X(T)} \text{mult}_\pi \lambda \cdot e_\lambda \end{aligned} \tag{10}$$

where e_λ is the function on T defined by $e_\lambda(\exp(t)) = e^{\langle t, \lambda \rangle}$, $t \in \mathfrak{t}$. Note that $e_{\lambda+\mu} = e_\lambda e_\mu$ and that $e_0 = 1$.

Weyl Character Formula

Theorem 19 (Weyl character formula). *Let $\lambda \in X_+$. Then*

$$\chi_{L(\lambda)} = \frac{A_{\lambda+\rho}}{A_\rho}, \quad A_\mu = \sum_{w \in W} \varepsilon(w) e_{w(\mu)}$$

where, for $w \in W$, we denote $\varepsilon(w) = \det w$ considered as a linear map $\mathfrak{t}^* \rightarrow \mathfrak{t}^*$, and $\rho = (1/2) \sum_{R_+} \alpha$.

In particular, computing the value of the character at point $t=0$ by L’Hopital’s rule, it is possible to deduce the following formula for the dimension of irreducible representations:

$$\dim L(\lambda) = \prod_{\alpha \in R_+} \frac{\langle \alpha^\vee, \lambda + \rho \rangle}{\langle \alpha^\vee, \rho \rangle} \tag{11}$$

Example 9 Let $G = \text{SU}(2)$. Then Weyl character formula gives, for irreducible representation Π_k with highest weight $k \cdot \omega$,

$$\begin{aligned} \chi_{\Pi_k} &= \frac{x^{k+1} - x^{-(k+1)}}{x - x^{-1}} \\ &= x^k + x^{k-2} + \dots + x^{-k}, \quad x = e_\omega \end{aligned}$$

which implies $\dim \Pi_k = k + 1$.

Weyl character formula is equivalent to the following formula for weight multiplicities, due to Kostant:

$$\text{mult}_{L(\lambda)} \mu = \sum_{w \in W} \varepsilon(w) K(w(\lambda + \rho) - \rho - \mu)$$

where K is Kostant’s partition function: $K(\tau)$ is the number of ways of writing τ as a sum of positive roots (with repetitions).

For classical Lie groups such as $G = \text{U}(n)$, there are more explicit combinatorial formulas for weight multiplicities; for $\text{U}(n)$, the answer can be written in terms of the number of “Young tableaux” of a given shape. Details can be found in [Fulton and Harris \(1991\)](#).

Tensor Product Multiplicities

Let (π, V) be a f.d. representation of G . By complete reducibility, one can write $V = \sum n_\lambda L(\lambda)$. The coefficients n_λ are called multiplicities; finding them is an important problem in many applications. In particular, a special case of this is finding the multiplicities in tensor product of two unirreps:

$$L(\lambda) \otimes L(\mu) = \sum N_{\lambda\mu}^\nu L(\nu)$$

Characters provide a practical tool for computing multiplicities: since characters of unirreps are linearly independent, multiplicities can be found from the condition that $\chi_V = \sum n_\lambda \chi_{L(\lambda)}$. In particular,

$$\chi_{L(\lambda)} \chi_{L(\mu)} = \sum N_{\lambda\mu}^\nu \chi_{L(\nu)}$$

Example 10 For $G = \text{SU}(2)$, tensor product multiplicities are given by

$$\Pi_n \otimes \Pi_m = \oplus \Pi_l$$

where the sum is taken over all l such that $|m - n| \leq l \leq m + n$, $m + n + l$ is even.

For $G = \text{U}(n)$, there is an algorithm for finding the tensor product multiplicities, formulated in the language of Young tableaux (Littlewood–Richardson rule). There are also tables and computer programs for computing these multiplicities; some of them are listed in the bibliography.

See also: Classical Groups and Homogeneous Spaces; Combinatorics: Overview; Equivariant Cohomology and

the Cartan Model; Finite Group Symmetry Breaking; Lie Groups: General Theory; Ljusternik–Schnirelman Theory; Noncommutative Geometry and the Standard Model; Optimal Cloning of Quantum States; Ordinary Special Functions; Quasiperiodic Systems; Symmetry Classes in Random Matrix Theory.

Further Reading

Bump D (2004) *Lie Groups*. New York: Springer.
 Bröcker T and tom Dieck T (1995) *Representations of Compact Lie Groups*, Graduate Texts in Mathematics, vol. 98. New York: Springer.

Fulton W and Harris J (1991) *Representation Theory*. New York: Springer.
 Knapp A (2002) *Lie Groups beyond an Introduction*, 2nd edn. Boston: Birkhäuser.
 LiE: A Computer algebra package for Lie group computations, available from <http://young.sp2mi.univ-poitiers.fr>
 McKay WG, Patera J, and Rand DW (1990) *Tables of Representations of Simple Lie Algebras, vol. I. Exceptional Simple Lie Algebras*. Montreal: CRM.
 Serre J-P (2001) *Complex Semisimple Lie Algebras*. Berlin: Springer.
 Simon B (1996) *Representations of Finite and Compact Groups*. Providence, RI: American Mathematical Society.
 Zelobenko DP (1973) *Compact Lie Groups and Their Representations*. Providence, RI: American Mathematical Society.

Compactification of Superstring Theory

M R Douglas, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Superstring theories and M-theory, at present the best candidate quantum theories which unify gravity, Yang–Mills fields, and matter, are directly formulated in ten and eleven spacetime dimensions. To obtain a candidate theory of our four-dimensional universe, one must find a solution of one of these theories whose low-energy physics is well described by a four-dimensional effective field theory (EFT), containing the well-established standard model (SM) of particle physics coupled to Einstein’s general relativity (GR). The standard paradigm for finding such solutions is compactification, along the lines originally proposed by Kaluza and Klein in the context of higher-dimensional general relativity. One postulates that the underlying D -dimensional spacetime is a product of four-dimensional Minkowski spacetime, with a $(D - 4)$ -dimensional compact and small Riemannian manifold K . One then finds that low-energy physics effectively averages over K , leading to a four-dimensional EFT whose field content and Lagrangian are determined in terms of the topology and geometry of K .

Of the huge body of prior work on this subject, the part most relevant for string/M-theory is supergravity compactification, as in the limit of low energies, small curvatures and weak coupling, the various string theories and M-theory reduce to ten- and eleven-dimensional supergravity theories. Many of the qualitative features of string/M-theory compactification, and a good deal of what is known quantitatively, can be

understood simply in terms of compactification of these field theories, with the addition of a few crucial ingredients from string/M-theory. Thus, most of this article will restrict attention to this case, leaving many “stringy” topics to the articles on conformal field theory, topological string theory, and so on. We also largely restrict attention to compactifications based on Ricci-flat compact spaces. There is an equally important class in which K has positive curvature; these lead to anti-de Sitter (AdS) spacetimes and are discussed in the article on AdS/CFT (*see* AdS/CFT Correspondence).

After a general review, we begin with compactification of the heterotic string on a three complex dimensional Calabi–Yau manifold. This was the first construction which led convincingly to the SM, and remains one of the most important examples. We then survey the various families of compactifications to higher dimensions, with an eye on the relations between these compactifications which follow from superstring duality. We then discuss some of the phenomena which arise in the regimes of large curvature and strong coupling. In the final section, we bring these ideas together in a survey of the various known four-dimensional constructions.

General Framework

Let us assume we are given a D - ($=d + k$) dimensional field theory \mathcal{T} . A compactification is then a D -dimensional spacetime which is topologically the product of a d -dimensional spacetime with an k -dimensional manifold K , the compactification or “internal” manifold, carrying a Riemannian metric and with definite expectation values for all other fields in \mathcal{T} . These must solve the equations of motion, and preserve d -dimensional Poincaré invariance (or, perhaps another d -dimensional symmetry group).

The most general metric ansatz for a Poincaré invariant compactification is

$$G_{IJ} = \begin{pmatrix} f & \eta_{\mu\nu} & 0 \\ 0 & 0 & G_{ij} \end{pmatrix}$$

where the tangent space indices are $0 \leq I < d + k = D$, $0 \leq \mu < d$, and $1 \leq i \leq k$. Here $\eta_{\mu\nu}$ is the Minkowski metric, G_{ij} is a metric on K , and f is a real-valued function on K called the “warp factor.”

As the simplest example, consider pure D -dimensional GR. In this case, Einstein’s equations reduce to Ricci flatness of G_{IJ} . Given our metric ansatz, this requires f to be constant, and the metric G_{ij} on K to be Ricci flat. Thus, any K which admits such a metric, for example, the k -dimensional torus, will lead to a compactification.

Typically, if a manifold admits a Ricci-flat metric, it will not be unique; rather there will be a moduli space of such metrics. Physically, one then expects to find solutions in which the choice of Ricci-flat metric is slowly varying in d -dimensional spacetime. General arguments imply that such variations must be described by variations of d -dimensional fields, governed by an EFT. Given an explicit parametrization of the family of metrics, say $G_{ij}(\phi^\alpha)$ for some parameters ϕ^α , in principle the EFT could be computed explicitly by promoting the parameters to d -dimensional fields, substituting this parametrization into the D -dimensional action, and expanding in powers of the d -dimensional derivatives. In pure GR, we would find the four-dimensional effective Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{EFT}} = & \int d^k y \sqrt{\det G(\phi)} R^{(4)} \\ & + \sqrt{\det G(\phi)} G^{ik}(\phi) G^{jl}(\phi) \frac{\partial G_{ij}}{\partial \phi^\alpha} \frac{\partial G_{kl}}{\partial \phi^\beta} \partial_\mu \phi^\alpha \partial_\mu \phi^\beta \\ & + \dots \end{aligned} \quad [1]$$

While this is easily evaluated for K a symmetric space or torus, in general a direct computation of \mathcal{L}_{EFT} is impossible. This becomes especially clear when one learns that the Ricci-flat metrics G_{ij} are not explicitly known for the examples of interest. Nevertheless, clever indirect methods have been found that give a great deal of information about \mathcal{L}_{EFT} ; this is much of the art of superstring compactification. However, in this section, let us ignore this point and continue as if we could do such computations explicitly.

Given a solution, one proceeds to consider its small perturbations, which satisfy the linearized equations of motion. If these include exponentially growing modes (often called “tachyons”), the solution is unstable. (Note that this criterion is modified

for AdS compactifications). The remaining perturbations can be divided into massless fields, corresponding to zero modes of the linearized equations of motion on K , and massive fields, the others. General results on eigenvalues of Laplacians imply that the masses of massive fields depend on the diameter of K as $m \sim 1/\text{diam}(K)$, so at energies far smaller than m , they cannot be excited (this is not universal; given strong negative curvature on K , or a rapidly varying warp factor, one can have perturbations of small nonzero mass). Thus, the massive fields can be “integrated out,” to leave an EFT with a finite number of fields. In the classical approximation, this simply means solving their equations of motion in terms of the massless fields, and using these solutions to eliminate them from the action. At leading order in an expansion around a solution, these fields are zero and this step is trivial; nevertheless, it is useful in making a systematic definition of the interaction terms in the EFT.

As we saw in pure GR, the configuration space parametrized by the massless fields in the EFT, is the moduli space of compactifications obtained by deforming the original solution. Thus, from a mathematical point of view, low-energy EFT can be thought of as a sort of enhancement of the concept of moduli space, and a dictionary set up between mathematical and physical languages. To give its next entry, there is a natural physical metric on moduli space, defined by restriction from the metric on the configuration space of the theory \mathcal{T} ; this becomes the sigma-model metric for the scalars in the EFT. Because the theories \mathcal{T} arising from string theory are geometrically natural, this metric is also natural from a mathematical point of view, and one often finds that much is already known about it. For example, the somewhat fearsome two derivative terms in eqn [1], are (perhaps) less so when one realizes that this is an explicit expression for the Weil–Petersson metric on the moduli space of Ricci-flat metrics. In any case, knowing this dictionary is essential for taking advantage of the literature.

Another important entry in this dictionary is that the automorphism group of a solution translates into the gauge group in the EFT. This can be either continuous, leading to the gauge symmetry of Maxwell and Yang–Mills theories, or discrete, leading to discrete gauge symmetry. For example, if the metric on K has continuous isometry group G , the resulting EFT will have gauge symmetry G , as in the original example of Kaluza and Klein with $K \cong S^1$ and $G \cong U(1)$. Mathematically, these phenomena of “enhanced symmetry” are often treated using the languages of equivariant theories (cohomology, K-theory, etc.), stacks, and so on.

To give another example, obstructed deformations (solutions of the linearized equations which do not correspond to elements of the tangent space of the true moduli space) correspond to scalar fields which, while massless, appear in the effective potential in a way which prevents giving them expectation values. Since the quadratic terms V'' are masses, this dependence must be at cubic or higher order.

While the preceding concepts are general and apply to compactification of all local field theories, string and M-theory add some particular ingredients to this general recipe. In the limits of small curvatures and weak coupling, string and M-theory are well described by the ten- and 11-dimensional supergravity theories, and thus the string/M-theory discussion usually starts with Kaluza–Klein compactification of these theories, which we denote I, IIa, IIb, HE, HO and M. Let us now discuss a particular example.

Calabi–Yau Compactification of the Heterotic String

Contact with the SM requires finding compactifications to $d = 4$ either without supersymmetry, or with at most $N = 1$ supersymmetry, because the SM includes chiral fermions, which are incompatible with $N > 1$. Let us start with the $E_8 \times E_8$ heterotic string or “HE” theory. This choice is made rather than HO because only in this case can we find the SM fermion representations as subrepresentations of the adjoint of the gauge group.

Besides the metric, the other bosonic fields of the HE supergravity theory are a scalar Φ called the dilaton, Yang–Mills gauge potentials for the group $G \equiv E_8 \times E_8$, and a 2-form gauge potential B (often called the “Neveu–Schwarz” or “NS” 2-form) whose defining characteristic is that it minimally couples to the heterotic string world-sheet. We will need their gauge field strengths below: for Yang–Mills, this is a 2-form F_{IJ}^a with a indexing the adjoint of Lie G , and for the NS 2-form this is a 3-form H_{IJK} . Denoting the two Majorana–Weyl spinor representations of $SO(1, 9)$ as S and C , then the fermions are the gravitino $\psi_I \in S \otimes V$, a spin 1/2 “dilatino” $\lambda \in C$, and the adjoint gauginos $\chi^a \in S$. We use Γ_I to denote Dirac matrices contracted with a “zehnbein,” satisfying $\{\Gamma_I, \Gamma_J\} = 2G_{IJ}$, and $\Gamma_{IJ} = (1/2)[\Gamma_I, \Gamma_J]$, etc.

A local supersymmetry transformation with parameter ϵ is then

$$\delta\psi_I = D_I\epsilon + \frac{1}{8}H_{IJK}\Gamma^{JK}\epsilon \quad [2]$$

$$\delta\lambda = \partial_I\Phi\Gamma^I\epsilon - \frac{1}{12}H_{IJK}\Gamma^{IJK}\epsilon \quad [3]$$

$$\delta\chi^a = F_{IJ}^a\Gamma^{IJ}\epsilon \quad [4]$$

We now assume $N = 1$ supersymmetry. An unbroken supersymmetry is a spinor ϵ for which the left-hand side is zero, so we seek compactifications with a unique solution of these equations.

We first discuss the case $H = 0$. Setting $\delta\psi_\mu$ in eqn [2] to zero, we find that the warp factor f must be constant. The vanishing of $\delta\psi_i$ requires ϵ to be a covariantly constant spinor. For a six-dimensional M to have a unique such spinor, it must have $SU(3)$ holonomy; in other words, M must be a Calabi–Yau manifold. In the following, we use basic facts about their geometry.

The vanishing of $\delta\lambda$ then requires constant dilaton Φ , while the vanishing of $\delta\chi^a$ requires the gauge field strength F to solve the hermitian Yang–Mills equations,

$$F^{2,0} = F^{0,2} = F^{1,1} = 0$$

By the theorem of Donaldson and Uhlenbeck–Yau, such solutions are in one-to-one correspondence with μ -stable holomorphic vector bundles with structure group H contained in the complexification of G . Choose such a bundle E ; by the general discussion above, the commutant of H in G will be the automorphism group of the connection on E and thus the low-energy gauge group of the resulting EFT. For example, since E_8 has a maximal $E_6 \times SU(3)$ subgroup, if E has structure group $H = SL(3)$, there is an embedding such that the unbroken gauge symmetry is $E_6 \times E_8$, realizing one of the standard grand unified groups E_6 as a factor.

The choice of E is constrained by anomaly cancellation. This discussion (Green *et al.* 1987) modifies the Bianchi identity for H to

$$dH = \text{tr } R \wedge R - \frac{1}{30} \sum_a F^a \wedge F^a \quad [5]$$

where R is the matrix of curvature 2-forms. The normalization of the $F \wedge F$ term is such that if we take $E \cong TK$ the holomorphic tangent bundle of K , with isomorphic connection, then using the embedding we just discussed, we obtain a solution of eqn [5] with $H = 0$.

Thus, we have a complete solution of the equations of motion. General arguments imply that supersymmetric Minkowski solutions are stable, so the small fluctuations consist of massless and massive fields. Let us now discuss a few of the massless fields. Since the EFT has $N = 1$ supersymmetry, the massless scalars live in chiral multiplets, which are local coordinates on a complex Kähler manifold.

First, the moduli of Ricci-flat metrics on K will lead to massless scalar fields: the complex structure

moduli, which are naturally complex, and Kähler moduli, which are not. However, in string compactification the latter are complexified to the periods of the 2-form $B + iJ$ integrated over a basis of $H_2(K, \mathbb{Z})$, where J is the Kähler form and B is the NS 2-form. In addition, there is a complex field pairing the dilaton (actually, $\exp(-\Phi)$) and the “model-independent axion,” the scalar dual in $d=4$ to the 2-form $B_{\mu\nu}$. Finally, each complex modulus of the holomorphic bundle E will lead to a chiral multiplet. Thus, the total number of massless uncharged chiral multiplets is $1 + h^{1,1}(K) + h^{2,1}(K) + \dim H^1(K, \text{End}(E))$.

Massless charged matter will arise from zero modes of the gauge field and its supersymmetric partner spinor χ^a . It is slightly easier to discuss the spinor, and then appeal to supersymmetry to get the bosons. Decomposing the spinors of $\text{SO}(6)$ under $\text{SU}(3)$, one obtains $(0, p)$ forms, and the Dirac equation becomes the condition that these forms are harmonic. By the Hodge theorem, these are in one-to-one correspondence with classes in Dolbeault cohomology $H^{0,p}(K, V)$, for some bundle V . The bundle V is obtained by decomposing the spinor into representations of the holonomy group of E . For $H = \text{SU}(3)$, the decomposition of the adjoint under the embedding of $\text{SU}(3) \times E_6$ in E_8 ,

$$248 = (8, 1) + (1, 78) + (3, 27) + (\bar{3}, \bar{27}) \quad [6]$$

implies that charged matter will form “generations” in the 27, of number $\dim H^{0,1}(K, E)$, and “antigenerations” in the $\bar{27}$, of number $\dim H^{0,1}(K, \bar{E}) = \dim H^{0,2}(K, E)$. The difference in these numbers is determined by the Atiyah–Singer index theorem to be

$$N_{\text{gen}} \equiv N_{27} - N_{\bar{27}} = \frac{1}{2}c_3(E)$$

In the special case of $E \cong TK$, these numbers are separately determined to be $N_{27} = b^{1,1}$ and $N_{\bar{27}} = b^{2,1}$, so their difference is $\chi(K)/2$, half the Euler number of K . In the real world, this number is $N_{\text{gen}} = 3$, and matching this under our assumptions so far is very constraining.

Substituting these zero modes into the ten-dimensional Yang–Mills action and integrating, one can derive the $d=4$ EFT. For example, the cubic terms in the superpotential, usually called Yukawa couplings after the corresponding fermion–boson interactions in the component Lagrangian, are obtained from the cubic product of zero modes

$$\int_K \Omega \wedge \text{tr}(\phi_1 \wedge \phi_2 \wedge \phi_3)$$

where Ω is the holomorphic $\phi_i \in H^{0,1}(K, \text{Rep } E)$ are the zero modes, and tr arises from decomposing the E_8 cubic group invariant.

Note the very important fact that this expression only depends on the cohomology classes of the ϕ_i (and Ω). This means the Yukawa couplings can be computed without finding the explicit harmonic representatives, which is not possible (we do not even know the explicit metric). More generally, one expects to be able to explicitly compute the superpotential and all other holomorphic quantities in the effective Lagrangian solely from “topological” information (the Dolbeault cohomology ring, and its generalizations within topological string theory). On the other hand, computing the Kähler metric in an $N=1$ EFT is usually out of reach as it would require having explicit normalized zero modes. Most results for this metric come from considering closely related compactifications with extended supersymmetry, and arguing that the breaking to $N=1$ supersymmetry makes small corrections to this.

There are several generalizations of this construction. First, the necessary condition to solve eqn [5] is that the left-hand side be exact, which requires

$$c_2(E) = c_2(TK) \quad [7]$$

This allows for a wide variety of E ’s to be used, so that $N_{\text{gen}} = 3$ can be attained with many more K ’s. This class of models is often called “ $(0, 2)$ compactification” to denote the world-sheet supersymmetry of the heterotic string in these backgrounds. One can also use bundles with larger structure group; for example, $H = \text{SL}(4)$ leads to unbroken $\text{SO}(10) \times E_8$, and $H = \text{SL}(5)$ leads to unbroken $\text{SU}(5) \times E_8$.

The subsequent breaking of the grand unified group to the SM gauge group is typically done by choosing K with nontrivial π_1 , so that it admits a flat line bundle W with nontrivial holonomy (usually called a “Wilson line”). One then uses the bundle $E \otimes W$ in the above discussion, to obtain the commutant of $H \otimes W$ as gauge group. For example, if $\pi_1(K) \cong \mathbb{Z}_5$, one can use W whose holonomy is an element of order 5 in $\text{SU}(5)$, to obtain as commutant the SM gauge group $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$.

Another generalization is to take the 3-form $H \neq 0$. This discussion begins by noting that, for supersymmetry, we still require the existence of a unique spinor ϵ ; however, it will no longer be covariantly constant in the Levi-Civita connection. One way to structure the problem is to note that the right-hand side of eqn [2] takes the form of a connection with torsion; the resulting equations have been discussed mathematically in (Li and Yau 2004).

Another recent approach to these compactifications (Gauntlett 2004) starts out by arguing that ϵ cannot vanish on K , so it defines a weak $\text{SU}(3)$ structure, a local reduction of the structure group of

$T K$ to $SU(3)$ which need not be integrable. This structure must be present in all $N = 1, d = 4$ supersymmetric compactifications and there are hopes that it will lead to a useful classification of the possible local structures and corresponding partial differential equations (PDEs) on K .

Higher-Dimensional and Extended Supersymmetric Compactifications

While there are similar quasirealistic constructions which start from the other string theories and M-theory, before we discuss these, let us give an overview of compactifications with $N \geq 2$ supersymmetry in four dimensions, and in higher dimensions. These are simpler analog models which can be understood in more depth; their study led to one of the most important discoveries in string/M-theory, the theory of superstring duality.

As before, we require a covariantly constant spinor. For Ricci-flat K with other background fields zero, this requires the holonomy of K to be one of trivial, $SU(n)$, $Sp(n)$, or the exceptional holonomies G_2 or $Spin(7)$. In Table 1 we tabulate the possibilities with spacetime dimension d greater or equal to 3, listing the supergravity theory, the holonomy type of K , and the type of the resulting EFT: dimension d , total number of real supersymmetry parameters N_s , and the number of spinor supercharges N (in $d = 6$, since left- and right-chirality Majorana spinors are inequivalent, there are two numbers).

The structure of the resulting supergravity EFTs is heavily constrained by N_s . We now discuss the various possibilities.

Table 1 String/M-theories, holonomy groups and the resulting supersymmetry

Theory	Holonomy	d	N_s	N
M, II	Torus	Any	32	Max
M	SU(2)	7	16	1
	SU(3)	5	8	1
	G_2	4	4	1
	Sp(4)	3	6	3
	SU(4)	3	4	2
	Spin(7)	3	2	1
IIa	SU(2)	6	16	(1, 1)
	SU(3)	4	8	2
	G_2	3	4	2
IIb	SU(2)	6	16	(0, 2)
	SU(3)	4	8	2
	G_2	3	4	2
HE, HO, I	Torus	Any	16	Max/2
	SU(2)	6	8	1
	SU(3)	4	4	1
	G_2	3	2	1

$N_s = 32$

Given the supersymmetry algebra, if such a supergravity exists, it is unique. Thus, toroidal compactifications of $d = 11$ supergravity, IIa and IIb supergravity lead to the same series of maximally supersymmetric theories. Their structure is governed by the exceptional Lie algebra E_{11-d} ; the gauge charges transform in a fundamental representation of this algebra, while the scalar fields parametrize a coset space G/H , where G is the maximally split real form of the Lie group E_{11-d} , and H is a maximal compact subgroup of G . Nonperturbative duality symmetries lead to a further identification by a maximal discrete subgroup of G .

$N_s = 16$

This supergravity can be coupled to maximally supersymmetric super Yang–Mills theory, which given a choice of gauge group G is unique. Thus, these theories (with zero cosmological constant and thus allowing super-Poincaré symmetry) are uniquely determined by the choice of G .

In $d = 10$, the choices $E_8 \times E_8$ and $Spin(32)/\mathbb{Z}_2$ which arise in string theory, are almost uniquely determined by the Green–Schwarz anomaly cancellation analysis. Compactification of these HE, HO and type I theories on T^n produces a unique theory with moduli space

$$\mathbb{R}^+ \times SO(n, n + 16; \mathbb{Z}) \backslash SO(n, n + 16; \mathbb{R}) / SO(n, \mathbb{R}) \times SO(n + 16, \mathbb{R}) \quad [8]$$

In Kaluza–Klein (KK) reduction, this arises from the choice of metric g_{ij} , the antisymmetric tensor B_{ij} and the choice of a flat $E_8 \times E_8$ or $Spin(32)/\mathbb{Z}_2$ connection on T^n , while a more unified description follows from the heterotic string world-sheet analysis. Here the group $SO(n, n + 16)$ is defined to preserve an even self-dual quadratic form η of signature $(n, n + 16)$; for example, $\eta = (-E_8) \oplus (-E_8) \oplus I \oplus I \oplus I$, where I is the form of signature $(1, 1)$ and E_8 is the Cartan matrix. In fact, all such forms are equivalent under orthogonal integer similarity transformation; so, the resulting EFT is unique. It has a rank $16 + 2n$ gauge group, which at generic points in moduli space is $U(1)^{16+2n}$, but is enhanced to a nonabelian group G at special points. To describe G , we first note that a point p in moduli space determines an n -dimensional subspace V_p of \mathbb{R}^{16+2n} , and an orthogonal subspace V_p^\perp (of varying dimension). Lattice points of length squared -2 contained in V_p^\perp then correspond to roots of the Lie algebra of G_p .

The other compactifications with $N_s=16$ is M-theory on K3 and its further toroidal reductions, and IIB on K3. M-theory compactification to $d=7$ is dual to heterotic on T^3 , with the same moduli space and enhanced gauge symmetry. As we discuss at the end of the section “Stringy and quantum corrections,” the extra massless gauge bosons of enhanced gauge symmetry are M2 branes wrapped on 2-cycles with topology S^2 . For such a cycle to have zero volume, the integral of the Kähler form and holomorphic 2-form over the cycle must vanish; expressing this in a basis for $H^2(K3, \mathbb{R})$ leads to exactly the same condition we discussed for enhanced gauge symmetry above. The final result is that all such K3 degenerations lead to one- of the two-dimensional canonical singularities, of types A, D or E, and the corresponding EFT phenomenon is the enhanced gauge symmetry of corresponding Dynkin type A, D, or E.

IIB on K3 is similar, but reducing the self-dual Ramond–Ramond (RR) 4-form potential on the 2-cycles leads to self-dual tensor multiplets instead of Maxwell theory. The moduli space is eqn [8] but with $n=5$, not $n=4$, incorporating periods of RR potentials and the $SL(2, \mathbb{Z})$ duality symmetry of IIB theory.

One may ask if the $N_s=16$ I/HE/HO theories in $d=8$ and $d=9$ have similar duals. For $d=8$, these are obtained by a pretty construction known as “F-theory.” Geometrically, the simplest definition of F-theory is to consider the special case of M-theory on an elliptically fibered Calabi–Yau, in the limit that the Kähler modulus of the fiber becomes small. One check of this claim for $d=8$ is that the moduli space of elliptically fibered K3s agrees with eqn [8] with $n=2$.

Another definition of F-theory is the particular case of IIB compactification using Dirichlet 7-branes, and orientifold 7-planes. This construction is T -dual to the type I theory on T^2 , which provides its simplest string theory definition. As discussed in Polchinski (1999), one can think of the open strings giving rise to type I gauge symmetry as living on 32 Dirichlet 9-branes (or D9-branes) and an orientifold nineplane. T -duality converts Dirichlet and orientifold p -branes to $(p-1)$ -branes; thus this relation follows by applying two T -dualities.

These compactifications can also be parametrized by elliptically fibered Calabi–Yaus, where K is the base, and the branes correspond to singularities of the fibration. The relation between these two definitions follows fairly simply from the duality between M-theory on T^2 , and IIB string on S^1 . There is a partially understood generalization of this to $d=9$.

Finally, these constructions admit further discrete choices, which break some of the gauge symmetry. The simplest to explain is in the toroidal compactification of I/HE/HO. The moduli space of theories we discussed uses flat connections on the torus which are continuously connected to the trivial connection, but in general the moduli space of flat connections has other components. The simplest example is the moduli space of flat $E_8 \times E_8$ connections on S^1 , which has a second component in which the holonomy exchanges the two E_8 ’s. On T^3 , there are connections for which the holonomies cannot be simultaneously diagonalized. This structure and the M-theory dual of these choices is discussed in (de Boer *et al.* 2001).

$N_s=8, d < 6$

Again, the gravity multiplet is uniquely determined, so the most basic classification is by the gauge group G . The full low-energy EFT is determined by the matter content and action, and there are two types of matter multiplets. First, vector multiplets contain the Yang–Mills fields, fermions and $6-d$ scalars; their action is determined by a prepotential which is a G -invariant function of the fields. Since the vector multiplets contain massless adjoint scalars, a generic vacuum in which these take nonzero distinct vacuum expectation values (VEVs) will have $U(1)^f$ gauge symmetry, the commutant of G with a generic matrix (for $d < 5$, while there are several real scalars, the potential forces these to commute in a supersymmetric vacuum). Vacua with this type of gauge symmetry breaking, which does not reduce the rank of the gauge group, are usually referred to as on a “Coulomb branch” of the moduli space. To summarize, this sector can be specified by n_V , the number of vector multiplets, and the prepotential \mathcal{F} , a function of the n_V VEVs which is cubic in $d=5$, and holomorphic in $d=4$.

Hypermultiplets contain scalars which parametrize a quaternionic Kähler manifold, and partner fermions. Thus, this sector is specified by a $4n_H$ real dimensional quaternionic Kähler manifold. The G action comes with triholomorphic moment maps; if nontrivial, VEVs in this sector can break gauge symmetry and reduce it in rank. Such vacua are usually referred to as on a “Higgs branch.”

The basic example of these compactifications is M-theory on a Calabi–Yau 3-fold (CY_3). Reduction of the 3-form leads to $h^{1,1}(K)$ vector multiplets, whose scalar components are the CY Kähler moduli. The CY complex structure moduli pair with periods of the 3-form to produce $h^{2,1}(K)$ hypermultiplets. Enhanced gauge symmetry then appears when the

CY_3 contains ADE singularities fibered over a curve, from the same mechanism involving wrapped M2 branes we discussed under $N_s = 16$. If degenerating curves lead to other singularities (e.g., the ODP or “conifold”), it is possible to obtain extremal transitions which translate physically into Coulomb–Higgs transitions. Finally, singularities in which surfaces degenerate lead to nontrivial fixed-point theories.

Reduction on S^1 leads to IIA on CY_3 , with the spectrum above plus a “universal hypermultiplet” which includes the dilaton. Perhaps the most interesting new feature is the presence of worldsheet instantons, which correct the metric on vector multiplet moduli space. This metric satisfies the restrictions of special geometry and thus can be derived from a prepotential.

The same theory can be obtained by compactification of IIB theory on the mirror CY_3 . Now vector multiplets are related to the complex structure moduli space, while hypermultiplets are related to Kähler moduli space. In this case, the prepotential derived from variation of complex structure receives no instanton corrections, as we discuss in the next section.

Finally, one can compactify the heterotic string on $K3 \times T^{6-d}$, but this theory follows from toroidal reduction of the $d = 6$ case we discuss next.

$N_s = 8, d = 6$

These supergravities are similar to $d < 6$, but there is a new type of matter multiplet, the self-dual tensor (in $d < 6$ this is dual to a vector multiplet). Since fermions in $d = 6$ are chiral, there is an anomaly cancellation condition relating the numbers of the three types of multiplets (Aspinwall 1996, section 6.6),

$$n_H - n_V + 29n_T = 273 \quad [9]$$

One class of examples is the heterotic string compactified on K3. In the original perturbative constructions, to satisfy eqn [7], we need to choose a vector bundle with $c_2(V) = \chi(K3) = 24$. The resulting degrees of freedom are a single self-dual tensor multiplet and a rank-16 gauge group. More generally, one can introduce N_{5B} heterotic 5-branes, which generalize eqn [7] to $c_2(E) + N_{5B} = c_2(TK)$. Since this brane carries a self-dual tensor multiplet, this series of models is parametrized by n_T . They are connected by transitions in which an E_8 instanton shrinks to zero size and becomes a 5-brane; the resulting decrease in the dimension of the moduli space of E_8 bundles on K3 agrees with eqn [9].

Another class of examples is F-theory on an elliptically fibered CY_3 . These are related to

M-theory on an elliptically fibered CY_3 in the same general way we discussed under $N_s = 16$. The relation between F-theory and the heterotic string on K3 can be seen by lifting M-theory–heterotic duality; this suggests that the two constructions are dual only if the CY_3 is a K3 fibration as well. Since not all elliptically fibered CY_3 s are K3 fibered, the F-theory construction is more general.

We return to $d = 4$ and $N_s = 4$ in the final section. The cases of $N_s < 4$ which exist in $d \leq 3$ are far less studied.

Stringy and Quantum Corrections

The D -dimensional low-energy effective supergravity actions on which we based our discussion so far are only approximations to the general story of string/M-theory compactification. However, if Planck’s constant is small, K is sufficiently large, and its curvature is small, then they are controlled approximations.

In M-theory, as in any theory of quantum gravity, corrections are controlled by the Planck scale parameter M_P^{D-2} , which sits in front of the Einstein term of the D -dimensional effective Lagrangian, and plays the role of \hbar . In general, this is different from the four-dimensional Planck scale, which satisfies $M_{P4}^2 = \text{Vol}(K)M_P^{D-2}$. After taking the low-energy limit $E \ll M_P$, the remaining corrections are controlled by the dimensionless parameters l_P/R , where R can any characteristic length scale of the solution: a curvature radius, the length of a nontrivial cycle, and so on.

In string theory, one usually thinks of the corrections as a double series expansion in g_s , the dimensionless (closed) string coupling constant, and α' , the inverse string tension parameter, of dimensions (length)². The ten-dimensional Planck scale is related to these parameters as $M_P^8 = 1/g_s^2(\alpha')^4$, up to a constant factor that depends on conventions.

Besides perturbative corrections, which have power-like dependence on these parameters, there can be world sheet and “brane” instanton corrections. For example, a string world sheet can wrap around a topologically nontrivial spacelike 2-cycle Σ in K , leading to an instanton correction to the effective action which is suppressed as $\exp(-\text{Vol}(\Sigma)/2\pi\alpha')$. More generally, any p -brane wrapping a p -cycle can produce a similar effect. As for which terms in the effective Lagrangian receive corrections, this depends largely on the number and symmetries of the fermion zero modes on the instanton world volumes.

Let us start by discussing some cases in which one can argue that these corrections are not present.

First, extended supersymmetry can serve to eliminate many corrections. This is analogous to the familiar result that the superpotential in $d=4, N=1$ supersymmetric field theory does not receive (or “is protected from”) perturbative corrections, and in many cases follows from similar formal arguments. In particular, supersymmetry forbids corrections to the potential and two derivative terms in the $N_s=32$ and $N_s=16$ Lagrangians.

In $N_s=8$, the superpotential is protected, but the two derivative terms can receive corrections. However, there is a simple argument which precludes many such corrections – since vector multiplet and hypermultiplet moduli spaces are decoupled, a correction whose control parameter sits in (say) a vector multiplet, cannot affect hypermultiplet moduli space. This fact allows for many exact computations in these theories.

As an example, in IIB on CY_3 , the metric on vector multiplet moduli space is precisely eqn [1] as obtained from supergravity (in other words, the Weil–Petersson metric on complex structure moduli space). First, while in principle it could have been corrected by world-sheet instantons, since these depend on Kähler moduli which sit in hypermultiplets, it is not. The only other instantons with the requisite zero modes to modify this metric are wrapped Dirichlet branes. Since in IIB theory these wrap even-dimensional cycles, they also depend on Kähler moduli and thus leave vector moduli space unaffected.

As previously discussed, for K3-fibered CY_3 , this theory is dual to the heterotic string on $K3 \times T^2$. There, the vector multiplets arise from Wilson lines on T^2 , and reduce to an adjoint multiplet of $N=2$ supersymmetric Yang–Mills theory. Of course, in the quantum theory, the metric on this moduli space receives instanton corrections. Thus, the duality allows deriving the exact moduli space metric, and many other results of the Seiberg–Witten theory of $N=2$ gauge theory, as aspects of the geometry of Calabi–Yau moduli space.

In $N_s=4$, only the superpotential is protected, and that only in perturbation theory; it can receive nonperturbative corrections. Indeed, it appears that this is fairly generic, suggesting that the effective potentials in these theories are often sufficiently complicated to exhibit the structure required for supersymmetry breaking and the other symmetry breakings of the SM. Understanding this is an active subject of research.

We now turn from corrections to novel physical phenomena which arise in these regimes. While this is too large a subject to survey here, one of the basic principles which governs this subject is the idea that

string/M-theory compactification on a singular manifold K is typically consistent, but has new light degrees of freedom in the EFT, not predicted by KK arguments. We implicitly touched on one example of this in the discussion of M-theory compactification on K3 above, as the space of Ricci-flat K3 metrics has degeneration limits in which curvatures grow without bound, while the volumes of 2-cycles vanish. On the other hand, the structure of $N_s=16$ supersymmetry essentially forces the $d=7$ EFT in these limits to be non-singular. Its only noteworthy feature is that a nonabelian gauge symmetry is restored, and thus certain charged vector bosons and their superpartners become massless.

To see what is happening microscopically, we must consider an M-theory membrane (or 2-brane), wrapped on a degenerating 2-cycle. This appears as a particle in $d=7$, charged under the vector potential obtained by reduction of the $D=11$ 3-form potential. The mass of this particle is the volume of the 2-cycle multiplied by the membrane tension, so as this volume shrinks to zero, the particle becomes massless. Thus, the physics is also well defined in 11 dimensions, though not literally described by 11-dimensional supergravity.

This phenomenon has numerous generalizations. Their common point is that, since the essential physics involves new light degrees of freedom, they can be understood in terms of a lower-dimensional quantum theory associated with the region around the singularity. Depending on the geometry of the singularity, this is sometimes a weakly coupled field theory, and sometimes a nontrivial conformal field theory. Occasionally, as in IIB on K3, the lightest wrapped brane is a string, leading to a “little string theory” (Aharony 2000).

$N=1$ Supersymmetry in Four Dimensions

Having described the general framework, we conclude by discussing the various constructions which lead to $N=1$ supersymmetry. Besides the heterotic string on a CY_3 , these compactifications include type IIA and IIB on orientifolds of CY_3 , the related F-theory on elliptically fibered Calabi–Yau 4-folds (CY_4), and M-theory on G_2 manifolds. Let us briefly spell out their ingredients, the known nonperturbative corrections to the superpotential, and the duality relations between these constructions.

To start, we recap the heterotic string construction. We must specify a CY_3K , and a bundle E over K which admits a Hermitian Yang–Mills connection. The gauge group G is the commutant of the structure group of E in $E_8 \times E_8$ or $Spin(32)/\mathbb{Z}_2$,

while the chiral matter consists of metric moduli of K , and fields corresponding to a basis for the Dolbeault cohomology group $H^{0,1}(K, \text{Rep } E)$ where $\text{Rep } E$ is the bundle E embedded into an E_8 bundle and decomposed into G -reps.

There is a general (though somewhat formal) expression for the superpotential,

$$W = \int \Omega \wedge + \text{tr}(\bar{A}\partial\bar{A} + \frac{2}{3}\bar{A}^3) + \int \Omega \wedge H^{(3)} + W_{\text{NP}} \quad [10]$$

The first term is the holomorphic Chern–Simons action, whose variation enforces the $F^{0,2} = 0$ condition. The second is the “flux superpotential,” while the third term is the nonperturbative corrections. The best understood of these arise from supersymmetric gauge theory sectors. In some, but not all, cases, these can be understood as arising from gauge theoretic instantons, which can be shown to be dual to heterotic 5-branes wrapped on K . Heterotic world-sheet instantons can also contribute.

The HO theory is S -dual to the type I string, with the same gauge group, realized by open strings on Dirichlet 9-branes. This construction involves essentially the same data. The two classes of heterotic instantons are dual to D1- and D5-brane instantons, whose world-sheet theories are somewhat simpler.

If the $\text{CY}_3 K$ has a fibration by tori, by applying T -duality to the fibers along the lines discussed for tori under $N_s = 16$ above, one obtains various type II orientifold compactifications. On an elliptic fibration, double T -duality produces a IIb compactification with D7s and O7s. Using the relation between IIb theory on T^2 and F-theory on K3 fiberwise, one can also think of this as an F-theory compactification on a K3-fibered CY_4 . More generally, one can compactify F theory on any elliptically fibered 4-fold to obtain $N=1$. These theories have D3-instantons, the T -duals of both the type I D1- and D5-brane instantons.

The theory of mirror symmetry predicts that all CY_3 s have T^3 fibration structures. Applying the corresponding triple T -duality, one obtains a IIA compactification on the mirror $\text{CY}_3 \tilde{K}$, with D6-branes and O6-planes. Supersymmetry requires these to wrap special Lagrangian cycles in \tilde{K} . As in all Dirichlet brane constructions, enhanced gauge symmetry arises from coincident branes wrapping the same cycle, and only the classical groups are visible in perturbation theory. Exceptional gauge symmetry arises as a strong coupling phenomenon of the sort described in the previous section. The superpotential can also be thought of as mirror to eqn [10], but now the first term is the sum of a real

Chern–Simons action on the special Lagrangian cycles, with disk world-sheet instanton corrections, as studied in open string mirror symmetry. The gauge theory instantons are now D2-branes.

Using the duality relation between the IIA string and 11-dimensional M-theory, this construction can be lifted to a compactification of M-theory on a seven-dimensional manifold L , which is an S^1 fibration over K . The D6 and O6 planes arise from singularities in the S^1 fibration. Generically, L can be smooth, and the only candidate in Table 1 for such an $N=1$ compactification is a manifold with G_2 holonomy; therefore, L must have such holonomy. Finally, both the IIA world-sheet instantons and the D2-brane instantons lift to membrane instantons in M-theory.

This construction implicitly demonstrates the existence of a large number of G_2 holonomy manifolds. Another way to arrive at these is to go back to the heterotic string on K , and apply the duality (again under $N_s = 16$) between heterotic on T^3 and M-theory on K3 to the T^3 fibration structure on K , to arrive at M-theory on a K3-fibered manifold of G_2 holonomy. Wrapping membranes on 2-cycles in these fibers, we can see enhanced gauge symmetry in this picture fairly directly. It is an illuminating exercise to work through its dual realizations in all of these constructions.

Our final construction uses the interpretation of the strong coupling limit of the HE theory as M-theory on a one-dimensional interval I , in which the two E_8 factors live on the two boundaries. Thus, our original starting point can also be interpreted as the heterotic string on $K \times I$. This construction is believed to be important physically as it allows generalizing a heterotic string tree-level relation between the gauge and gravitational couplings which is phenomenologically disfavored. One can relate it to a IIA orientifold as well, now with D8- and O8-branes.

These multiple relations are often referred to as the “web” of dualities. They lead to numerous relations between compactification manifolds, moduli spaces, superpotentials, and other properties of the EFTs, whose full power has only begun to be appreciated.

Suggestions for further reading

Original references for all but the most recent of these topics can be found in the following textbooks and proceedings. We have also referenced a few research articles which are good starting points for the more recent literature. There are far more reviews than we could reference here, and a partial listing of these appears at <http://www.slac.stanford.edu/spires/reviews/>

See also: Brane Construction of Gauge Theories; Random Algebraic Geometry, Attractors and Flux Vacua;

String Theory: Phenomenology; Superstring Theories; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras; Viscous Incompressible Fluids: Mathematical Theory.

Further Reading

- Aharony O (2000) A brief review of “little string theories.” *Classical and Quantum Gravity* 17: 929–938.
- Aspinwall PS (1996) K3 surfaces and string duality, 1996 preprint, arXiv:hep-th/9611137.
- Bachas C *et al.* (eds.) (2002) *Les Houches 2001: Unity from Duality: Gravity, Gauge Theory and Strings*. Berlin: Springer.
- de Boer J *et al.* (2002) Triples, fluxes, and strings. *Advances in Theoretical and Mathematical Physics* 4: 995.

- Connes A and Gawędzki K (eds.) (1998) *Les Houches 1995: Quantum Symmetries*. Amsterdam: North-Holland.
- Deligne P *et al.* (eds.) (1999) *Quantum Fields and Strings: A Course for Mathematicians*. Providence, RI: American Mathematical Society.
- Douglas M *et al.* (eds.) (2004) *Strings and Geometry: Proceedings of the 2002 Clay School*. Providence, RI: American Mathematical Society.
- Gauntlett J (2004) Branes, calibrations and supergravity. In: Douglas M *et al.* (eds.) *Strings and Geometry*, pp. 79–126. Providence, RI: American Mathematical Society.
- Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*, 2 vols. Cambridge: Cambridge University Press.
- Li J and Yau S-T (2004) The existence of supersymmetric string theory with torsion, 2004 preprint, arXiv:hep-th/0411136.
- Polchinski J (1998) *String Theory*, 2 vols. Cambridge: Cambridge University Press.

Compressible Flows: Mathematical Theory

G-Q Chen, Northwestern University, Evanston, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Euler equations for compressible fluids consist of the conservation laws of mass, momentum, and energy:

$$\partial_t \rho + \nabla_x \cdot \mathbf{m} = 0, \quad \mathbf{x} \in \mathbf{R}^d \quad [1]$$

$$\partial_t \mathbf{m} + \nabla_x \cdot \left(\frac{\mathbf{m} \otimes \mathbf{m}}{\rho} \right) + \nabla_x p = 0 \quad [2]$$

$$\partial_t E + \nabla_x \cdot \left(\frac{\mathbf{m}}{\rho} (E + p) \right) = 0 \quad [3]$$

Equivalently, these correspond to the general form of nonlinear hyperbolic systems of conservation laws:

$$\partial_t \mathbf{u} + \nabla_x \cdot \mathbf{f}(\mathbf{u}) = 0, \quad \mathbf{x} \in \mathbf{R}^d, \quad \mathbf{u} \in \mathbf{R}^n \quad [4]$$

System [1]–[3] is closed by the following constitutive relations:

$$p = p(\rho, e), \quad E = \frac{1}{2} \frac{|\mathbf{m}|^2}{\rho} + \rho e \quad [5]$$

In [1]–[3] and [5], $\tau = 1/\rho$ is the deformation gradient (specific volume for fluids, strain for solids), $\mathbf{v} = (v_1, \dots, v_d)^\top$ is the fluid velocity with $\rho \mathbf{v} = \mathbf{m}$ the momentum vector, p is the scalar pressure, and E is the total energy with e the internal energy which is a given function of (τ, p) or (ρ, p) defined through thermodynamical relations. The other two thermodynamic variables are temperature θ and entropy S . If (ρ, S) are chosen as

independent variables, then the constitutive relations can be written as

$$(e, p, \theta) = (e(\rho, S), p(\rho, S), \theta(\rho, S)) \quad [6]$$

governed by $\theta dS = de + pd\tau = de - pd\rho/\rho^2$. For polytropic gases,

$$\begin{aligned} p &= p(\rho, S) = \kappa \rho^\gamma e^{S/c_v} \\ e &= \frac{p}{(\gamma - 1)\rho} \\ \theta &= \frac{p}{R\rho} \end{aligned} \quad [7]$$

where $R > 0$ may be taken to be the universal gas constant divided by the effective molecular weight of the particular gas, $c_v > 0$ is the specific heat at constant volume, $\gamma = 1 + R/c_v > 1$ is the adiabatic exponent, and κ can be any positive constant under scaling.

The most important criterion of applicability of any mathematical model is its well-posedness: existence, uniqueness, and stability. The well-posedness theory for compressible fluid flows is far from being complete, and many further issues are still unexplored. In particular, the global existence and uniqueness of solutions in \mathbf{R}^d , $d \geq 2$, is still a major open problem, and only partial results shed some lights on the amazing complexity of the problem. Below, we will mainly focus on the well-posedness issues with emphasis on the Cauchy problem, the initial value problem:

$$\mathbf{u}|_{t=0} = \mathbf{u}_0 \quad [8]$$

first for inviscid compressible fluid flows and then for viscous compressible fluid flows.

Throughout this article, where a cited reference is not shown in the “Further reading” section, it may usually be found by consulting Bressan (2000),

Chen (2005), Dafermos (2005), Feireisl (2004), Lions (1986, 1988) or Liv (2000).

Inviscid Compressible Fluid Flows: Euler Equations

Solutions to the Euler equations [1]–[3] are generically discontinuous functions obeying the Clausius–Duhem inequality, the second law of thermodynamics:

$$\partial_t(\rho S) + \nabla_x \cdot (\mathbf{m}S) \geq 0 \quad [9]$$

in the sense of distributions. Such discontinuous solutions are called entropy solutions.

When a flow is isentropic, that is, entropy S is a uniform constant S_0 in the flow, then the Euler equations for the flow take the simpler form:

$$\begin{aligned} \partial_t \rho + \nabla_x \cdot \mathbf{m} &= 0 \\ \partial_t \mathbf{m} + \nabla_x \cdot (\mathbf{m} \otimes \mathbf{m} / \rho) + \nabla_x p &= 0 \end{aligned} \quad [10]$$

where the pressure is a function of the density, $p = p(\rho, S_0)$, with constant S_0 . For a polytropic gas,

$$p(\rho) = \kappa \rho^\gamma, \quad \gamma > 1 \quad [11]$$

where κ can be any positive constant by scaling. This system can be derived from [1] to [3] as follows: for smooth solutions of [1]–[3], entropy $S(\rho, \mathbf{m}, E)$ is conserved along fluid particle trajectories, that is,

$$\partial_t(\rho S) + \nabla_x \cdot (\mathbf{m}S) = 0$$

If the entropy is initially a uniform constant and the solution remains smooth, then the energy equation can be eliminated and entropy S keeps the same constant in later time. Thus, under constant initial entropy, a smooth solution of [1]–[3] satisfies the equations in [10]. Furthermore, solutions of system [10] are also a good approximation to solutions of system [1]–[3] even after shocks form, since the entropy increases across a shock to the third order in wave strength for solutions of [1]–[3], while in [10] the entropy is constant. Moreover, system [10] is an excellent model for the isothermal fluid flow with $\gamma = 1$ and for the shallow-water flow with $\gamma = 2$. For such barotropic flows (i.e., $p = p(\rho)$), the energy equation [3] serves as an entropy inequality (see Lax (1973)):

$$\begin{aligned} \partial_t E + \nabla_x \cdot (\mathbf{m}(E + p(\rho)) / \rho) &\leq 0 \\ \text{in the sense of distributions} \end{aligned}$$

In the one-dimensional case, system [1]–[3] in Eulerian coordinates is

$$\begin{aligned} \partial_t \rho + \partial_x m &= 0, & \partial_t m + \partial_x (m^2 / \rho + p) &= 0 \\ \partial_t E + \partial_x (m(E + p) / \rho) &= 0 \end{aligned} \quad [12]$$

The system above can be rewritten in Lagrangian coordinates:

$$\begin{aligned} \partial_t \tau - \partial_x v &= 0, & \partial_t v + \partial_x p &= 0 \\ \partial_t (e + v^2 / 2) + \partial_x (pv) &= 0 \end{aligned} \quad [13]$$

with $v = m / \rho$, where the coordinates (t, x) are the Lagrangian coordinates, which are different from the Eulerian coordinates for [12]; for simplicity of notations, we do not distinguish them. For the barotropic case, systems [12] and [13] reduce to

$$\partial_t \rho + \partial_x m = 0, \quad \partial_t m + \partial_x (m^2 / \rho + p) = 0 \quad [14]$$

and

$$\partial_t \tau - \partial_x v = 0, \quad \partial_t v + \partial_x p = 0 \quad [15]$$

respectively, where pressure $p = p(\rho) = \tilde{p}(\tau)$, $\tau = 1 / \rho$. The solutions of [12] and [13], as well as [14] and [15], are equivalent even for entropy solutions with vacuum where $\rho = 0$.

The potential flow is well known in transonic aerodynamics, beyond the isentropic approximation [10] from [1] to [3]. Denote $D_t = \partial_t + \sum_{k=1}^d v_k \partial_{x_k}$ the convective derivative along fluid particle trajectories. From [1] to [3], we have

$$D_t S = 0 \quad [16]$$

and, by taking the curl of the momentum equations,

$$D_t \left(\frac{\omega}{\rho} \right) = \frac{\omega}{\rho} \cdot \nabla_x \mathbf{v} + \frac{p_S(\rho, S)}{\rho^3} \nabla_x \rho \times \nabla_x S \quad [17]$$

The identities [16] and [17] imply that a smooth solution of [1]–[3] which is both isentropic and irrotational at time $t = 0$ remains isentropic and irrotational for all later times, as long as this solution stays smooth. Then, the conditions $S = S_0 = \text{const.}$ and $\omega = \text{curl}_x \mathbf{v} = 0$ are reasonable for smooth solutions. For a smooth irrotational solution, we integrate the d -momentum equations in [10] through Bernoulli's law:

$$\partial_t v + \nabla_x (|v|^2 / 2) + \nabla_x h(\rho) = 0$$

where $h'(\rho) = p_\rho(\rho, S_0) / \rho$. On a simply connected space region, the condition $\text{curl}_x \mathbf{v} = 0$ implies that there exists Φ such that $\mathbf{v} = \nabla_x \Phi$. Then,

$$\begin{aligned} \partial_t \rho + \nabla_x \cdot (\rho \nabla_x \Phi) &= 0 \\ \partial_t \Phi + \frac{1}{2} |\nabla_x \Phi|^2 + h(\rho) &= K \end{aligned} \quad [18]$$

for some constant K . From the second equation in [18], we have

$$\rho(D\Phi) = h^{-1}(K - (\partial_t \Phi + \frac{1}{2} |\nabla_x \Phi|^2))$$

Then, system [18] can be rewritten as the following time-dependent potential flow equation of second order:

$$\partial_t \rho(D\Phi) + \nabla_x \cdot (\rho(D\Phi)\nabla_x \Phi) = 0 \quad [19]$$

For a steady solution $\Phi = \Phi(\mathbf{x})$, that is, $\partial_t \Phi = 0$, we obtain the celebrated steady potential flow equation of aerodynamics:

$$\nabla_x \cdot (\rho(\nabla_x \Phi)\nabla_x \Phi) = 0 \quad [20]$$

In applications in aerodynamics, [18] or [19] is used for discontinuous solutions, and the empirical evidence is that entropy solutions of [18] or [19] are fairly good approximations to entropy solutions for [1]–[3] provided that (1) the shock strengths are small, (2) the curvature of shock fronts is not too large, and (3) there is a small amount of vorticity in the region of interest. Model [19] or [18] is an excellent model to capture multidimensional shock waves by ignoring vorticity waves, while the incompressible Euler equations are an excellent model to capture multidimensional vorticity waves by ignoring shock waves.

Local Well-Posedness for Classical Solutions

Consider the Cauchy problem for the Euler equations [1]–[3] with Cauchy data [8]:

Assume that $\mathbf{u}_0 : \mathbf{R}^d \rightarrow \mathcal{D}$ is in $H^s \cap L^\infty$ with $s > d/2 + 1$. Then, for the Cauchy problem [1]–[3] and [8], there exists a finite time $T = T(\|\mathbf{u}_0\|_s, \|\mathbf{u}_0\|_{L^\infty}) \in (0, \infty)$ such that there is a unique, stable bounded classical solution $\mathbf{u} \in C^1([0, T] \times \mathbf{R}^d)$ with $\mathbf{u}(t, \mathbf{x}) \in \mathcal{D}$ for $(t, \mathbf{x}) \in [0, T] \times \mathbf{R}^d$ and $\mathbf{u} \in C([0, T]; H^s) \cap C^1([0, T]; H^{s-1})$. Moreover, the interval $[0, T)$ with $T < \infty$ is the maximal interval of the classical H^s existence for [1]–[3] if and only if either $\|(\mathbf{u}_t, \nabla_x \mathbf{u})\|_{L^\infty} \rightarrow \infty$ or $\mathbf{u}(t, \mathbf{x})$ escapes every compact subset $K \Subset \mathcal{D}$ as $t \rightarrow T$.

This local existence can be established by relying solely on the elementary linear existence theory for symmetric hyperbolic systems with smooth coefficients (cf. Majda (1984)), or by the abstract semigroup theory (Kato 1975).

Formation of Singularities

For the one-dimensional case, singularities include the development of shock waves and formation of vacuum states. For the multidimensional case, the situation is much more complicated: besides shock waves and vacuum states, singularities can also be generated from vortex sheets, focusing and breaking of waves, among others.

Consider the Cauchy problem of the Euler equations [1]–[3] in \mathbf{R}^3 for polytropic gases with smooth initial data:

$$\begin{aligned} (\rho, \mathbf{v}, S)|_{t=0} &= (\rho_0, \mathbf{v}_0, S_0)(\mathbf{x}) \\ \rho_0(\mathbf{x}) &> 0, \quad \mathbf{x} \in \mathbf{R}^3 \end{aligned} \quad [21]$$

satisfying $(\rho_0, \mathbf{v}_0, S_0)(\mathbf{x}) = (\bar{\rho}, 0, \bar{S})$ for $|\mathbf{x}| \geq L$, where $\bar{\rho} > 0$, \bar{S} , and L are given constants. The equations possess a unique local C^1 solution $(\rho, \mathbf{v}, S)(t, \mathbf{x})$ with $\rho(t, \mathbf{x}) > 0$ provided that the initial data [21] is sufficiently regular. The support of the smooth disturbance $(\rho_0(\mathbf{x}) - \bar{\rho}, \mathbf{v}_0(\mathbf{x}), S_0(\mathbf{x}) - \bar{S})$ propagates with speed at most $\sigma = \sqrt{p_\rho(\bar{\rho}, \bar{S})}$ (the sound speed), that is,

$$(\rho, \mathbf{v}, S)(t, \mathbf{x}) = (\bar{\rho}, 0, \bar{S}) \quad \text{if } |\mathbf{x}| \geq L + \sigma t \quad [22]$$

Define

$$\begin{aligned} P(t) &= \int_{\mathbf{R}^3} (p(\rho(t, \mathbf{x}), S(t, \mathbf{x}))^{1/\gamma} - p(\bar{\rho}, \bar{S})^{1/\gamma}) \, d\mathbf{x} \\ F(t) &= \int_{\mathbf{R}^3} (\rho \mathbf{v})(t, \mathbf{x}) \cdot \mathbf{x} \, d\mathbf{x} \end{aligned}$$

which, roughly speaking, measure the entropy and the radial component of momentum. Then, if $(\rho, \mathbf{v}, S)(t, \mathbf{x})$ is a C^1 solution of [1]–[3] and [21] for $0 < t < T$, and

$$\begin{aligned} P(0) &\geq 0, \quad F(0) > \alpha \sigma R^4 \max_{\mathbf{x}} \rho_0(\mathbf{x}) \\ \text{with } \alpha &= 16\pi/3 \end{aligned} \quad [23]$$

then the lifespan T of the C^1 solution is finite (Sideris 1985).

To illustrate a way in which the conditions in [23] may be satisfied, consider the initial data: $\rho_0 = \bar{\rho}$, $S_0 = \bar{S}$. Then $P(0) = 0$, and [23] holds if

$$\int_{|\mathbf{x}| < R} \mathbf{v}_0(\mathbf{x}) \cdot \mathbf{x} \, d\mathbf{x} > \alpha \sigma R^4$$

Comparing both sides, one finds that the initial velocity must be supersonic in some region relative to the sound speed at infinity. The formation of a singularity (presumably a shock wave) is detected as the disturbance overtakes the wave front forcing the front to propagate with supersonic speed.

Singularities are formed even without the condition of largeness, such as [23], being satisfied. For example, if $S_0(\mathbf{x}) \geq \bar{S}$ and, for some $0 < R_0 < R$,

$$\begin{aligned} \int_{|\mathbf{x}| > r} |\mathbf{x}|^{-1} (|\mathbf{x}| - r)^2 (\rho_0(\mathbf{x}) - \bar{\rho}) \, d\mathbf{x} &> 0 \\ \int_{|\mathbf{x}| > r} |\mathbf{x}|^{-3} (|\mathbf{x}|^2 - r^2) \rho_0(\mathbf{x}) \mathbf{v}_0(\mathbf{x}) \cdot \mathbf{x} \, d\mathbf{x} &\geq 0 \end{aligned} \quad [24]$$

for $R_0 < r < R$, then the lifespan T of the C^1 solution of [1]–[3] and [21] is finite. The

assumptions in [24] mean that, in an average sense, the gas must be slightly compressed and outgoing directly behind the wave front.

Local Well-Posedness for Shock-Front Solutions

For a general hyperbolic system of conservation laws [4], shock-front solutions are discontinuous, piecewise smooth entropy solutions with the following structure:

1. There exists a C^2 spacetime hypersurface $\mathcal{S}(t)$ defined in (t, \mathbf{x}) for $0 \leq t \leq T$ with spacetime normal $(\nu_t, \nu_x) = (\nu_t, \nu_1, \dots, \nu_d)$ as well as two C^1 vector-valued functions: $\mathbf{u}^+(t, \mathbf{x})$ and $\mathbf{u}^-(t, \mathbf{x})$, defined on respective domains \mathcal{D}^+ and \mathcal{D}^- on either side of the hypersurface $\mathcal{S}(t)$ and satisfying $\partial_t \mathbf{u}^\pm + \nabla_x \cdot \mathbf{f}(\mathbf{u}^\pm) = 0$ in \mathcal{D}^\pm ;
2. The jump across the hypersurface $\mathcal{S}(t)$ satisfies the Rankine–Hugoniot condition:

$$\{\nu_t(\mathbf{u}^+ - \mathbf{u}^-) + \nu_x \cdot (\mathbf{f}(\mathbf{u}^+) - \mathbf{f}(\mathbf{u}^-))\}_{\mathcal{S}} = 0$$

For [4], the surface \mathcal{S} is not known in advance and must be determined as part of the solution of the problem; thus, the two equations in (1)–(2) describe a multidimensional, highly nonlinear, free-boundary-value problem. The initial data yielding shock-front solutions is defined as follows. Let \mathcal{S}_0 be a smooth hypersurface parametrized by α , and let $\nu(\alpha) = (\nu_1, \dots, \nu_d)(\alpha)$ be a unit normal to \mathcal{S}_0 . Define the piecewise smooth initial values for respective domains \mathcal{D}_0^+ and \mathcal{D}_0^- on either side of the hypersurface \mathcal{S}_0 as

$$\mathbf{u}_0(\mathbf{x}) = \begin{cases} \mathbf{u}_0^+(\mathbf{x}), & \mathbf{x} \in \mathcal{D}_0^+ \\ \mathbf{u}_0^-(\mathbf{x}), & \mathbf{x} \in \mathcal{D}_0^- \end{cases} \quad [25]$$

It is assumed that the initial jump in [25] satisfies the Rankine–Hugoniot condition, that is, there is a smooth scalar function $\sigma(\alpha)$ so that

$$-\sigma(\alpha)(\mathbf{u}_0^+(\alpha) - \mathbf{u}_0^-(\alpha)) + \nu(\alpha) \cdot (\mathbf{f}(\mathbf{u}_0^+(\alpha)) - \mathbf{f}(\mathbf{u}_0^-(\alpha))) = 0 \quad [26]$$

and that $\sigma(\alpha)$ does not define a characteristic direction, that is,

$$\sigma(\alpha) \neq \lambda_i(\mathbf{u}_0^\pm), \quad \alpha \in \bar{\mathcal{S}}_0, \quad 1 \leq i \leq n \quad [27]$$

where $\lambda_i, i = 1, \dots, n$, are the eigenvalues of [4]. It is natural to require that $\mathcal{S}(0) = \mathcal{S}_0$.

Consider the Euler equations [1]–[3] in \mathbf{R}^3 for polytropic gases with piecewise smooth initial data:

$$(\rho, \mathbf{v}, E)|_{t=0} = \begin{cases} (\rho_0^+, \mathbf{v}_0^+, E^+)(\mathbf{x}), & \mathbf{x} \in \mathcal{D}_0^+ \\ (\rho_0^-, \mathbf{v}_0^-, E^-)(\mathbf{x}), & \mathbf{x} \in \mathcal{D}_0^- \end{cases} \quad [28]$$

Assume that \mathcal{S}_0 is a smooth compact surface in \mathbf{R}^3 and that $(\rho_0^\pm, \mathbf{v}_0^\pm, E_0^\pm)(\mathbf{x})$ belongs to the uniform local

Sobolev space $H_{\text{ul}}^s(\mathcal{D}_0^+)$, while $(\rho_0^-, \mathbf{v}_0^-, E_0^-)(\mathbf{x})$ belongs to the Sobolev space $H^s(\mathcal{D}_0^-)$, for some fixed $s \geq 10$. Assume also that there is a function $\sigma(\alpha) \in H^s(\mathcal{S}_0)$ so that [26] and [27] hold, and the compatibility conditions up to order $s - 1$ are satisfied on \mathcal{S}_0 by the initial data, together with the entropy condition:

$$\begin{aligned} \mathbf{v}_0^+ \cdot \nu(\alpha) + \sqrt{p_\rho(\rho_0^+, \mathcal{S}_0^+)} &< \sigma(\alpha) \\ &< \mathbf{v}_0^- \cdot \nu(\alpha) + \sqrt{p_\rho(\rho_0^-, \mathcal{S}_0^-)} \end{aligned} \quad [29]$$

Then, there are a C^2 hypersurface $\mathcal{S}(t)$ and C^1 functions $(\rho^\pm, \mathbf{v}^\pm, E^\pm)(t, \mathbf{x})$ defined for $t \in [0, T]$, with T sufficiently small, so that

$$(\rho, \mathbf{v}, E)(t, \mathbf{x}) = \begin{cases} (\rho^+, \mathbf{v}^+, E^+)(t, \mathbf{x}), & (t, \mathbf{x}) \in \mathcal{D}^+ \\ (\rho^-, \mathbf{v}^-, E^-)(t, \mathbf{x}), & (t, \mathbf{x}) \in \mathcal{D}^- \end{cases} \quad [30]$$

is the discontinuous shock-front solution of the Cauchy problem [1]–[3] and [28]. Here a vector function \mathbf{u} is in H_{ul}^s , provided that there exists some $r > 0$ so that $\max_{\mathbf{y} \in \mathbf{R}^d} \|w_{r, \mathbf{y}} \mathbf{u}\|_{H^s} < \infty$ with $w_{r, \mathbf{y}}(\mathbf{x}) = w((\mathbf{x} - \mathbf{y})/r)$, where $w \in C_0^\infty(\mathbf{R}^d)$ is a function so that $w(\mathbf{x}) \geq 0, w(\mathbf{x}) = 1$ when $|\mathbf{x}| \leq 1/2$, and $w(\mathbf{x}) = 0$ when $|\mathbf{x}| > 1$.

The compatibility conditions are needed in order to avoid the formation of discontinuities in higher derivatives along other characteristic surfaces emanating from \mathcal{S}_0 : Once the main condition [26] is satisfied, the compatibility conditions are automatically guaranteed for a wide class of initial data. The idea of the proof is to use the existence of a strictly convex entropy and the symmetrization of [4]; the shock-front solutions are defined as the limit of a convergent classical iteration scheme based on a linearization by using the theory of linearized stability for shock fronts (Majda 1984). The uniform existence time of shock-front solutions in shock strength can be achieved (Métivier 1990).

Global Theory in L^∞ for the Isentropic Euler Equations for $x \in \mathbf{R}$

Consider the Cauchy problem for [14] with initial data:

$$(\rho, m)|_{t=0} = (\rho_0, m_0)(x) \quad [31]$$

where ρ_0 and m_0 are in the physical region $\{(\rho, m) : \rho \geq 0, |m| \leq C_0 \rho\}$ for some $C_0 > 0$. System [14] is strictly hyperbolic at the states with $\rho > 0$, and strict hyperbolicity fails at the vacuum states $V := \{(\rho, m/\rho) : \rho = 0, |m/\rho| < \infty\}$. Then, we have:

1. There exists a global solution $(\rho, m)(t, x)$ of the Cauchy problem [14] and [31] satisfying

$$0 \leq \rho(t, x) \leq C, \quad |m(t, x)| \leq C\rho(t, x) \quad [32]$$

for some $C > 0$ depending only on C_0 and γ , and the entropy inequality

$$\partial_t \eta(\rho, m) + \partial_x q(\rho, m) \leq 0 \tag{33}$$

in the sense of distributions for any convex weak entropy–entropy flux pair (η, q) , that is,

$$\nabla q(\rho, m) = \nabla \eta(\rho, m) \nabla f(\rho, m)$$

with

$$\nabla^2 \eta(\rho, m) \geq 0 \quad \text{and} \quad \eta|_V = 0$$

2. The solution operator $(\rho, m)(t, \cdot) = S_t(\rho_0, m_0)(\cdot)$, determined by (1), is compact in $L^1_{loc}(\mathbf{R})$ for $t > 0$;
3. Furthermore, if $(\rho_0, m_0)(x)$ is periodic with period P , then there exists a global periodic solution $(\rho, m)(t, x)$ with [32] such that $(\rho, m)(t, x)$ asymptotically decays to

$$\frac{1}{|P|} \int_P (\rho_0, m_0)(x) dx$$

in L^1 .

The convergence of the Lax–Friedrichs scheme, the Godunov scheme, and the vanishing viscosity method for system [14] have also been established.

The results are based on a compensated compactness framework to replace the BV compactness framework. For a gas obeying the γ -law, the case $\gamma = (N + 2)/N$, $N \geq 5$ odd, was first studied by DiPerna (1983), and the case $1 < \gamma \leq 5/3$ for usual gases was first solved by Chen (1986) and Ding–Chen–Luo (1985). The cases $\gamma \geq 3$ and $5/3 < \gamma < 3$ were treated by Lions–Perthame–Tadmor (1994) and Lions–Perthame–Souganidis (1996), respectively. The case of general pressure laws was solved by Chen–LeFloch (2000, 2003). All the results for entropy solutions to [14] in Eulerian coordinates can equivalently be presented as the corresponding results for entropy solutions to [15] in Lagrangian coordinates. The isothermal case $\gamma = 1$ was treated by Huang–Wang (2002).

Global Theory in BV for the Adiabatic Euler Equations for $x \in \mathbf{R}$

Consider the Euler equations [13] for polytropic gases with the Cauchy data:

$$(\tau, v, S)|_{t=0} = (\tau_0, v_0, S_0)(x) \tag{34}$$

Then we have (Liu 1977, Temple 1981, Chen and Wagner 2003):

Let $K \subset \{(\tau, v, S) : \tau > 0\}$ be a compact set in $\mathbf{R}_+ \times \mathbf{R}^2$, and let $N \geq 1$ be any constant. Then there exists a constant $C_0 = C_0(K, N)$, independent of $\gamma \in (1, 5/3]$,

such that, for every initial data $(\tau_0, v_0, S_0) \in K$ with $TV_{\mathbf{R}}(\tau_0, v_0, S_0) \leq N$, when

$$(\gamma - 1)TV_{\mathbf{R}}(\tau_0, v_0, S_0) \leq C_0 \quad \text{for any } \gamma \in (1, 5/3]$$

the Cauchy problem [13] and [34] has a global entropy solution $(\tau, v, S)(t, x)$ which is bounded and satisfies

$$TV_{\mathbf{R}}(\tau, v, S)(t, \cdot) \leq C TV_{\mathbf{R}}(\tau_0, v_0, S_0)$$

for some constant $C > 0$ independent of γ .

This result specially includes that for the barotropic case (Nishida 1968, Nishida–Smoller 1973, DiPerna 1973). Some efforts in the direction of relaxing the requirement of small total variation have been made. Some extensions to the initial-boundary value problems have also been made. In addition, an entropy solution in BV with periodic data or compact support decays when $t \rightarrow 0$. Furthermore, even for a general hyperbolic system [4] for $x \in \mathbf{R}$, we have:

If the initial data functions $u_0(x)$ and $v_0(x)$ have sufficiently small total variation and $u_0 - v_0 \in L^1(\mathbf{R})$, then, for the corresponding exact Glimm, or wave-front tracking, or vanishing viscosity solutions $u(t, x)$ and $v(t, x)$ of the Cauchy problem [4] and [8], there exists a constant $C > 0$ such that

$$\|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbf{R})} \leq C \|u_0 - v_0\|_{L^1(\mathbf{R})} \tag{35}$$

for all $t > 0$

An immediate consequence is that the whole sequence of the approximate solutions constructed by the Glimm (1965) scheme, as well as the wave-front tracking method and the vanishing viscosity method, converges to a unique entropy solution of [4] and [8] when the mesh size or the viscosity coefficient tends to zero. More detailed discussions and extensive references about the L^1 -stability of BV entropy solutions and related topics can be found in Bressan (2000) and Dafermos (2000); also see Chen and Wang (2002). Furthermore, the Riemann solution is unique and asymptotically stable in the class of entropy solutions to [13] with large variation satisfying only one physical entropy inequality (Chen–Frid–Li 2002).

Multidimensional Steady Theory

The mathematical study of two-dimensional steady supersonic flows past wedges, whose vertex angles are less than the critical angle, can date back to the 1940s, since the stability of such flows is fundamental in applications (cf. Courant–Friedrichs (1948)). Local solutions around the wedge vertex were first constructed (Gu 1962, Schaeffer 1976, Li 1980).

Such global potential solutions were constructed when the wedge has some convexity, or is a small perturbation of the straight wedge with fast decay in the flow direction (Chen 2001, Chen-Xin-Yin 2002), or is piecewise smooth which is a small perturbation of straight wedge (Zhang 2003). For the two-dimensional steady supersonic flows governed by the full Euler equations past Lipschitz wedges, it indicates (Chen-Zhang-Zhu 2005a) that, when the wedge vertex angle is less than the critical angle, the strong shock front emanating from the wedge vertex is nonlinearly stable in structure globally, although there may be many weak shocks and vortex sheets between the wedge boundary and the strong shock front, under the BV perturbation of the wedge so that the total variation of the tangent function along the wedge boundary is suitably small. This asserts that any supersonic shock for the wedge problem is nonlinearly stable.

A self-similar gas flow past an infinite cone in \mathbf{R}^3 with small vertex angle is also nonlinearly stable upon the BV perturbation of the obstacle (Lien-Liu 1999). It is still open for the nonlinear stability when the infinite cone in \mathbf{R}^3 has arbitrary vertex angle. The stability issues of supersonic vortex sheets have been studied by classical linearized stability analysis, large-scale numerical simulations, and asymptotic analysis. In particular, the nonlinear development of instabilities of supersonic vortex sheets at high Mach number was predicted as time evolves (Woodward 1985, Artola-Majda 1989). In contrast with the prediction of evolution instability, steady supersonic vortex sheets, as time-asymptotics, are stable globally in structure, even under the BV perturbation of the Lipschitz walls, although there may be many weak shocks and supersonic vortex sheets away from the strong vortex sheet (Chen-Zhang-Zhu 2005b).

Transonic shock problems for steady fluid flows are important in applications (cf. Courant and Friedrichs (1948)). A program on the existence and stability of multidimensional transonic shocks has been initiated and three new analytical approaches have been developed (Chen-Feldman 2003, 2004). The transonic problems include the existence and stability of transonic shocks in the whole \mathbf{R}^d , the existence and stability of transonic flows past finite or infinite nozzles, the stability of transonic flows past infinite nonsmooth wedges, and the existence of regular shock reflection solutions. The first approach is an iteration scheme based on the nondegeneracy of the free boundary condition: the jump of the normal derivative of a solution across

the free boundary has a strictly positive lower bound (Chen-Feldman 2003, 2004), which works for the nonlinear equations whose coefficients may depend on not only the solution itself but also the gradients of the solution. The second approach is a partial hodograph procedure, with which the existence and stability of multidimensional transonic shocks that are not nearly orthogonal to the flow direction can be handled (Chen-Feldman 2004): one of the main ingredients in this approach is to employ a partial hodograph transform to reduce the free boundary problem into a conormal boundary value problem for the corresponding nonlinear equations of divergence form and then develop techniques to solve the conormal boundary value problem. When the regularity of the steady perturbation is $C^{3,\alpha}$ or higher, the third approach is to employ the implicit function theorem to deal with the existence and stability problem. Another iteration approach, which works well for the two-dimensional equations whose coefficients depend only on the solution itself, has also been developed (Canic-Keyfitz-Lieberman 2000).

Further longstanding open problems include the existence of global transonic flows past an airfoil or a smooth obstacle (Morawetz 1956–58, 1985).

Multidimensional Unsteady Problems

Now we present some multidimensional time-dependent problems with a simplifying feature that the data (domain and/or the initial data) coupled with the structure of the underlying equations obey certain geometric structure so that the multidimensional problems can be reduced to lower-dimensional problems with more complicated couplings. Different types of geometric structure call for different techniques.

The Euler equations for compressible fluids with geometric structure describe many important fluid flows, including spherically symmetric flows and self-similar flows. Such geometric flows are motivated by many physical problems such as shock diffractions, supernovas formation in stellar dynamics, inertial confinement fusion, and underwater explosions. For the initial data with large amplitude having geometric structure, the required physical insight is: (1) whether the solution has the same geometric structure globally and (2) whether the solution blows up to infinity in a finite time. These questions are not easily understood in physical experiments and numerical simulations, especially for the blow-up, because of the limited capacity of available instruments and computers.

The first type of geometric structure is spherical symmetry. A criterion for L^∞ Cauchy data functions of arbitrarily large amplitude was observed to guarantee the existence of spherically symmetric solutions in L^∞ in the large for the isentropic flows, which model outgoing blast waves and large-time asymptotic solutions (Chen 1997). On the other hand, it is evident that the density blows up as $|\mathbf{x}| \rightarrow 0$ in general, especially for the focusing case; the singularity at the origin makes the problem truly multidimensional due to the reflection of waves from infinity and their strengthening as they move radially inwards. One of the important open questions is to understand the order of singularity, $\rho(t, |\mathbf{x}|) \sim |\mathbf{x}|^{-\alpha}$, at the origin for bounded Cauchy data.

The second type of geometric structure is self-similarity, that is, the solutions with initial data functions that give rise to self-similar solutions, especially including Riemann solutions. Compressible flow equations in \mathbf{R}^d , $d \geq 2$, with one or more linearly degenerate modes of wave propagation have additional difficulties. In that case, the global flow is governed by a reduced (self-similar) system which is of composite (hyperbolic–elliptic) type in the subsonic region. The linearly degenerate waves give rise to one or more families of degenerate characteristics which remain real in the subsonic region. In some cases, the reduced equations couple an elliptic (degenerate elliptic) problem for the density with a hyperbolic (transport) equation for the vorticity.

An important prototype for both practical applications and the theory of multidimensional complex wave patterns is the problem of diffraction of a shock wave which is incident along an inclined ramp (see Glimm and Majda (1991)). When a plane shock hits a wedge head-on, a self-similar reflected shock moves outward as the original shock moves forward. The computational and asymptotic analysis shows that various patterns of reflected shocks may occur, including regular reflection and (simple, double, and complex) Mach reflections. The main part or whole reflected shock is a transonic shock in the self-similar coordinates, for which the corresponding equation changes the type from hyperbolic to elliptic across the shock. There are few rigorous mathematical results on the global existence and stability of shock reflection solutions and the transition among regular, simple Mach, double Mach, and complex Mach reflections for the potential flow equation [19] and the full Euler equations [1]–[3]. Some results were recently obtained for simplified models including the transonic small-disturbance equation near the reflection point and the pressure

gradient equation when the wedge is close to a flat wall.

For the potential flow equation [19], a self-similar solution is a solution of the form: $\Psi = t\phi(\mathbf{y})$, $\mathbf{y} = \mathbf{x}/t$. Letting $\varphi(\mathbf{y}) = -\mathbf{y}^2/2 + \phi(\mathbf{y})$, then the system can be rewritten in the form of a second-order equation of mixed hyperbolic–elliptic type in $\mathbf{y} \in \mathbf{R}^d$ by scaling:

$$\nabla_{\mathbf{y}} \cdot (\rho(|\nabla_{\mathbf{y}}\varphi|^2, \varphi)\nabla_{\mathbf{y}}\varphi) + d\rho(|\nabla_{\mathbf{y}}\varphi|^2, \varphi) = 0 \quad [36]$$

with $\rho(q^2, z) = (1 - (q^2 + 2z)/2)^{1/(\gamma-1)}$. Equation [36] at $|\nabla_{\mathbf{y}}\varphi| = q$ is hyperbolic (pseudosupersonic) if $\rho(q^2, z) + q\rho_q(q^2, z) < 0$ and elliptic (pseudosubsonic) if $\rho(q^2, z) + q\rho_q(q^2, z) > 0$. Under this framework, the nature of the shock reflection pattern has been explored for weak incident shocks (strength b) and small wedge angles $2\theta_w$ by a number of different scalings, a study of mixed equations, and matching asymptotics for the different scalings, where the parameter $\beta = c_1\theta_w^2/b(\gamma + 1)$ ranges from 0 to ∞ and c_1 is the speed of sound behind the incident shock (Morawetz 1994). For $\beta > 2$, a regular reflection of both strong and weak kinds is possible as well as a Mach reflection; for $\beta < 1/2$, a Mach reflection occurs and the flow behind the reflection is subsonic and can be constructed in principle (with an elliptic problem) and matched; and for $1/2 < \beta < 2$, the flow behind a Mach reflection may be transonic which is a solution of a nonlinear boundary-value problem of mixed type. The basic pattern of reflection has been shown to be an almost semicircular shock issuing, for a regular reflection, from the reflection point on the wedge and, for a Mach reflection, matched with a local interaction flow. Some related observations were also made (Keller-Blank 1951, Hunter-Keller 1984, Hunter 1988). It is important to establish rigorous proofs. Recently, a rigorous existence proof was established for global solutions to shock reflection by large-angle wedges in Chen and Feldman (2005).

Analytical Frameworks for Entropy Solutions

The recent great progress for entropy solutions for one-dimensional time-dependent Euler equations and two-dimensional steady Euler equations, based on BV, L^1 , or even L^∞ estimates, naturally arises the expectation that a similar approach may also be effective for the multidimensional Euler equations, or more generally, hyperbolic systems of conservation laws, especially,

$$\|\mathbf{u}(t, \cdot)\|_{\text{BV}} \leq C\|\mathbf{u}_0\|_{\text{BV}} \quad [37]$$

Unfortunately, this is not the case. The necessary condition for [37] to be held for $p \neq 2$ (Rauch 1986) is

$$\begin{aligned} \nabla f_k(\mathbf{u}) \nabla f_l(\mathbf{u}) &= \nabla f_l(\mathbf{u}) \nabla f_k(\mathbf{u}) \\ \text{for all } k, l &= 1, 2, \dots, d \end{aligned} \quad [38]$$

The analysis suggests that only systems in which the commutativity relation [38] holds offer any hope for treatment in the framework of BV. This special case includes the scalar case $n=1$ and the case of one space dimension $d=1$. Beyond that, it contains very few systems of physical interest.

In this regard, it is important to identify effective analytical frameworks for studying entropy solutions of the multidimensional Euler equations [1]–[3], which are not in BV. Naturally, we want to approach the questions of existence, stability, uniqueness, and long-time behavior of entropy solutions with as much generality as possible. For this purpose, a theory of divergence-measure fields to construct such a global framework has been developed for studying entropy solutions (Chen-Frid 1999, 2000, Chen-Torres 2005, Chen-Torres-Ziemer 2005). For more details, see Chen (2005).

Viscous Compressible Fluid Flows: Navier–Stokes Equations

Compressible fluid flows that are viscous and conduct heat are governed by the following Navier–Stokes equations:

$$\partial_t \rho + \nabla_x \cdot \mathbf{m} = 0, \quad \mathbf{x} \in \mathbf{R}^d \quad [39]$$

$$\partial_t \mathbf{m} + \nabla_x \cdot \left(\frac{\mathbf{m} \otimes \mathbf{m}}{\rho} \right) + \nabla_x p = \nabla_x \cdot \boldsymbol{\Sigma} \quad [40]$$

$$\partial_t E + \nabla_x \cdot \left(\frac{\mathbf{m}}{\rho} (E + p) \right) = \nabla_x \cdot \left(\frac{\mathbf{m}}{\rho} \cdot \boldsymbol{\Sigma} \right) - \nabla_x \cdot \mathbf{q} \quad [41]$$

Here, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\nabla_x \mathbf{v}, \rho, \theta)$ is the viscous stress tensor which is symmetric from the conservation of angular momentum and \mathbf{q} is the heat flux. If the fluid is isotropic and the viscous tensor $\boldsymbol{\Sigma}$ is a linear function of $\nabla_x \mathbf{v}$ and invariant under a change of reference frame (translation and rotation), then we deduce from elementary algebraic manipulations that necessarily

$$\boldsymbol{\Sigma} = \lambda(\rho, \theta) \nabla_x \cdot \mathbf{v} + 2\mu(\rho, \theta) D \quad [42]$$

which corresponds to the Newtonian fluids, where $D = (\nabla_x \mathbf{v} + (\nabla_x \mathbf{v})^\top)/2$ is the deformation tensor and λ and μ are the Lamé viscosity coefficients.

Furthermore, since the fluid is isotropic, we are led to the Fourier law:

$$\mathbf{q} = -k(\rho, \theta, |\nabla_x \theta|) \nabla_x \theta$$

for scalar function k which, in most cases, is taken to be simply a function of ρ and θ , or even a constant called the thermal conduction coefficient. Again, system [39]–[41] is closed by the constitutive relations in [5]. The equation for entropy S is

$$\begin{aligned} \partial_t(\rho S) + \nabla_x \cdot \left(\mathbf{m} S + \frac{\mathbf{q}}{\theta} \right) \\ = \frac{\boldsymbol{\Sigma}(\nabla_x \mathbf{v}) : \nabla_x \mathbf{v}}{\theta} - \frac{\mathbf{q} \cdot \nabla_x \theta}{\theta^2} \end{aligned} \quad [43]$$

The second law of thermodynamics indicates that the right-hand side of [43] should be non-negative which yields the restriction:

$$k(\rho, \theta, |\nabla_x \theta|) \geq 0, \quad \mu \geq 0, \quad \lambda + 2\mu/d \geq 0$$

The case $\mu > 0$ and $\lambda + \mu > 0$ is the viscous case with heat conductivity $k > 0$. In particular, the kinetic theory indicates that the Stokes relationship should hold, namely $\lambda = -2\mu/d$ and the adiabatic component $\gamma = 5/3$ for monatomic gases.

In mathematical viscous fluid dynamics, an important model is the barotropic model for viscous fluids, that is, $p = p(\rho)$. Then, the specific energy E can be taken in the form of $E = (1/2)\rho|\mathbf{v}|^2 + \rho e(\rho)$ with $e'(\rho) = p(\rho)/\rho^2$. For classical solutions, the energy of a barotropic flow satisfies the equality:

$$\partial_t E + \nabla_x \cdot ((E + p)\mathbf{v}) = \nabla_x \cdot (\boldsymbol{\Sigma} \mathbf{v}) - \boldsymbol{\Sigma} : \nabla_x \mathbf{v}$$

which is now a direct consequence of [39] and [40].

The question of local existence of classical solutions to [39]–[41] for regular initial data was addressed by Nash (1962), where there is no indication whether or not these solutions exist for all times.

In the case of one space dimension, the well-posedness is largely settled. The basic result for the existence of classical solutions is that of Kazhikhov (1976); see Lions (1998) and Feireisl (2004) for extensive references. The discontinuous solutions have been constructed (Shelukhin 1979, Serre 1986, Hoff 1987, Chen-Hoff-Trivisa 2000).

For the Navier–Stokes equations in \mathbf{R}^3 with general equation of state, the global classical solutions for the Cauchy problem and various initial-boundary value problems whose initial data is small around a constant state have been

constructed (Matsumura-Nishida 1980, 1983). The approach is to obtain *a priori* estimates via energy methods for extending the local solution or for a difference method globally. These results have been extended to the Cauchy problem or the initial-boundary value problems with small discontinuous initial data (Hoff 1997).

For the Navier–Stokes equations in \mathbf{R}^d for barotropic flows with [11] and large initial data, the global existence of solutions containing vacuum for the Cauchy problem or various initial-boundary value problems was first established by Lions (1998) for $\gamma \geq 3/2$ if $d=2$, $\gamma \geq 9/5$ if $d=3$, and $\gamma > d/2$ if $d \geq 4$. The gap was closed by Feireisl–Novotný–Petzeltová (2001) for the full range $\gamma > d/2$. These results have been extended to the full Navier–Stokes equations describing the motion of a general compressible, viscous, and heat conducting fluid (see Feireisl (2004)). The physically relevant isothermal case, $\gamma=1$, is completely open even if $d=2$. The only large data existence result is that for radially symmetric data (Hoff 1992). The general case $\gamma \geq 1$ and $d=3$ for radially symmetric data was solved only recently (Jiang-Zhang 2001).

The lower-bound estimate on the density is a delicate issue. Weak solutions containing vacuum for the isentropic viscous flows with constant viscosity are unstable in general (Hoff-Serre 1991). Hence, it is important to see whether vacuum will never develop if the initial data is away from vacuum; this has been shown for the one-dimensional case for large initial data and for the multidimensional case with small data. On the other hand, from the kinetic theory, if solutions contain vacuum, then the viscosity coefficients in the Navier–Stokes equations should depend on the density near vacuum; this indeed stabilizes the solutions for the one-dimensional case.

The stability of viscous shock waves has been studied for the one-dimensional case (see Liu (2000) and the references therein). The compressible–incompressible limits from the isentropic compressible to incompressible Navier–Stokes equations when the Mach number tends to zero have been established for arbitrarily weak solutions (Lions-Masmoudi 1998) and for smooth solutions and a class of initial data functions (Hoff 1998).

The inviscid limits from the Navier–Stokes equations to the Euler equations have been established as long as the solutions of the Euler equations are smooth, when the viscosity and heat conductivity coefficients tend to zero (Klainerman-Majda 1982). It is completely open for general entropy solutions, even in the one-dimensional case.

See also: Breaking Water Waves; Capillary Surfaces; Fluid Mechanics: Numerical Methods; Geophysical Dynamics; Incompressible Euler Equations: Mathematical Theory; Inviscid Flows; Magnetohydrodynamics; Newtonian Fluids and Thermohydraulics; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stability of Flows; Viscous Incompressible Fluids: Mathematical Theory.

Further Reading

- Bressan A (2000) *Hyperbolic Systems of Conservation Laws: The One-Dimensional Cauchy Problem*. Oxford: Oxford University Press.
- Chen G-Q (2005) Euler equations and related hyperbolic conservation laws. In: Dafermos CM and Feireisl E (eds.) *Handbook of Differential Equations II: Evolutionary Differential Equations*, Chapter 1, pp. 1–104. Amsterdam: Elsevier.
- Chen G-Q and Wang D (2002) The Cauchy problem for the Euler equations for compressible fluids. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. 1, ch. 5, pp. 421–543. Amsterdam: Elsevier Science B.V.
- Courant R and Friedrichs KO (1948) *Supersonic Flow and Shock Waves*. New York: Springer.
- Dafermos CM (2005) *Hyperbolic Conservation Laws in Continuum Physics* (2nd edn). Berlin: Springer.
- Feireisl E (2004) *Dynamics of Viscous Compressible Fluids*. Oxford: Oxford University Press.
- Glimm J (1965) Solutions in the large for nonlinear hyperbolic system of equations. *Communications on Pure and Applied Mathematics* 18: 95–105.
- Glimm J and Majda A (1991) *Multidimensional Hyperbolic Problems and Computations*. New York: Springer.
- Lax PD (1973) *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*. Philadelphia: SIAM.
- Lions PL (1996, 1998) *Mathematical Topics in Fluid Mechanics*, vols. 1–2. New York: Oxford University Press.
- Liu T-P (2000) *Hyperbolic and Viscous Conservation Laws*, CBMS-NSF RCSAM, vol. 72. Philadelphia: SIAM.
- Majda A (1984) *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*. New York: Springer.

Computational Methods in General Relativity: The Theory

M W Choptuik, University of British Columbia,
Vancouver, Canada

© 2006 Elsevier Ltd. All rights reserved.

Conventions and Units

This article adopts many of the conventions and notations of Misner, Thorne, and Wheeler (1973) – hereafter denoted MTW – including metric signature $(-+++)$; definitions of Christoffel symbols and curvature tensors (up to index permutations permitted by standard symmetries of the tensors in a coordinate basis); the use of Greek indices $\alpha, \beta, \gamma, \dots$, ranging over the spacetime coordinate values $(0, 1, 2, 3) \rightarrow (t, x^1, x^2, x^3)$, to denote the components of spacetime tensors such as $g_{\mu\nu}$; the similar use of Latin indices i, j, k, \dots , ranging over the spatial coordinate values $(1, 2, 3) \rightarrow (x^1, x^2, x^3)$, for spatial tensors such as γ_{ij} ; the use of the Einstein summation convention for both types of indices; the use of standard Kronecker delta symbols (tensors), δ^μ_ν and δ^i_j ; the choice of geometric units, $G = c = 1$; and, finally, the normalization of the matter fields implicit in the choice of the constant 8π in [1].

The majority of the equations that appear in this article are tensor equations, or specific components of tensor equations, written in traditional index (not abstract index) form. Thus, these equations are generally valid in any coordinate system, (t, x^i) , but, of course do require the introduction of a coordinate basis and its dual. This approach is also largely a matter of convention, since all of what follows can be derived in a variety of fashions, some of them purely geometrical, and there are also approaches to numerical relativity based, for example, on frames rather than coordinate bases.

This article departs from MTW in its use of α, β^i , and γ_{ij} to denote the lapse, shift, and spatial metric, respectively, rather than MTW's N, N^i , and ${}^{(3)}g_{ij}$.

Finally, the operations of partial differentiation with respect to coordinates x^μ, t , and x^i are denoted ∂_μ, ∂_t , and ∂_i , respectively.

Introduction

The numerical analysis of general relativity, or numerical relativity, is concerned with the use of computational methods to derive approximate solutions to the Einstein field equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \quad [1]$$

Here, $G_{\mu\nu}$ is the Einstein tensor – that contracted piece of the Riemann curvature tensor that has vanishing divergence – and $T_{\mu\nu}$ is the stress tensor of the matter content of the spacetime. $T_{\mu\nu}$ likewise has vanishing divergence, an expression of the principle of local conservation of stress–energy that general relativity embodies.

The elegant tensor formulation [1] belies the fact that, ultimately, the field equations are generically a complicated and nonlinear set of partial differential equations (PDEs) for the components of the spacetime metric tensor, $g_{\mu\nu}(x^\alpha)$, in some coordinate system x^α . Moreover, implicit in a numerical solution of [1] is the numerical solution of the equations of motion for any matter fields that couple to the gravitational field – that is, that contribute to $T_{\mu\nu}$. The reader is reminded that it is a hallmark of general relativity that, in principle, all matter fields – including massless ones such as the electromagnetic field – contribute to $T_{\mu\nu}$.

Now, in the $3+1$ approach to general relativity that is described below, the task of solving the field equations [1] is formulated as an initial-value or Cauchy problem. Specifically, the spacetime metric, $g_{\mu\nu}(x^\alpha) = g_{\mu\nu}(t, x^k)$, which encodes all geometric information concerning the spacetime, \mathcal{M} , is viewed as the time history, or dynamical evolution, of the spatial metric, $\gamma_{ij}(0, x^k)$, of an initial space-like hypersurface, $\Sigma(0)$. In any practical calculation, the degree to which the matter fields “back-react” on the gravitational field, that is, contribute to $T_{\mu\nu}$ substantially enough to cause perturbations in $g_{\mu\nu}$ at or above the desired accuracy threshold, will thus depend on the specifics of the initial configuration.

In astrophysics, there are relatively few well-identified environments in which it is generally thought to be crucial to the faithful emulation of the physics that the matter fields be fully coupled to the gravitational field. However, both observationally and theoretically, the existence of gravitationally compact objects is quite clear. Gravitationally compact means that a star with mass, M , has a radius, R , comparable to its Schwarzschild radius, R_M , which is defined by

$$R_M = \frac{2G}{c^2} M \approx 10^{-27} \text{ kg m}^{-1} \quad [2]$$

Here, and only here, G and c – Newton's gravitational constant and the speed of light, respectively – have been explicitly reintroduced. The fact that R_M/R is about 10^{-6} and 10^{-9} at the surfaces of the sun and earth, respectively, is a reminder of just how

weak gravity is in the locality of Earth. However, as befits anything of Einsteinian nature, the weakness of gravity is relative, so that at the surface of a neutron star, one would find

$$\frac{R_M}{R} \sim 0.4 \quad [3]$$

while for black holes, one has

$$\frac{R_M}{R} = 1 \quad [4]$$

In such circumstances, gravity is anything but weak! Furthermore, in situations where the matter–energy distribution has a highly time-dependent quadrupole moment – such as occurs naturally with a compact-binary system (i.e., a gravitationally bound two-body system, in which each of the bodies is either a black hole or a neutron star) – the dynamics of the gravitational field, including, crucially, the dynamics of the radiative components of the gravitational field, can be expected to dominate the dynamics of the overall system, matter included. For scenarios such as these, it should come as no surprise that the solution of the combined gravitohydrodynamical system begs for numerical analysis.

In addition, both from the physical and mathematical perspectives, it is also natural to study the strong, field dynamic regimes ($R \rightarrow R_M$ and/or $v \rightarrow c$, where v is the typical speed characterizing internal bulk motion of the matter) of general relativity within the context of a variety of matter models. Typical processes addressed by these theoretical studies include the process of black hole formation, end-of-life events for various types of model stars, and, again, the interaction, including collisions, of gravitationally compact objects. Note that it is another hallmark of general relativity that highly dynamical spacetimes need not contain any matter; indeed, the interaction of two black holes – the natural analog of the Kepler problem in relativity – is a vacuum problem; that is, it is described by a solution of [1] with $T_{\mu\nu} = 0$.

Motivated in significant part by the large-scale efforts currently underway to directly detect gravitational radiation (gravitational waves), much of the contemporary work in numerical relativity is focused on precisely the problem of the late phases of compact-binary inspiral and merger. Such binaries are expected to be the most likely candidates for early detection by existing instruments such as TAMA, GEO, VIRGO, LIGO, and, more likely, by planned detectors including LIGO II and LISA (see, e.g., [Hough and Rowan \(2000\)](#)). Detailed and accurate predictions of expected waveforms from

these events – using the techniques of numerical relativity – have the potential to substantially hasten the discovery process, on the basis of the general principle that if one knows what signal to look for, it is much easier to extract that signal from the experimental noise.

The computational task facing numerical relativists who study problems such as binary inspiral is formidable. In particular, such problems are intrinsically “3D,” to use the CFD (computational fluid dynamics) nomenclature in which time dependence is always assumed. That is, the PDEs that must be solved govern functions, $F(t, x^k)$, that depend on all three spatial coordinates, x^k , as well as on time, t . Unfortunately, even a cursory description of 3D work in numerical relativity as it stands at this time is far beyond the scope of this article.

What follows, then, is an outline of a traditional approach to numerical relativity that underpins many of the calculations from the early years of the field (1970s and 1980s), most of which were carried out with simplifying restrictions to either spherical symmetry or axisymmetry. The mathematical development, which will hereafter be called the 3 + 1 approach to general relativity, has the advantage of using tensors and an associated tensor calculus that are reasonably intuitive for the physicist. This “standard” 3 + 1 approach is also sufficient in many instances (particularly those with symmetry) in the sense that it leads to well-posed sets of PDEs that can be discretized and then solved computationally in a convergent (stable) fashion. In addition, a thorough understanding of the 3 + 1 approach will be of significant help to the reader wishing to study any of the current literature in numerical relativity, including the 3D work.

However, the reader is strongly cautioned that the blind application of any of the equations that follow, especially in a 3D context, may well lead to “ill-posed systems,” numerical analysis of which is useless. Anyone specifically interested in using the methods of numerical relativity to generate discrete, approximate solutions to [1], particularly in the generic 3D case, is thus urged to first consult one of the comprehensive reviews of numerical relativity that continue to appear at fairly regular intervals (see, e.g., [Lehner \(2001\)](#), or [Baumgarte and Shapiro \(2003\)](#)). Most such references will also provide a useful overview of many of the most popular numerical techniques that are currently being used to discretize (convert to algebraic form) the Einstein equations, as well as the main algorithms that are used to solve the resulting discrete equations. These subjects are not

described below, not least since discussion of the available discretization techniques only makes sense in the context of PDEs of specific systems with specific boundary conditions, while there is only space here to describe the general mathematical setting for 3 + 1 numerical relativity.

The 3 + 1 Spacetime Split

At least at the current time, computations in numerical relativity are restricted to the case of globally hyperbolic spacetimes. A spacetime (four-dimensional pseudo-Riemannian manifold), \mathcal{M}_Σ , endowed with a metric, $g_{\mu\nu}$, is globally hyperbolic if there is at least one edgeless, spacelike hypersurface, $\Sigma(0)$, that serves as a Cauchy surface. That is, provided that the initial data for the gravitational field are set consistently on $\Sigma(0)$ – so that the four constraint equations are satisfied (see below) – the entire metric $g_{\mu\nu}(t, x^i)$ can be determined from the field equations [1] (with appropriate boundary conditions), and thus, so can the complete geometric structure of the spacetime manifold.

To be sure, global hyperbolicity is restrictive. It excludes, for example, the highly interesting Gödel universe. However, particularly from the point of view of studying asymptotically flat solutions (or solutions asymptotic to any of the currently popular cosmologies), as is usually the case in astrophysics, the requirement of global hyperbolicity is natural.

The 3 + 1 split is based on the complete foliation of \mathcal{M}_Σ based on level surfaces of a scalar function, t – the time function. That is, the $t = \text{const.}$ slices, are three-dimensional spacelike (Riemannian) hypersurfaces, and, as t ranges from $-\infty$ to $+\infty$, completely fill the spacetime manifold, \mathcal{M}_Σ . In order for the $\Sigma(t)$ to be everywhere spacelike, t must be everywhere timelike:

$$g_{\mu\nu} \nabla^\mu t \nabla^\nu t < 0 \quad [5]$$

Here ∇_μ is the spacetime covariant derivative operator compatible with the four metric, $g_{\mu\nu}$, thus satisfying $\nabla_\alpha g_{\mu\nu} = 0$, and $g^{\mu\nu}$ is the inverse metric tensor, which satisfies $g^{\mu\alpha} g_{\alpha\nu} = \delta^\mu_\nu$. The reader is reminded that δ^μ_ν is a Kronecker delta symbol; that is, δ^μ_ν has the value 1 if $\mu = \nu$, and the value 0 otherwise.

Furthermore, the scalar function t is now adopted as the temporal coordinate, so that $x^\mu = (t, x^i)$, where the x^i are the three spatial coordinates. As noted implicitly above, since the problem under consideration is a pure Cauchy evolution, the range

of t should nominally be infinite, both to the future as well as to the past; that is, the solution domain is

$$-\infty < t < \infty \quad [6]$$

$$|X| \equiv (\gamma_{ij} x^i x^j)^{1/2} < \infty \quad [7]$$

However, this assumes that one has global existence for arbitrarily strong initial data, which is decidedly not always the case in general relativity. Indeed, “continued” or “catastrophic” gravitational collapse – that is, the process of black hole formation – signaled, in modern language, by the appearance of a trapped surface, inexorably leads to a physical singularity, which – the somewhat vague nature of the singularity theorems of Penrose, Hawking, and others notwithstanding – in actual numerical computations invariably turns out to be “catastrophic” in terms of Cauchy evolution.

Such behavior in time-dependent nonlinear PDEs is quite familiar in the mathematical community at large, where it is frequently known as finite-time blow-up (or finite-time singularity). However, despite the fact that such behavior is one of the most fascinating aspects of solutions of the Einstein equations, the following discussion will be, implicitly at least, restricted to the case of weak initial data, that is, to initial data for which there is global existence.

With the manifold \mathcal{M}_Σ sliced into an infinite stack of spacelike hypersurfaces, $\Sigma(t)$, attention shifts to any single surface, as well as to the manner in which such a generic surface is embedded in the spacetime.

First, each spacelike hypersurface, $\Sigma(t)$, is itself a three-dimensional Riemannian differential manifold with a metric $\gamma_{ij}(t, x^k)$. (Note that in this discussion, the symbol t is to be understood to represent any specific value of coordinate time.) From this metric, one can construct an inverse metric, $\gamma^{ij}(t, x^k)$, defined, as usual, so that

$$\gamma^{ik} \gamma_{kj} = \delta^i_j \quad [8]$$

Associated with the spatial metric, γ_{ij} , is a natural spatial covariant derivative operator, D_i , that is compatible with γ_{ij} :

$$D_k \gamma_{ij} = 0 \quad [9]$$

With the spatial metric, γ_{ij} , and its inverse, γ^{ij} , in hand, the standard formulas of tensor analysis can be applied to compute the usual suite of geometrical tensors. All tensors thus computed, and indeed, all tensors defined intrinsically to the

hypersurfaces $\Sigma(t)$ are called “spatial” tensors, and have their indices (if any) raised and lowered with γ^{ij} and γ_{ij} , respectively.

Thus, the Christoffel symbols of the second kind, Γ^i_{jk} , are given by

$$\Gamma^i_{jk} = \frac{1}{2} \gamma^{il} (\partial_k \gamma_{lj} + \partial_j \gamma_{lk} - \partial_l \gamma_{jk}) \quad [10]$$

Note that these quantities are symmetric in their last two indices

$$\Gamma^i_{jk} = \Gamma^i_{kj} \quad [11]$$

and that they can be used, as usual, in explicit calculation of the action of the spatial covariant derivative operator on an arbitrary tensor. In particular, for the special cases of a spatial vector, V^i , and a covector (1-form), W_i , one has

$$D_i V^j = \partial_i V^j + \Gamma^j_{ik} V^k \quad [12]$$

and

$$D_i W_j = \partial_i W_j - \Gamma^k_{ij} W_k \quad [13]$$

respectively.

Given the Christoffel symbols, the components of the spatial Riemann tensor, denoted here $\mathcal{R}_{ijk}{}^l$, are computed using

$$\begin{aligned} \mathcal{R}_{ijk}{}^l = & \partial_j \Gamma^l_{ik} - \partial_i \Gamma^l_{jk} + \Gamma^m_{ik} \Gamma^l_{mj} \\ & - \Gamma^m_{jk} \Gamma^l_{mi} \end{aligned} \quad [14]$$

Finally, the Ricci tensor, \mathcal{R}^i_j , and Ricci scalar, \mathcal{R} , are defined in the usual fashion

$$\mathcal{R}^i_j = \gamma^{ik} \mathcal{R}_{kj} = \gamma^{ik} \mathcal{R}_{klj}{}^l \quad [15]$$

$$\mathcal{R} = \gamma^{ij} \mathcal{R}_{ij} \quad [16]$$

The reader should again note that all of the tensors just defined “live” on each and every single spacelike hypersurface, $\Sigma(t)$, and are thus known as hypersurface-intrinsic quantities. In particular, the spatial Riemann tensor, $\mathcal{R}_{ijk}{}^l$, which encodes all intrinsic geometric information about $\Sigma(t)$, in no way depends on how the slice is embedded in the spacetime \mathcal{M}_Σ .

The next step in the 3 + 1 approach involves rewriting the fundamental spacetime line element for the squared proper distance, ds^2 , between two spacetime events, \mathcal{P} and \mathcal{Q} , having coordinates x^μ and $x^\mu + dx^\mu$, respectively,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad [17]$$

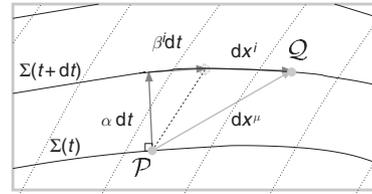


Figure 1 Spacetime displacement in the 3 + 1 approach, following Misner, Thorne, and Wheeler (1973). Solid lines represent surfaces of constant time, t ; that is, each solid line represents a single spacelike hypersurface, $\Sigma(t)$. Dotted lines denote trajectories of constant spatial coordinate, that is, trajectories with $x^k = \text{const}$. The lapse function, $\alpha(t, x^k)$, encodes the (local) ratio between elapsed coordinate time, dt , and elapsed proper time, $d\tau = \alpha dt$, for an observer moving normal to the slices (i.e., for an observer with a 4-velocity, u^μ , identical to the hypersurface normal, n^μ). Similarly, the shift vector, $\beta^i(t, x^k)$, describes the shift, $\beta^i(t, x^i) dt$, in trajectories of constant spatial coordinate – the dotted lines in the figure – relative to motion perpendicular to the slices. The 3 + 1 form of the line element [18] then follows immediately from an application of the spacetime version of the Pythagorean theorem.

As **Figure 1** illustrates, a quick route to the 3 + 1 decomposition of the above expression, and thus of the tensor $g_{\mu\nu}$ itself, is based on an application of the “four-dimensional Pythagorean theorem.” In setting up the calculation, one naturally identifies four functions, the scalar lapse, $\alpha(t, x^k)$, and the vector shift, $\beta^i(t, x^k)$, that encode the full coordinate (gauge) freedom of the theory. That is, complete specification of the lapse and shift is equivalent to completely fixing the spacetime coordinate system.

In light of the above discussion, and again referring to **Figure 1**, one readily deduces the 3 + 1 decomposition of the spacetime line element:

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij} (dx^i + \beta^i dt) (dx^j + \beta^j dt) \quad [18]$$

A rearranged form of this last expression is also often seen in the literature:

$$\begin{aligned} ds^2 = & (-\alpha^2 + \beta_k \beta^k) dt^2 + 2\beta_k dx^k dt \\ & + \gamma_{ij} dx^i dx^j \end{aligned} \quad [19]$$

The following useful identifications of the “time–time,” “time–space,” and “space–space” pieces of the spacetime metric, $g_{\mu\nu}$, follow immediately from [19]:

$$g_{00} = -\alpha^2 + \beta^i \beta_i \quad [20]$$

$$g_{0i} = g_{i0} = \beta_i = \gamma_{ik} \beta^k \quad [21]$$

$$g_{ij} = \gamma_{ij} \quad [22]$$

This last relation is an example of a useful general result; the purely spatial components, $Q_{ijk\dots}$, of a

completely covariant, but otherwise arbitrary, spacetime tensor, $Q_{\alpha\beta\gamma\dots}$, constitute the components of a completely covariant spatial tensor.

A straightforward calculation, which provides a good exercise in the use of the 3 + 1 calculus, yields the following equally useful identifications for various pieces of the inverse spacetime metric: $g^{\alpha\beta}$

$$g^{00} = -\alpha^{-2} \quad [23]$$

$$g^{0i} = g^{i0} = \alpha^{-2}\beta^i \quad [24]$$

$$g^{ij} = \gamma^{ij} - \alpha^{-2}\beta^i\beta^j \quad [25]$$

Since the Einstein field equations are equations with, loosely speaking, geometry on one side and matter on the other, tensors built from matter fields must also be decomposed. In particular, it is conventional to define tensors, ρ , j_i , and S_{ij} that result from various projections of the spacetime stress energy tensor, $T_{\mu\nu}$, onto the hypersurface:

$$\rho \equiv n_\mu n_\nu T^{\mu\nu} \quad [26]$$

$$j_i \equiv -n_\mu T^\mu{}_i \quad [27]$$

$$S_{ij} \equiv T_{ij} \quad [28]$$

For observers with 4-velocities u^μ equal to n^μ , and only for those observers with $u^\mu = n^\mu$, the above quantities have the interpretation of the locally and instantaneously measured energy density, momentum density, and spatial stresses, respectively. As with the geometric quantities, all of the matter variables, ρ , j_i , and S_{ij} defined in [26]–[28] are spatial tensors and thus have their indices (if any) raised and lowered with the 3-metric. Note that the identification $S_{ij} = T_{ij}$ is another illustration of the general result mentioned in the context of the previous identification of γ_{ij} and g_{ij} .

Finally, observing that time parameters are naturally defined in terms of level surfaces (equipotential surfaces), it should be no surprise that the covariant components, n_μ , of the hypersurface normal field,

$$n_\mu = (-\alpha, 0, 0, 0) \quad [29]$$

are simpler than the components, n^μ , of the normal itself,

$$n^\mu = (\alpha^{-1}, \alpha^{-1}\beta^i) \quad [30]$$

and, in fact, eqn [29] can also be deduced from a quick study of [Figure 1](#).

In the 3 + 1 approach, in addition to the 3-metric, $\gamma_{ij}(t, x^k)$, and coordinate functions, $\alpha(t, x^i)$ and $\beta(t, x^i)$, it is convenient to introduce an additional rank-2 symmetric spatial tensor, $K_{ij}(t, x^k)$, known as

the extrinsic curvature (or second fundamental form). This additional tensor is analogous to a time derivative of $\gamma_{ij}(t, x^k)$, or, from a Hamiltonian perspective, to a variable that is dynamically conjugate to $\gamma_{ij}(t, x^k)$.

As the name suggests, the extrinsic curvature describes the manner in which the slice $\Sigma(t)$ is embedded in the manifold (to be contrasted with $\mathcal{R}_{ijk}{}^l$ defined by [14] which is, as mentioned previously, completely insensitive to the manner in which the hypersurface is embedded in \mathcal{M}_Σ).

Geometrically, K_{ij} is computed by calculating the spacetime gradient of the normal covector field, n_μ , and projecting the result on to the hypersurface,

$$K_{ij} = -\frac{1}{2}\nabla_i n_j \quad [31]$$

where it must be stressed that ∇_μ is the spacetime covariant derivative operator compatible with the 4-metric, $g_{\alpha\beta}$; that is, $\nabla_\mu g_{\alpha\beta} = 0$. A straightforward tensor calculus calculation then yields the following, which can be viewed as a definition of the K_{ij} :

$$K_{ij} = \frac{1}{2\alpha} (\partial_t \gamma_{ij} + D_i \beta_j + D_j \beta_i) \quad [32]$$

Here, D_i is the spatial covariant metric, compatible with γ_{ij} ($D_k \gamma_{ij} = 0$), that was defined previously. Observe that this equation can be easily solved for $\partial_t \gamma_{ij}$ (this will be done below), and thus, in the 3 + 1 approach it is [32] that is the origin of the evolution equations for the 3-metric components, γ_{ij} .

Einstein's Equations in 3 + 1 Form

The Constraint Equations

As is well known, as a result of the coordinate (gauge) invariance of the theory, general relativity is overdetermined in a sense completely analogous to the situation in electrodynamics with the Maxwell equations. One of the ways that this situation is manifested is via the existence of the constraint equations of general relativity. Briefly, starting from the naive view that the ten metric functions, $g_{\mu\nu}(t, x^i)$, that completely determine the spacetime geometry are all dynamical – that is, that they satisfy second-order-in-time equations of motion – one finds that the Einstein equations do not provide dynamical equations of motion for the lapse, α , or the shift, β^i . Rather, four of the field equations [1] are equations of constraint for the “true” dynamical variables of the theory, $\{\gamma_{ij}, \partial_t \gamma_{ij}\}$, or, equivalently, $\{\gamma_{ij}, K^i{}_j\}$. Note that in the following, the mixed form, $K^i{}_j$, is at times used – again by convention – as the principal representation of the extrinsic curvature tensor (instead of K_{ij} as previously, or K^{ij}).

Thus, four of the components of [1] can be written in the form

$$C^\mu(\gamma_{ij}, K^i_j, \partial_k \gamma_{ij}, \partial_l \partial_k \gamma_{ij}, \partial_k K^i_j) = T^\mu \quad [33]$$

where T^μ depends only on the matter content in the spacetime. Note that in addition to having no dependence on $\partial_t^2 \gamma_{ij}$, the constraints are also independent of α and β^i .

If the Einstein equations [1] are to hold throughout the spacetime, then the constraints [33] must hold on each and every spacelike hypersurface, $\Sigma(t)$, including, crucially, the initial hypersurface, $\Sigma(0)$. From the point of view of Cauchy evolution, this means that the 12 functions, $\{\gamma_{ij}(0, \mathbf{x}^k), K^i_j(0, \mathbf{x}^k)\}$, constituting the gravitational part of the initial data, are not completely freely specifiable, but must satisfy the four constraints

$$C^\mu(\gamma_{ij}(0, \mathbf{x}^k), K^i_j(0, \mathbf{x}^k), \dots) = T^\mu(0, \mathbf{x}^k) \quad [34]$$

However, provided initial data that do satisfy the equations is chosen, then – as consistency of the theory demands – the dynamical equations of motion for the $\{\gamma_{ij}, K^i_j\}$ (eqns [37] and [38] below) guarantee that the constraints will be satisfied on all future (or past) hypersurfaces, $\Sigma(t)$. In this internal self-consistency, the geometrical Bianchi identities, $\nabla_\mu G^{\mu\nu} = 0$, and the local conservation of stress energy, $\nabla_\mu T^{\mu\nu} = 0$, play crucial roles.

In the 3 + 1 approach, as one would expect, the constraint equations further naturally subdivide into a scalar equation

$$\mathcal{R} - K_{ij}K^{ij} + K^2 = 16\pi\rho \quad [35]$$

and a (spatial) vector equation

$$D_j K^{ij} - D^i K = 8\pi j^i \quad [36]$$

where the energy and momentum densities, ρ and $j^i = \gamma^{ik} j_k$, are given by [26]–[28]. Equations [35] and [36] are often known as the Hamiltonian and momentum constraint, respectively, not least since the behavior of their solutions as $X \equiv \sqrt{\gamma_{ij} x^i x^j} \rightarrow \infty$ encodes the conserved mass and linear momentum (four numbers) that can be defined in asymptotically flat spacetimes.

In a general 3 + 1 coordinate system, and with an appropriate choice of variables, the constraints can be written as a set of quasilinear elliptic equations for four of the $\{\gamma_{ij}, K^i_j\}$ (or, more properly, for certain algebraic combinations of the $\{\gamma_{ij}, K^i_j\}$). Thus, especially for 2D and 3D calculations, the setting of initial data for the Cauchy problem in general relativity is itself a highly nontrivial mathematical and computational exercise. Readers wishing more details on this subject are directed to the comprehensive review by Cook (2000).

The Evolution Equations

As discussed above, in the 3 + 1 form of the Einstein equations [1], the spatial metric, γ_{ij} , and the extrinsic curvature, K^i_j , are viewed as the dynamical variables for the gravitational field. The remainder of the 3 + 1 equations are thus two sets of six first-order-in-time evolution equations; one set for γ_{ij} ,

$$\begin{aligned} \partial_t \gamma_{ij} = & -2\alpha \gamma_{ik} K^k_j + \beta^k \partial_k \gamma_{ij} \\ & + \gamma_{ik} \partial_j \beta^k + \gamma_{kj} \partial_i \beta^k \end{aligned} \quad [37]$$

and the other set for K^i_j ,

$$\begin{aligned} \partial_t K^i_j = & \beta^k \partial_k K^i_j - \partial_k \beta^i K^k_j + \partial_j \beta^k K^i_k - D^i D_j \alpha \\ & + \alpha (\mathcal{R}^i_j + K K^i_j + 8\pi (\frac{1}{2} \delta^i_j (S - \rho) - S^i_j)) \end{aligned} \quad [38]$$

As also noted previously, the evolution equations [37] for the spatial metric components, γ_{ij} , follow from the definition of the extrinsic curvature [31]. The derivation of the equations for the extrinsic curvature, on the other hand, require lengthy, but well-documented, manipulations of the spatial components of the field equations [1].

The (Naive) Cauchy Problem

A naive statement of the Cauchy problem for 3 + 1 numerical relativity is thus as follows: fix a specified number, N , of matter fields $\xi^A(t, \mathbf{x}^k)$, $A = 1, 2, \dots, N$, all minimally coupled to the gravitational field, with a total stress tensor, $T_{\mu\nu}$, given by

$$T_{\mu\nu} = \sum_{A=1}^N T_{\mu\nu}^A \quad [39]$$

where $T_{\mu\nu}^A$ is the stress tensor corresponding to the matter field ξ^A . Choose a topology for $\Sigma(0)$ (e.g., \mathcal{R}^3 with asymptotically flat boundary conditions; T^3 with no boundaries, etc.) This also fixes the topology of \mathcal{M}_Σ to be $\mathbb{R} \times$ the topology of $\Sigma(0)$.

Next, freely specify eight of the 12 $\{\gamma_{ij}(0, \mathbf{x}^k), K^i_j(0, \mathbf{x}^k)\}$, as well as initial values, $\xi^A(0, \mathbf{x}^k)$, for the matter fields. Then determine the remaining four dynamical gravitational fields from the constraints [35] and [36]. This completes the initial data specification.

One must now choose a prescription for the kinematical (coordinate) functions, α and β^i , so that either explicitly or implicitly, they are completely fixed; for the case of implicit specification, this may well mean that the coordinate functions themselves will satisfy PDEs, which, furthermore, can be of essentially any type in practice (i.e., elliptic, hyperbolic, parabolic, ...). Finally, with consistent initial data, $\{\gamma_{ij}(0, \mathbf{x}^k), K^i_j(0, \mathbf{x}^k); \xi_A(0, \mathbf{x}^k)\}$, in hand, and with a prescription for the coordinate functions, the evolution

equations [37] and [38] can be used to advance the dynamical variables forward or backward in time.

The above description is naive since, apart from a consistent mathematical specification, the most crucial issue in the solution of a time-dependent PDE as a Cauchy problem is that the problem be “well posed.” Roughly speaking, this means that solutions do not grow without bound (“blow-up”) without physical cause, and that small, smooth changes to initial data yield correspondingly small, smooth changes to the evolved data. In short, the Cauchy problem must be stable, and whether or not a particular subset of the equations displayed in this section yields a well-posed problem is a complicated and delicate issue, especially in the generic 3D case. The reader is thus again cautioned against blind application of any of the equations displayed in this article.

Boundary Conditions

In principle, because all spacelike hypersurfaces, $\Sigma(t)$, in a pure Cauchy evolution are edgeless – and provided that the initial data $\{\gamma_{ij}(0, x^k), K^i_j(0, x^k); \xi_A(0, x^k)\}$ is consistent with asymptotic flatness, or whatever other condition is appropriate given the topology of the $\Sigma(t)$ – there are essentially no boundary conditions to be imposed on the dynamical variables, $\{\gamma_{ij}(t, x^k), K^i_j(t, x^k)\}$, during Cauchy evolution. Note that asymptotic flatness generally requires that

$$\lim_{X \rightarrow \infty} \gamma_{ij} = f_{ij} + O\left(\frac{1}{X}\right) \quad [40]$$

and

$$\lim_{X \rightarrow \infty} K^i_j = O\left(\frac{1}{X^2}\right) \quad [41]$$

where X is defined by

$$X \equiv \sqrt{\gamma_{ij} x^i x^j} \quad [42]$$

as previously, and f_{ij} is the flat 3-metric. Similarly, should the lapse, α , and shift, β , be constrained by elliptic PDEs – as is frequently the case in practice – then the only natural place to set boundary conditions is at spatial infinity, and then, provided that the frame at spatial infinity is inertial, with coordinate time t measuring proper time, one should have

$$\lim_{x \rightarrow \infty} \alpha = 1 + O\left(\frac{1}{X}\right) \quad [43]$$

and

$$\lim_{X \rightarrow \infty} \beta^i = O\left(\frac{1}{X}\right) \quad [44]$$

It is critical to note at this point, however, that in the vast bulk of past and current work in numerical relativity, including most of the ongoing work in 3D, the Einstein equations [1] have been solved, not as a pure Cauchy problem, but as a mixed initial-value/boundary-value (IBVP) problem. That is, in the discretization process in which the continuum equations [1] are replaced with algebraic equations, the continuum domain [6]–[7] is typically replaced with a truncated spatial domain

$$|x^i| \leq X^i_{\max} \quad [45]$$

where the X^i_{\max} are *a priori* specified constants (parameters of the computational solution) that define the extremities of the “computational box.” As one might expect, the theory underlying stability and well-posedness of IBVP problems – especially for differential systems as complicated as [1] – is even more involved than for the pure initial-value case, and is another very active area of research in both mathematical and numerical relativity (see, e.g., Friedrich and Nagy (1999)).

See also: Critical Phenomena in Gravitational Collapse; Einstein Equations: Initial Value Formulation; Fluid Mechanics: Numerical Methods; General Relativity: Overview; Geometric Analysis and General Relativity; Gravitational Waves; Hamiltonian Reduction of Einstein’s Equations; Magnetohydrodynamics; Spacetime Topology, Causal Structure and Singularities; Symmetric Hyperbolic Systems and Shock Waves.

Further Reading

- Baumgarte T and Shapiro SL (2001) Numerical relativity and compact binaries. *Physics Reports* 376: 41–131.
- Cook G (2000) Initial data for numerical relativity. *Living Reviews of Relativity* 3: 5 (irr-2000-5).
- Font JA (2003) Numerical hydrodynamics in general relativity. *Living Reviews of Relativity* 6: 4 (irr-2003-4).
- Frauenfelder J (2004) Conformal infinity. *Living Reviews of Relativity* 7: 1 (irr-2004-1).
- Friedrich H and Nagy G (1999) The initial boundary value problem for Einstein’s vacuum field equation. *Communications in Mathematical Physics* 201: 619–655.
- Hough J and Rowan S (2000) Gravitational wave detection by interferometry (ground and space). *Living Reviews of Relativity* 3: 3 (irr-2000-3).
- Lehner L (2001) Numerical relativity: a review. *Classical and Quantum Gravity* 18: R25–R86.
- Misner CW, Thorne KS, and Wheeler JA (1973) *Gravitation*. San Francisco: W.H. Freeman.
- Reula OA (1998) Hyperbolic methods for Einstein’s equations. *Living Reviews of Relativity* 1: 3 (irr-1998-3).
- Winicour J (2001) Characteristic evolution and matching. *Living Reviews of Relativity* 4: 3 (irr-2001-3).

Confinement see Quantum Chromodynamics

Conformal Geometry see Two-dimensional Conformal Field Theory and Vertex Operator Algebras

Conservation Laws see Symmetries and Conservation Laws

Constrained Systems

M Henneaux, Université Libre de Bruxelles, Brussels, Belgium

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Consider a dynamical system with coordinates q^i ($i = 1, \dots, n$) and Lagrangian $L(q^i, \dot{q}^i)$ (field theory is formally covered by regarding the spatial coordinates as a continuous index). When going to the Hamiltonian formulation, it is usually assumed that the Legendre transformation between the velocities \dot{q}^i and the momenta

$$p_i = \frac{\partial L}{\partial \dot{q}^i} \quad [1]$$

can be inverted to yield the velocities as functions of the q 's and the p 's. This “regular” situation occurs for most systems appearing in standard classical mechanics and enables one to proceed to the Hamiltonian formulation of the theory without difficulty.

In field theory, however, the regular case is the exception rather than the rule. This is due to gauge invariance and first-order Lagrangians.

- *Gauge invariance* A system possesses gauge symmetries if it is invariant under transformations that involve arbitrary functions of time (gauge transformations). In that case, the solution of the equations of motion with given initial data is not unique, since it is always possible to perform a gauge transformation in the course of the evolution without changing the initial data. It is then clear that the Legendre transformation cannot be invertible, for if it were, one could rewrite the equations

of motion in the standard canonical form $\dot{q}^i = \partial H / \partial p_i, \dot{p}_i = -\partial H / \partial q^i$. These canonical equations are in normal form and have a unique solution for given initial data, which would contradict the presence of a gauge symmetry.

A simple example that illustrates this phenomenon is given by the following model for three variables q^1, q^2 , and λ , the Lagrangian of which reads

$$L = \frac{1}{2} \left((\dot{q}^1 - \lambda)^2 + (\dot{q}^2 - \lambda)^2 \right) \quad [2]$$

This model is inspired by electromagnetism: the variables q^1 and q^2 play a role somewhat similar to that of the spatial components of the vector potential, while λ corresponds to the temporal component. The Lagrangian is invariant under the gauge transformations

$$q^1 \rightarrow q^1 + \varepsilon, \quad q^2 \rightarrow q^2 + \varepsilon, \quad \lambda \rightarrow \lambda + \dot{\varepsilon} \quad [3]$$

where ε is an arbitrary function of time. The conjugate momenta are

$$p_1 = \dot{q}^1 - \lambda, \quad p_2 = \dot{q}^2 - \lambda, \quad \pi_\lambda = 0$$

One cannot invert the Legendre transformation since one cannot express the velocity $\dot{\lambda}$ in terms of the momenta.

- *First-order Lagrangians* Fermionic fields obey first-order equations. Their Lagrangian is linear in the derivatives, so that the conjugate momenta p_i depend on the coordinates q^i only. It is then clearly impossible to express the velocities in terms of the momenta through the Legendre transformation. More generally, any first-order Lagrangian with or without gauge symmetry leads to a noninvertible Legendre transformation.

A simple system that exhibits this feature is described by the Lagrangian

$$L = z^2 \dot{z}^1 - \frac{1}{2}(z^2)^2 \quad [4]$$

for two bosonic degrees of freedom (z^1, z^2). This is in fact the canonical form of the Lagrangian for a free particle in one dimension (z^2 is the momentum conjugate to the position z^1): the system is already in Hamiltonian form. There is no gauge invariance, but because the Lagrangian is first order, the Legendre transformation with [4] as starting point,

$$p_1 = z^2, \quad p_2 = 0 \quad [5]$$

is non invertible for the velocities (which do not even appear in the formulas for the momenta).

Dirac showed how to develop the Hamiltonian formalism in the case when the Legendre transformation is not invertible. One can still reformulate the equations in phase space and write them in terms of brackets with the Hamiltonian, but a new major feature emerges, namely the canonical variables are no longer free. Rather, the permissible phase-space points are constrained to be on the so-called “constrained surface.” For this reason, systems for which the Legendre transformation is not invertible are also called “constrained Hamiltonian systems.” We shall adopt this terminology here.

The purpose of this article is to explain the main ideas underlying the Dirac method. To simplify the discussions and to focus on the features peculiar to the Dirac construction, we shall assume as a rule that all necessary smoothness conditions are fulfilled by the functions, surfaces, etc., appearing in the formalism. How to develop the analysis when some of the smoothness conditions are not fulfilled is of definite interest but goes beyond the scope of this review. We shall also assume, for definiteness, that all the variables are bosonic in order to avoid straightforward but somewhat cumbersome sign factors in the formulas.

General Theory

Dirac Algorithm

Primary constraints When the Legendre transformation [1] cannot be inverted, the momenta p_i 's do not span an n -dimensional space but are constrained by relations

$$\phi_m(q, p) = 0, \quad m = 1, \dots, M \quad [6]$$

which follow from their definition. These equations reduce to identities when the momenta are replaced

by their expression [1] in terms of the coordinates and the velocities. They are called primary constraints. We shall assume that the matrix

$$\frac{\partial(\phi_m)}{\partial(p_i, q^i)}$$

is everywhere of constant (maximum) rank M on the phase-space surface defined by eqns [6] which is assumed to be smooth. This surface is of dimension $2n - M$.

Canonical Hamiltonian The next step in the Dirac procedure is to define the canonical Hamiltonian H through

$$H = \dot{q}^i p_i - L \quad [7]$$

As shown by Dirac, H can be re-expressed as a function $H(q, p)$ of the momenta and the coordinates, even when the Legendre transformation is not invertible: the canonical Hamiltonian H depends on the velocities only through the p_i 's. Furthermore, the original equations of motion in Lagrangian form are equivalent to the Hamiltonian equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + u^m \frac{\partial \phi_m}{\partial p_i} \quad [8]$$

$$\dot{p}_i = -\frac{\partial H}{\partial q^i} - u^m \frac{\partial \phi_m}{\partial q^i} \quad [9]$$

$$\phi_m(q, p) = 0 \quad [10]$$

where the u^m 's are parameters, some of which will be determined through the consistency algorithm to be discussed shortly. (In [7]–[9] and everywhere below, there is a summation over the repeated indices.)

Secondary constraints The equations of motion [8] and [9] can be rewritten as

$$\dot{F} = [F, H] + u^m [F, \phi_m] \quad [11]$$

where $F = F(q, p)$ is any function of the canonical variables. Here, the Poisson bracket is defined as usual by

$$[G, F] = \frac{\partial G}{\partial q^i} \frac{\partial F}{\partial p_i} - \frac{\partial G}{\partial p_i} \frac{\partial F}{\partial q^i} \quad [12]$$

If one takes for F one of the primary constraints ϕ_m , one should get zero, $\dot{\phi}_m = 0$. This yields the consistency conditions

$$[\phi_m, H] + u^{m'} [\phi_m, \phi_{m'}] = 0 \quad [13]$$

These conditions can imply further restrictions on the canonical variables and/or impose conditions on the

variables u^m . Any new relation $X(q, p) = 0$ on the canonical variables leads, in turn, to a further consistency condition $\dot{X} = [X, H] + u^{m'} [X, \phi_{m'}] = 0$, which can bring in either further restriction on the constraint surface or fix more variables u^m . Constraints that follow from the consistency algorithm are called “secondary constraints.” Finally, one is left with a certain number of secondary constraints, which are denoted by $\phi_k = 0$, $k = M + 1, \dots, M + K$. We assume again that all the constraints (primary and secondary) define a smooth surface, called the “constraint surface,” and fulfill the condition that $\partial(\phi_k)/\partial(q^i, p_i)$ is of maximum rank $J \equiv M + K$ on the constraint surface. (We also assume for simplicity that there is no branching in the consistency algorithm.)

Restrictions on the u 's Having a complete set of constraints

$$\phi_j = 0, \quad j = 1, \dots, M + K \equiv J \quad [14]$$

we can now investigate more precisely the restrictions on the variables u^m . These read

$$[\phi_j, H] + u^m [\phi_j, \phi_m] \approx 0, \quad j = 1, \dots, J \quad [15]$$

where the notation \approx means “equal modulo the constraints.” In [15], m is summed from 1 to M . Equations [15] are a set of J linear, inhomogeneous equations for the u 's, with coefficients that are functions of the canonical variables q^i, p_i . The general solution of this system is of the form

$$u^m = U^m + u^a V_a^m \quad [16]$$

where U^m is a particular solution and where the V_a^m ($a = 1, \dots, A$) provide a complete set of independent solutions of the homogeneous system

$$V_a^m [\phi_j, \phi_m] \approx 0 \quad [17]$$

The coefficients u^a ($a = 1, \dots, A$) are completely arbitrary.

We thus see the emergence of another new feature in the theory, in addition to the appearance of constraints. It is that the general solution of the equations of motion may contain arbitrary functions of time (when $A \neq 0$), in agreement with the possible presence of a gauge symmetry.

First- and Second-Class Constraints

First- and second-class functions A function $F(q, p)$ is called a first-class function if it generates a canonical transformation that maps the constraint surface on itself. Thus, $F(q, p)$ is first class if its

Poisson brackets with all the constraints vanish weakly (i.e., are zero on the constraint surface),

$$[F, \phi_j] \approx 0, \quad j = 1, \dots, J \quad [18]$$

A function is second class otherwise, that is, if there is at least one constraint ϕ_j such that $[F, \phi_j] \neq 0$ (not even weakly). Second-class functions generate canonical transformations that do not leave the constraint surface invariant. Since canonical transformations that map the constraint surface on itself form a group, the Poisson bracket of two first-class functions is itself a first-class function.

Because the system is constrained to lie on the constraint surface, the only allowed canonical transformations are those that are generated by first-class functions. The importance of the distinction between first-class and second-class functions stems from this elementary fact. Note, in particular, that the time evolution is generated – as it should – by a first-class generator since the equations of motion [11] can be rewritten as

$$\dot{F} \approx [F, H'] + u^a [F, V_a^m \phi_m] \quad [19]$$

with

$$H' = H + U^m \phi_m \quad [20]$$

One has both $[H', \phi_m] \approx 0$ and $[V_a^m \phi_m, \phi_j] \approx 0$.

Splitting of the constraints One can separate the constraints between first-class and second-class constraints. This can be achieved by considering the matrix $C_{j'j''}$ of the Poisson bracket of the constraints,

$$C_{j'j''} = [\phi_j, \phi_{j'}], \quad j, j' = 1, \dots, J \quad [21]$$

One has the following theorem due to Dirac.

Theorem 1 *If $\det C_{j'j''} \approx 0$, there exists at least one first-class constraint among the ϕ_j 's.*

Proof Straightforward: if $\det C_{j'j''} \approx 0$, one can find a nontrivial solution λ^j of $\lambda^j C_{j'j''} \approx 0$. The corresponding constraint $\lambda^j \phi_j$ is easily verified to be first class.

By redefining the constraints as $\phi_j \rightarrow \bar{\phi}_j = a_j^{j'} \phi_{j'}$ with $a_j^{j'}(q, p)$ invertible, one can bring the Poisson brackets of the constraints to the form

$$[\gamma_a, \gamma_b] = 0, \quad [\gamma_a, \chi_\alpha] = 0, \quad [\chi_\alpha, \chi_\beta] = C_{\alpha\beta} \quad [22]$$

with $(\bar{\phi}_j) \equiv (\gamma_a, \chi_\alpha)$ and where the matrix $C_{\alpha\beta}$ is invertible. (We assume, for simplicity, throughout that the rank of the matrix $C_{j'j''}$ is constant on the constraint surface (“regular case”).) In this representation, the constraints are completely split into first-class constraints (γ_a) and second-class

constraints (χ_α): there is no first-class constraint left among the χ_α 's, and the set $\{\gamma_a\}$ exhausts all the first-class constraints. Note that now the index $a=1, \dots, A, A+1, \dots, \bar{A}$ runs over all (primary and secondary) first-class constraints.

This separation of the constraints into first-class and second-class constraints is quite important because, as already seen above, the first-class constraints generate admissible canonical transformations, while the second-class constraints do not.

For a bosonic system, the matrix $C_{\alpha\beta}$ is antisymmetric. As $C_{\alpha\beta}$ is invertible, this implies that the number of second-class constraints is even. In the fermionic case, $C_{\alpha\beta}$ is symmetric (in the fermionic sector) and, therefore, the number of second-class constraints can be even or odd.

First-class constraints and gauge symmetries The first-class constraints not only map the constraint surface on itself, but generate, in fact, transformations that do not change the physical state of the system, that is, gauge transformations. Indeed, the presence of arbitrary functions in the solutions of the equations of motion indicates that the q 's and the p 's involve some redundancy and are not all physically distinct. Only those phase-space functions whose time evolution does not depend on the arbitrary functions u^a are observables.

That the first-class constraints generate gauge transformations is rather clear in the case of the first-class primary constraints, since these appear explicitly in the generator of the time evolution multiplied by arbitrary functions. That it also holds for the first-class secondary constraints is known as the ‘‘Dirac conjecture.’’ This conjecture can be proved under reasonable assumptions (see, e.g., Henneaux *et al.* 1990). The reason that the secondary first-class constraints also correspond to gauge transformations is that they appear in the brackets of the Hamiltonian with the primary first-class constraints. Thus, different choices of arbitrary functions u^a in the dynamical equations of motion will lead to phase-space points that differ by a canonical transformation whose generator involves the secondary first-class constraints as well.

In any case, as noted below, one must identify the phase-space points in the same orbit generated by all the first-class constraints (primary and secondary) in order to get a reduced space with a symplectic structure (‘‘reduced phase space’’). For this reason, one postulates that the first-class constraints always generate gauge transformations, even for systems which are counterexamples to the Dirac conjecture (i.e., in that case, one defines the gauge

transformations as being the transformations generated by the first-class constraints).

The extended Hamiltonian H_E is defined to be the sum of the first-class Hamiltonian [20] and of all the first-class constraints γ_a multiplied by an arbitrary Lagrange multiplier,

$$H_E = H' + v^a \gamma_a \quad [23]$$

(with a summed from 1 to \bar{A}). It is the generator of the time evolution in which the complete gauge symmetry is fully displayed.

Elimination of second-class constraints – Dirac brackets Second-class constraints do not generate permissible canonical transformations, since they do not map the constraint surface on itself. For this reason, it is convenient to eliminate them. This can consistently be done by using the Dirac brackets instead of the Poisson brackets. By definition, the Dirac bracket $[F, G]_D$ of two phase-space functions F and G is given by

$$[F, G]_D = [F, G] - [F, \chi_\alpha] C^{\alpha\beta} [\chi_\beta, G] \quad [24]$$

where $C^{\alpha\beta}$ is the inverse to $C_{\alpha\beta}$,

$$C^{\alpha\beta} C_{\beta\gamma} = \delta_\gamma^\alpha$$

(which exists since the χ_α 's are second class). As shown by Dirac, the bracket [24] is indeed a bracket (antisymmetry, derivation property, and Jacobi identity). Furthermore, it fulfills the crucial property that the Dirac bracket of anything with any second-class constraint is zero,

$$[F, \chi_\alpha]_D = 0 \quad (F \text{ arbitrary}) \quad [25]$$

Thus, one can consistently eliminate the second-class constraints and replace the Poisson bracket by the Dirac bracket. Once this is done, one has fewer canonical variables and only first-class constraints remain (if any). It also follows from the definition that the Dirac bracket of two first-class functions is equal to their Poisson bracket.

Gauge conditions One can push the reduction procedure further and eliminate the first-class constraints by means of gauge conditions. Gauge conditions $C_a=0$ are conditions on the phase-space variables which do not follow from the Lagrangian and which have the property that they cut each gauge orbit once and only once. Since the gauge transformations are generated by the first-class constraints, this requirement is (locally) equivalent to

$$[C_a, \gamma_b] \varepsilon^b \approx 0 \Rightarrow \varepsilon^b \approx 0 \quad [26]$$

That is, the constraints (γ_a, C_b) form together a second-class system: there is no first-class constraint left once the conditions $C_a = 0$ are included. One can then eliminate all the constraints and gauge conditions and introduce the corresponding Dirac bracket. For gauge-invariant functions, this Dirac bracket coincides with the original Poisson bracket.

The reduced phase space is the unconstrained space obtained after this reduction, equipped with the Dirac bracket. It has dimension $2n - s - 2\bar{A}$, where $2n$ is the dimension of the original phase space, s is the number of second-class constraints, and \bar{A} is the number of first-class constraints. In the bosonic case, this number is even (as it should) because s is even. One sees that “first-class constraints strike twice” since they need gauge conditions.

The observables of the theory are the reduced phase-space functions. They form a Poisson algebra, the relevant reduced phase-space bracket being the Dirac bracket associated with all the constraints and gauge conditions. The symplectic structure defined in the reduced phase space is nondegenerate because one has removed all the first-class constraints.

The definition of reduced phase space given above is useful in practice but has the conceptual drawback of relying on gauge conditions. This approach does not display clearly its intrinsic significance and, furthermore, in the case of the so-called Gribov problems (global obstructions to cutting each gauge orbit once and only once), may yield the incorrect expectation that the reduced phase space does not exist. We shall provide a more intrinsic definition below, which does not involve gauge conditions.

Examples

First example (see eqn [2]). There is here one primary constraint, namely $\pi_\lambda = 0$. The canonical Hamiltonian is $(1/2)((p_1)^2 + (p_2)^2) + \lambda(p_1 + p_2)$. The consistency algorithm yields the secondary constraint $p_1 + p_2 = 0$ and no condition on the u 's. The constraints are first class. They generate the gauge transformations $q^1 \rightarrow q^1 + \varepsilon$, $q^2 \rightarrow q^2 + \varepsilon$, and $\lambda \rightarrow \lambda + \eta$, which coincide with the Lagrangian gauge transformations if one identifies η with $\dot{\varepsilon}$ (ε and $\dot{\varepsilon}$ are, of course, independent at any given time). One can fix the gauge by means of the gauge conditions $\lambda = 0$, $q^1 + q^2 = 0$. The reduced phase space is two-dimensional and the observables can be identified with the functions of the gauge-invariant variables $(1/2)(q^1 - q^2)$ and $p_1 - p_2$, which are conjugate. Any other gauge condition leads to the same reduced phase space.

Second example (see eqn [4]). The primary constraints are $p_1 - z^2 = 0$ and $p_2 = 0$ and define a two-dimensional plane in the four-dimensional phase space (z^1, z^2, p_1, p_2) . The consistency algorithm forces $u^1 = z^2$ and $u^2 = 0$ and does not bring any further constraint. The constraints are second class since $[p_2, p_1 - z^2] = 1$. One can eliminate p_1 and p_2 through the constraints. The Dirac brackets of the remaining variables vanish, except $[z^1, z^2] = 1$. The reduced phase is the space of the z 's, with z^2 conjugate to z^1 . The Hamiltonian is the free-particle Hamiltonian, $H = (1/2)(z^2)^2$. Thus, one recovers the original description which was already in Hamiltonian form. (The recognition that a system is already in first-order form often enables one to shortcut some aspects of the Dirac procedure by not introducing the unnecessary momenta which would in any case be eliminated in the end.)

Quantization

The phase space of physical interest is the reduced phase space and the physical algebra is the algebra of the observables. The quantization of the theory then amounts to quantizing the algebra of the observables. This can be achieved along two different lines:

1. *Reduce then quantize*: In this direct approach, one represents as quantum operators only the reduced phase-space functions. There is no operator associated with non-gauge-invariant functions.
2. *Quantize then reduce*: In this approach, one represents as quantum operators the bigger algebra of functions of all the phase-space variables. One must then take into account the constraints. The second-class constraints are enforced as operator equations, which is consistent with the correspondence rule that the commutator in the quantum theory is $i\hbar$ times the Dirac bracket,

$$AB - BA = i\hbar[A, B]_D \quad [27]$$

(plus higher-order terms in \hbar). The first-class constraints are implemented in a more subtle way. It would be inconsistent to impose them as operator equations since in general $[\gamma_a, F]_D \neq 0$ (even in the Dirac bracket). What one does is to impose them as conditions on the physical states: these are defined as the states annihilated by the first-class constraints,

$$\gamma_a|\psi\rangle = 0 \quad [28]$$

For simple systems, it is easy to verify that the two procedures are equivalent. There is yet another

approach, in which one extends the system rather than reduce it. This is the Becchi–Rouet–Stora–Tyutin (BRST) approach, in which the new variables are called ghosts.

Geometric Description

We defined above first-class and second-class constraints through algebraic means. It turns out that these definitions also have a geometrical interpretation, which sheds considerable insight into their nature.

The phase-space symplectic 2-form σ induces, by pullback, a 2-form σ_Σ on the constraint surface Σ . While σ is of maximal rank, this may not be the case for the induced σ_Σ , which may be degenerate. In fact, the rank of σ_Σ fails to be equal to the maximum rank $2n - J$ (where J is the total number of constraints) by precisely the number \bar{A} of first-class constraints.

Indeed, the Hamiltonian vector fields X_{γ_a} associated with the first-class constraints are tangent to the constraint surface Σ and are null eigenvectors of σ_Σ ,

$$\sigma_\Sigma(X_{\gamma_a}, Y) = 0 \quad \forall Y \text{ tangent to } \Sigma \quad [29]$$

as an immediate consequence of the first-class property. Here, all first-class constraints (primary and secondary) yield a null eigenvector. The integral surfaces of the vector fields X_{γ_a} are the gauge orbits. The reduced phase space is nothing else but the quotient space of the constraint surface by the gauge orbits. The 2-form induced in the quotient space is invertible because one has removed all degeneracy directions (including the ones associated with secondary first-class constraints). Reaching the reduced phase space falls under the scope of Hamiltonian reduction. The observables are the functions on the reduced phase space.

Thus, the reduced phase space is obtained through a two-step procedure. First, one restricts the functions to functions on the constraint surface Σ . One may view the algebra $C^\infty(\Sigma)$ of smooth functions on Σ as the quotient algebra $C^\infty(P)/\mathcal{N}$ of the algebra $C^\infty(P)$ of smooth phase-space functions by the ideal \mathcal{N} of phase-space functions that vanish on the constraint surface σ . The second step in the reduction procedure is to impose the gauge-invariant condition on the

functions in $C^\infty(\Sigma)$, that is, to impose that they are constant along the gauge orbits \mathcal{O} . Assuming all necessary smoothness and regularity conditions to be fulfilled (i.e., that the orbits fiber which is, for instance, the case if the gauge orbits are the orbits of a free and proper group action), one may denote the algebra of observables as $C^\infty(\Sigma/\mathcal{O})$. This algebra is a Poisson algebra because the induced 2-form on the quotient space Σ/\mathcal{O} is nondegenerate. The algebraic description of the observables underlies the BRST construction.

It is interesting to note that in the covariant approach to phase space, a similar two-step reduction procedure occurs. What plays the role of the constraint surface is the stationary surface in the space of all histories $q^i(t)$ of the dynamical variables. The gauge symmetry acts on this space and the reduced phase space is just the quotient space. One can establish the equivalence of the two descriptions (Barnich *et al.* 1991).

See also: Batalin–Vilkovisky Quantization; BRST Quantization; Canonical General Relativity; Operads; Perturbative Renormalization Theory and BRST; Quantum Dynamics in Loop Quantum Gravity; Quantum Field Theory: A Brief Introduction.

Further Reading

- Anderson JL and Bergmann PG (1951) Constraints in covariant field theories. *Physical Review* 83: 1018.
- Barnich G, Henneaux M, and Schombld C (1991) On the covariant description of the canonical formalism. *Physical Review D* 44: 939.
- Dirac PAM (1950) Generalized Hamiltonian dynamics. *Canadian Journal of Mathematics* 2: 129.
- Dirac PAM (1967) *Lectures on Quantum Mechanics*. New York: Academic Press.
- Flato M, Lichnerowicz A, and Sternheimer D (1976) Deformations of Poisson brackets, Dirac brackets and applications. *Journal of Mathematical Physics* 17: 1754.
- Hanson A, Regge T, and Teitelboim C (1976) *Constrained Hamiltonian Systems*. Rome: Accad. Naz. dei Lincei.
- Henneaux M and Teitelboim C (1992) *Quantization of Gauge Systems*. Princeton: Princeton University Press.
- Henneaux M, Teitelboim C, and Zanelli J (1990) Gauge invariance and degree of freedom count. *Nuclear Physics B* 332: 169.
- Marsden JE and Weinstein A (1974) Reduction of symplectic manifolds with symmetry. *Reports on Mathematical Physics* 5: 121.

Constructive Quantum Field Theory

G Gallavotti, Università di Roma “La Sapienza,”
Rome, Italy

© 2006 G Gallavotti. Published by Elsevier Ltd.
All rights reserved.

Euclidean Quantum Fields

The construction of a relativistic quantum field is still an open problem for fields in spacetime dimension $d \geq 4$. The conceptual difficulty that sometimes led to fear an incompatibility between nontrivial quantum systems and special relativity has however been solved in the case of dimension $d=2,3$ although, so far, has not influenced the corresponding debate on the foundations of quantum mechanics, still much alive.

It began in the early 1960s with Wightman’s work on the axioms and the attempts at understanding the mathematical aspects of renormalization theory and with Hepp’s renormalization theory for scalar fields. The breakthrough idea was, perhaps, Nelson’s realization that the problem could really be studied in Euclidean form. A solution in dimensions $d=2,3$ has been obtained in the 1960s and 1970s through a remarkable series of papers by Nelson, Glimm, Jaffe, and Guerra. While the works of Nelson and Guerra relied on the “Euclidean approach” (see below) and on $d=2$, the early works of Glimm and Jaffe dealt with $d=3$ making use of the “Minkowskian approach” (based on second quantization) but making already use of a “multiscale analysis” technique. The latter received great impulsion and systematization by the adoption of Wilson’s views and methods on renormalization: in physics terminology, renormalization group methods; a point of view taken here following the Euclidean approach. The solution dealt initially with scalar fields but it has been subsequently considerably extended.

The Euclidean approach studies quantum fields through the following problems:

1. existence of the functional integrals defining the generating functions (see below) of the probability distribution of the interacting fields in finite volume: the “ultraviolet stability problem,”
2. existence of the infinite-volume limit of the generating functions: the “infrared problem,” and
3. check that the infinite volume generating functions satisfy the axioms needed to pass from the Euclidean, probabilistic, formulation to a Minkowskian formulation guaranteeing the existence of the Hamiltonian operator,

relativistic covariance, Ruelle–Haag scattering theory: the “reconstruction problem.”

The characteristic problem for the construction of quantum fields is (1) and here attention will be confined to it with the further restriction to the paradigmatic massive scalar fields cases. The dimension d of the spacetime will be $d=2,3$ unless specified otherwise.

Given a cube Λ of side L , $\Lambda \subset \mathbb{R}^d$, consider the following functional integral on the space of the fields on Λ , that is, on functions $\varphi_\xi^{(\leq N)}$ defined for $\xi \in \Lambda$,

$$Z_N(\Lambda, f) = \int \exp \left(- \int_\Lambda \left(\lambda_N \varphi_\xi^{(\leq N)^4} + \mu_N \varphi_\xi^{(\leq N)^2} + \nu_N + f_\xi \varphi_\xi^{(\leq N)} \right) d\xi \right) P_N(d\varphi^{(\leq N)}) \quad [1]$$

The fields $\varphi_\xi^{(\leq N)}$ are called “Euclidean” fields with ultraviolet cutoff $N > 0$, f_ξ is a smooth function with compact support bounded by $|f_\xi| \leq 1$ (for definiteness), the constants $\lambda_N > 0, \mu_N, \nu_N$ are called “bare couplings,” and P_N is a Gaussian probability distribution defining the free-field distribution with mass m and ultraviolet cutoff N ; the probability distribution P_N is determined by its “covariance” $C_{\xi, \eta}^{(\leq N)} \stackrel{\text{def}}{=} \int \varphi_\xi^{(\leq N)} \varphi_\eta^{(\leq N)} dP_N$, which in the physics literature is called a “propagator,” given by

$$C_{\xi, \eta}^{(\leq N)} = \frac{1}{(2\pi)^d} \sum_{n \in \mathbb{Z}^d} \int \frac{e^{ip \cdot (\xi - \eta + nL)}}{p^2 + m^2} \chi_N(|p|) d^d p \quad [2]$$

The sum over the integers $n \in \mathbb{Z}^d$ is introduced so that the field $\varphi_\xi^{(\leq N)}$ is periodic over the box Λ : this is not really necessary as in the limit $L \rightarrow \infty$ either translation invariance would be recovered or lack of it properly understood, but it makes the problem more symmetric and generates a few technical simplifications; here $\chi_N(z)$ is a “regularizer” and a standard choice is

$$\chi_N(|p|) = \frac{m^2(\gamma^{2N} - 1)}{p^2 + \gamma^{2N}m^2}$$

with $\gamma > 1$, which is such that

$$\begin{aligned} \frac{\chi_N(|p|)}{p^2 + m^2} &\equiv \frac{1}{p^2 + m^2} - \frac{1}{p^2 + \gamma^{2N}m^2} \\ &\equiv \sum_{b=1}^N \left(\frac{1}{p^2 + \gamma^{2(b-1)}m^2} - \frac{1}{p^2 + \gamma^{2b}m^2} \right) \quad [3] \end{aligned}$$

here $\gamma > 1$ can be chosen arbitrarily: so $\gamma=2$. If $d > 3$, the above regularization will not be sufficient and a χ_N decaying faster than p^{-2} would be needed.

A simple estimate yields, if $\varepsilon \in (0, 1)$ is fixed and c is suitably chosen,

$$\begin{aligned} \left| C_{\xi, \eta}^{(\leq N)} \right| &\leq c \gamma^{(d-2)N} e^{-m|\xi-\eta|} \\ \left| C_{\xi, \eta}^{(\leq N)} - C_{\xi, \eta'}^{(\leq N)} \right| &\leq c \gamma^{(d-2)N} (\gamma^N m |\eta - \eta'|)^\varepsilon \end{aligned} \quad [4]$$

with $\gamma^{(d-2)N}$ interpreted as N if $d = 2$.
The

$$\zeta(f) = \log \frac{Z_N(\Lambda, f)}{Z_N(\Lambda, 0)}$$

defines a “generating function” of a probability distribution P_{int} over the fields on Λ which will be called the “distribution with φ^4 -interaction” regularized on Λ and at length scale $m^{-1}\gamma^{-N}$: the integral, in [1],

$$\begin{aligned} V_N(\varphi^{(\leq N)}) \stackrel{\text{def}}{=} &\int_{\Lambda} \left(\lambda_N \varphi_{\xi}^{(\leq N)^4} + \mu_N \varphi_{\xi}^{(\leq N)^2} \right. \\ &\left. + \nu_N + f_{\xi} \varphi_{\xi}^{(\leq N)} \right) d^d \xi \end{aligned} \quad [5]$$

will be called the “interaction potential” with external field f . The regularization is introduced to guarantee that the integral [1], $\int e^{V_N} dP_N$, is well defined if $\lambda_N > 0$. The momenta of P_{int} are the functional derivatives of $\zeta(f)$: they are called “Schwinger functions.”

The problem (1) can now be made precise: it is to show the existence of λ_N, μ_N, ν_N so that the limit

$$\lim_{N \rightarrow \infty} \frac{Z_N(\Lambda, f)}{Z_N(\Lambda, 0)}$$

exists for all f and is not Gaussian, that is, it is not the exponential of a quadratic form in f : which would be the case if $\lambda_N, \mu_N \rightarrow 0$ fast enough: the last requirement is of course essential because the Gaussian case describes, in the physical interpretation, free fields and noninteracting particles, that is, it is trivial. Note that ν_N does not play a role: its introduction is useful to be able to study separately the numerator and the denominator of the fraction

$$\frac{Z_N(\Lambda, f)}{Z_N(\Lambda, 0)}$$

For more details, the reader is referred to [Wightman and Gårding \(1965\)](#), [Streater and Wightman \(1964\)](#), [Nelson \(1966\)](#), [Guerra \(1972\)](#), [Osterwalder and Schrader \(1973\)](#), and [Simon \(1974\)](#).

The Regularized Free Field

Since the propagator, see [4], decays exponentially over a scale m^{-1} and is smooth over a scale $m^{-1}\gamma^{-N}$,

the fields $\varphi_{\xi}^{(\leq N)}$ sampled with distribution P_N are rather singular objects. Their properties cannot be described by a single length scale: they are extremely large for large N , take independent values only beyond distances of order m^{-1} but, at the same time, they look smooth only on the much smaller scale $m^{-1}\gamma^{-N}$. Their essential feature is that fixed $\varepsilon < 1$, for example, $\varepsilon = 1/2$, with P_N -probability 1 there is $B > 0$ such that (interpreting $\gamma^{(d-2)/2N}$ as N if $d = 2$)

$$\begin{aligned} \left| \varphi_{\xi}^{(\leq N)} \right| &\leq B \gamma^{N(d-2)/2} \\ \left| \varphi_{\xi}^{(\leq N)} - \varphi_{\eta}^{(\leq N)} \right| &< B \gamma^{N(d-2)/2} (\gamma^N m |\xi - \eta|)^{\varepsilon/2} \end{aligned} \quad [6]$$

and furthermore the probability of the relations in [6] will be N -independent, that is, $\varphi_{\xi}^{(\leq N)}$ are bounded and roughly of size $\gamma^{N(d-2)/2}$ as $N \rightarrow \infty$ and, on a very small length scale $m^{-1}\gamma^{-N}$, almost constant.

Substantial control on the field $\varphi_{\xi}^{(\leq N)}$ statistically sampled with distribution P_N can be obtained by decomposing it, through [3], into “components of various scales”: that is, as a sum of statistically mutually independent fields whose properties are entirely characterized by a single scale of length. This means that they have size of order 1 and are independent and smooth on the same length scale.

Assuming the side of Λ to be an integer multiple of m^{-1} , let \mathcal{Q}_b be a pavement of Λ into boxes of side $m^{-1}\gamma^{-b}$, imagined hierarchically arranged so that the boxes of \mathcal{Q}_b are exactly paved by those of \mathcal{Q}_{b+1} .

Define $z_{\xi}^{(b)}$ to be the random field with propagator $C_{\xi, \eta}^{(b)}$ with Fourier transform

$$\sum_{p \in \mathbb{Z}^d} \left(\frac{1}{p^2 + \gamma^{-2} m^2} - \frac{1}{p^2 + m^2} \right) e^{i p \cdot L \gamma^b}$$

so that $\varphi_{\xi}^{(\leq N)}$ and its propagator $C_{\xi, \eta}^{(\leq N)}$ can be represented, see [2], [3], as

$$\begin{aligned} \varphi_{\xi}^{(\leq N)} &\equiv \sum_{b=1}^N \gamma^{b(d-2)/2} z_{\gamma^b \xi}^{(b)} \\ C_{\xi, \eta}^{(\leq N)} &= \sum_{b=1}^N \gamma^{b(d-2)} C_{\gamma^b \xi, \gamma^b \eta}^{(b)} \end{aligned} \quad [7]$$

where the fields $z^{(b)}$ are independently distributed Gaussian fields. Note that the fields $z^{(b)}$ are also almost identically distributed because their propagator is obtained by periodizing over the period $\gamma^b L$ the same function

$$\overline{C}_{\xi, \eta}^{(0)} \stackrel{\text{def}}{=} \int \frac{e^{i p \cdot (\xi - \eta)} d p}{(2\pi)^d} \left(\frac{1}{p^2 + \gamma^{-2} m^2} - \frac{1}{p^2 + m^2} \right)$$

that is, their propagator is

$$C_{\xi, \eta}^{(b)} = \sum_{n \in \mathbb{Z}^d} \bar{C}_{\xi, \eta + \gamma^b n}^{(0)}$$

The reason why they are not exactly equally distributed is that the field $z_{\xi}^{(b)}$ is periodic with period $\gamma^b L$ rather than L . But proceeding with care the sum over n in the above expressions can be essentially ignored: this is a little price to pay if one wants translation invariance built in the analysis since the beginning.

The representation [7] defines a “multiscale representation” of the field $\varphi_{\xi}^{(\leq N)}$. Smoothness properties for the field $\varphi_{\xi}^{(\leq N)}$ can be read from those of its “components” $z_{\xi}^{(b)}$. Define, for $\Delta \in \mathcal{Q}_0$,

$$\|z^{(b)}\|_{\Delta} = \max_{\substack{\xi \in \Delta, \eta \in \Lambda \\ |\xi - \eta| \leq m^{-1}}} \left(|z_{\xi}^{(b)}| + \tau \frac{|z_{\xi}^{(b)} - z_{\eta}^{(b)}|}{|\xi - \eta|^{1/4}} \right) \quad [8]$$

and τ will be chosen $\tau = 0$ or $\tau = 1$ as needed (in practice $\tau = 0$ if $d = 2$ and $\tau = 1$ if $d = 3$): $\tau = 1$ will allow us to discuss some smoothness properties of the fields which will be necessary (e.g., if $d = 3$). Then the size $\|z\|_{\Delta}$ of any field $z^{(b)}$, for all $b \geq 1$, is estimated by

$$\begin{aligned} P\left(\max_{\Delta \in \mathcal{Q}_0} \|z\|_{\Delta} \leq B\right) &\geq e^{-ce^{-c'B^2}|\Lambda|} \\ P(\|z\|_{\Delta} \geq B_{\Delta}, \forall \Delta \in \mathcal{D}) &\leq \prod_{\Delta \in \mathcal{D}} ce^{-c'B_{\Delta}^2} \end{aligned} \quad [9]$$

where P is the Gaussian probability distribution of z , \mathcal{D} is any collection of boxes $\Delta \in \mathcal{Q}_0$ and $c, c' > 0$ are suitable constants. The [9] imply in particular [6]. The estimates [9] follow from the Markovian nature of the Gaussian field $z^{(b)}$, that is, from the fact that the propagator is the Green’s function of an elliptic operator (of fourth order, see the first of [3]), with constant coefficients which implies also the inequalities (fixing $\varepsilon \in (0, 1)$)

$$\begin{aligned} |C_{\xi, \eta}^{(b)}| &\equiv \left| \int z_{\xi} z_{\eta} P(dz) \right| \leq ce^{-m|\xi - \eta|} \\ |C_{\xi, \eta}^{(b)} - C_{\xi, \eta'}^{(b)}| &\leq c(m|\eta - \eta'|)^{\varepsilon} \end{aligned} \quad [10]$$

where $|\xi - \eta|$ is reinterpreted as the distance between ξ, η measured over the periodic box $\gamma^b \Lambda$ (hence $|\xi - \eta|$ differs from the ordinary distance only if the latter is of the order of $\gamma^b L$). The interpretation of [10] is that $z_{\xi}^{(b)}$ are essentially bounded variables which, on scale $\sim m^{-1}$, are essentially constant and furthermore beyond length $\sim m^{-1}$ are essentially independently distributed.

For more details, the reader is referred to Wilson (1970, 1972) and Gallavotti (1981, 1985).

Perturbation Theory

The naive approach to the problem is to fix $\lambda_N \equiv \lambda > 0$ and to develop $Z_N(\Lambda, f)$ or, more conveniently and equivalently, $(1/|\Lambda|) \log Z_N(\Lambda, f)$ in powers of λ . If one fixes *a priori* μ_N, ν_N independent of N , however, even a formal power series is not possible: this is trivially due to the divergence of the coefficients of the power series, already to second order, for generic f in the limit $N \rightarrow \infty$. Nevertheless it is possible to determine $\mu_N(\lambda), \nu_N(\lambda)$ as functions of N and λ so that a formal power series exists (to all orders in λ): this is the key result of renormalization theory.

To find the perturbative expansion, the simplest is to use a graphical representation of the coefficients of the power expansion in λ, μ_N, ν_N, f and the Gaussian integration rules which yield (after a classical computation) that the coefficient of $\lambda^n \mu_N^p f_{\xi_1} \dots f_{\xi_r}$ is obtained by considering the graph elements shown in Figure 1, where the segments will be called half-lines and the graph elements will be called, respectively, “coupling” or “ φ^4 -vertex,” “mass vertex,” “vacuum vertex,” and “external vertex.”

The half-lines of the graph elements are considered distinct (i.e., imagine a label attached to distinguish them). Then consider all possible *connected* graphs G obtained by first drawing, respectively, n, p, r graph elements in Figure 1, which are not vacuum vertices, with their nodes marked by points in Λ named $\xi_1, \dots, \xi_n, \xi_{n+1}, \dots, \xi_{n+p+r}$; and form all possible graphs obtained by attaching pairs of half-lines emerging from the vertices of the graph elements. These are the “nontrivial graphs.” Furthermore, consider also the single “trivial” graph formed just by the third graph element and consisting of a single point. All graphs obtained in this way are particular Feynman graphs.

Given a nontrivial graph G (there are many of them) we define its value to be the product

$$\begin{aligned} W_G(\xi_1, \dots, \xi_n, \xi_{n+1}, \dots, \xi_{n+p+r}) \\ = (-1)^{n+p+r} \frac{\lambda^n \mu_N^p \prod f_{\xi_{n+p+i}}}{n! p! r!} \prod_{\ell} C_{\xi_{\ell}, \eta_{\ell}}^{(\leq N)} \end{aligned} \quad [11]$$

where the last product runs over all pairs $\ell = (\xi_{\ell}, \eta_{\ell})$ of half-lines of G that are joined and connect two vertices labeled by points ξ_{ℓ}, η_{ℓ} : “call line of G ” any such pair. If the graph consists of the single vacuum

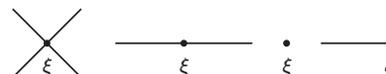


Figure 1 The graph elements to representing $\varphi_{\xi}^{(\leq N)4}, \varphi_{\xi}^{(\leq N)2}$, a constant $\varphi_{\xi}^{(\leq N)}$.

vertex its value will be ν_N . The series for $(1/|\Lambda|) \log Z_N(\Lambda, f)$ is then

$$-\nu_N + \frac{1}{|\Lambda|} \sum_G \int W_G(\xi_1, \dots, \xi_{n+p+r}) \prod_{j=1}^{n+p+r} d\xi_j \quad [12]$$

and the integral will be called the integrated graph value.

Suppose first that $\mu_N = \nu_N = 0$. Then if a graph G contains subgraphs like in **Figure 2**, the corresponding respective contribution to the integral in [12] (considering only the integrals over η and suitably taking care of the combinatorial factors) is a factor obtained by integrating over ξ the quantities

$$-6\lambda C_{\alpha\xi}^{(\leq N)} C_{\xi\xi}^{(\leq N)} C_{\xi\beta}^{(\leq N)}$$

or

$$\frac{4^2 \cdot 3!}{2!} \lambda^2 C_{\alpha\xi}^{(\leq N)} \int C_{\xi\eta}^{(\leq N)3} C_{\eta\beta}^{(\leq N)} d\eta \quad [13]$$

which if $d=3$ diverge as $N \rightarrow \infty$ as γ^N or, respectively, as N ; the second factor does not diverge in dimension $d=2$ while the first still diverges as N . The divergences arise from the fact that as $\xi - \eta \rightarrow 0$ the propagator behaves as $|\xi - \eta|^{-N}$ if $d=3$ or as $-\log|\xi - \eta|$ if $d=2$, all the way until saturation occurs at distance $|\xi - \eta| \simeq m^{-1} \gamma^{-N}$: for this reason the latter divergences are called “ultraviolet divergences.”

However, if we set $\mu_N \neq 0$, then for every graph containing a subgraph like those in **Figure 2** there is another one identical except that the points α, β are connected via a mass vertex, see **Figure 1**, with the vertex in ξ , by a line $\alpha\xi$ and a line $\xi\beta$; the new graph value receives a contribution from the mass vertex inserted in ξ between α and β simply given by a factor $-\mu_N$. Therefore if we fix, for $d=3$,

$$\mu_N = -6\lambda C_{\xi\xi}^{(\leq N)} + \frac{4^2 \cdot 3!}{2} \lambda^2 \int_{\Lambda} C_{\xi\eta}^{(\leq N)3} d\eta \stackrel{\text{def}}{=} -6\lambda C_{\xi\xi}^{(\leq N)} + \delta\mu_N \quad [14]$$

we can simply consider graphs which do not contain any mass graph element and in which there are no subgraphs like the first in **Figure 2** while the subgraphs like the second in **Figure 2** do not contribute a factor $\int C_{\alpha\xi}^{(\leq N)} C_{\xi\eta}^{(\leq N)3} C_{\eta\beta}^{(\leq N)} d\eta$ but a renormalized factor

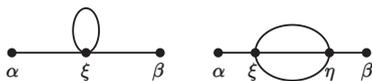


Figure 2 Divergent subgraphs, if $d=3$. If $d=2$ only the first diverges.

$\int C_{\alpha\xi}^{(\leq N)} C_{\xi\eta}^{(\leq N)3} (C_{\eta\beta}^{(\leq N)} - C_{\xi\beta}^{(\leq N)}) d\eta$. If $d=2$, we only need to define μ_N as the first term on the right-hand side (RHS) of [14] and we can leave the subgraphs like the second in **Figure 2** as they are (without any renormalization).

Graphs without external lines are called vacuum graphs and there are a few such graphs which are divergent. Namely, if $d=3$, they are the first three drawn in **Figure 3**; furthermore, if μ_N is set to the above nonzero value a new vacuum graph, the fourth in **Figure 3**, can be formed. Such graphs contribute to the graph value, respectively, the terms in the sum

$$-3\lambda C_{\xi_1\xi_1}^{(\leq N)2} + \frac{4!}{2} \lambda^2 \int C_{\xi_1\xi_2}^{(\leq N)4} d\xi_2 - \frac{2^3 \cdot 3!^3}{3!} \lambda^3 \times \int C_{\xi_1\xi_2}^{(\leq N)2} C_{\xi_2\xi_3}^{(\leq N)2} C_{\xi_3\xi_1}^{(\leq N)2} d\xi_2 d\xi_3 - \mu_N C_{\xi_1\xi_1}^{(\leq N)} \quad [15]$$

and diverge, respectively, as $\gamma^{2N}, \gamma^N, N, \gamma^{2N}$ if $d=3$ while, if $d=2$, only the first and the last (see [14]) diverge, like N^2 .

Therefore, if we fix ν_N as minus the quantity in [15] we can disregard graphs like those in **Figure 3**; if $d=2$ ν_N can be defined to be the sum of the first and last terms in [15].

The formal series in λ and f thus obtained is called the “renormalized series” for the field φ^4 in dimension $d=2$ or, respectively, $d=3$. Note that with the given definitions and choices of μ_N, ν_N the only graphs G that need to be considered to construct the expansion in λ and f are formed by the first and last graph elements in **Figure 1**, paying attention that the graphs in **Figure 3** do not contribute and, if $d=3$, the graphs with subgraphs like the second in **Figure 2** have to be computed with the modification described.

In the next section, it will be shown that the above are the only sources of divergences as $N \rightarrow \infty$ and therefore the problem of studying [1] is solved at the level of formal power series by the subtraction in [14]. This also shows that giving a meaning to the series thus obtained is likely to be much easier if $d=2$ than if $d=3$.

The coefficients of order k of the expansion in λ of $(1/|\Lambda|) \log Z_N(\Lambda, f)$ can be ordered by the number $2n$ of vertices representing external fields: and have the form $\int S_{2n}^{(k)}(\xi_1, \dots, \xi_{2n}) \prod_{i=1}^{2n} (f_{\xi_i} d\xi_i)$: the kernels $S_{2n}^{(k)}$ are the Schwinger functions of order $2n$, see the section “Euclidean quantum fields.”

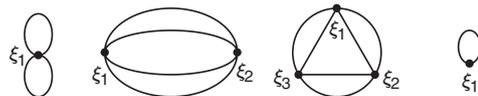


Figure 3 Divergent vacuum graphs.

Remark If $d=4$, the regularization at cutoff N in [2] is not sufficient as in the subtraction procedure smoothness of the first derivatives of the field $\varphi^{(\leq N)}$ is necessary, while the regularization [2] does not even imply [6], that is, not even Hölder continuity. A higher regularization (i.e., using a χ_N like the square of the χ_N in [3]). Furthermore, the subtractions discussed in the case $d=3$ are not sufficient to generate a formal power series and many more subtractions are needed: for instance, graphs with a subgraph like the one in Figure 4 would give a contribution to the graph value which is a factor

$$\lambda^2 \ell_N \stackrel{\text{def}}{=} \frac{2 \cdot 6^2}{2!} \lambda^2 \int_{\Lambda} C_{\xi\eta}^{(\leq N)2} d\eta$$

also divergent as $N \rightarrow \infty$ proportionally to N . Although this divergence could be canceled by changing λ into $\lambda_N = \lambda + \lambda^2 \ell_N$ the previously discussed cancelations would be affected and a change in the value of μ_N would become necessary; furthermore, the subtraction in [14] will not be sufficient to make finite the graphs, not even to second order in λ , unless a new term $-\alpha_N \int (\partial_{\xi} \varphi_{\xi}^{(\leq N)})^2 d\xi$ with $\alpha_N = (1/2)\lambda^2 \int \partial_{\eta} C_{\xi\eta}^{(\leq N)3} (\xi - \eta)^2$ is added in the exponential in [1].

But all this will not be enough and still new divergences, proportional to λ^3 , will appear.

And so on indefinitely, the consequence being that it will be necessary to define $\lambda_N, \mu_N, \alpha_N, \nu_N$ as formal power series in λ (with coefficients diverging as $N \rightarrow \infty$) in order to obtain a formal power series in λ for [1] in which all coefficients have a finite limit as $N \rightarrow \infty$. Thus, the interpretation of the formal renormalized series in the case $d=4$ is substantially different and naturally harder than the cases $d=2, 3$. Beyond formal perturbation expansions, the case $d=4$ is still an open problem: the most widespread conjecture is that the series cannot be given a meaning other than setting to 0 all coefficients of $\lambda^j, j > 0$. In other words, the conjecture claims, there should be no nontrivial solution to the ultraviolet problem for scalar φ^4 fields in $d=4$. But this is far from being proved, even at a heuristic level. The situation is simpler if $d \geq 5$: in such cases, it is impossible to find formal power series in λ for $(1/|\Lambda|) \log Z_N(\Lambda, f)$, even allowing $\lambda_N, \mu_N, \alpha_N, \nu_N$ to be formal power series in λ with divergent coefficients.



Figure 4 The simplest new divergent subgraph on $d=4$.

The distinctions between the cases $d=2, 3, 4, >4$ explain the terminology given to the φ^4 -scalar field theories calling them super-renormalizable if $d=2, 3$, renormalizable if $d=4$ and nonrenormalizable if $d > 4$. Since the (divergent) coefficients in the formal power series defining $\lambda_N, \mu_N, \alpha_N, \nu_N$ are called counter-terms, the φ^4 -scalar fields require finitely many counter-terms (see [14]) in the super-renormalizable cases and infinitely many in the renormalizable case. The nonrenormalizable cases ($d > 4$) cannot be treated in a way analogous to the renormalizable ones.

For more details, the reader is referred to Gallavotti (1985), Aizenman (1982), and Fröhlich (1982).

Finiteness of the Renormalized Series, $d=2, 3$: “Power Counting”

Checking that the renormalized series is well defined to all orders is a simple dimensional estimate characteristic of many multiscale arguments that in physics have become familiar with the name of “renormalization group arguments.”

Consider a graph G with $n+r$ vertices built over n graph elements with vertices ξ_1, \dots, ξ_n each with four half-lines and r graph elements with vertices $\xi_{n+1}, \dots, \xi_{n+r}$ representing the external fields: as remarked in the previous section, these are the only graphs to be considered to form the renormalized series.

Develop each propagator into a sum of propagators as in [7]. The graph G value will, as a consequence, be represented as a sum of values of new graphs obtained from G by adding scale labels on its lines and the value of the graph will be computed as a product of factors in which a line joining $\xi\eta$ and bearing a scale label h will contribute with $C_{\xi\eta}^{(h)}$ replacing $C_{\xi\eta}^{(\leq N)}$. To avoid proliferation of symbols, we shall call the graphs obtained in this way, i.e., with the scale labels attached to each line, still G : no confusion should arise as we shall, henceforth, only consider graphs G with each line carrying also a scale label.

The scale labels added on the lines of the graph G allow us to organize the vertices of G into “clusters”: a cluster of scale h consists in a maximal set of vertices (of the graph elements in the graph) connected by lines of scale $h' \geq h$ among which one at least has scale h .

It is convenient to consider the vertices of the graph elements as “trivial” clusters of highest scale: conventionally call them clusters of scale $N+1$.

The clusters can be of “first generation” if they contain only trivial clusters, of “second generation”

if they contain only clusters which are trivial or of the first generation, and so on.

Imagine to enclose in a box the vertices of graph elements inside a cluster of the first generation and then into a larger box the vertices of the clusters of the second generation and so on: the set of boxes ordered by inclusion can then be represented by a rooted tree graph whose nodes correspond to the clusters and whose “top points” are nodes representing the trivial clusters (i.e., the vertices of the graph).

If the maximum number of nodes that have to be crossed to reach a top point of the tree starting from a node v is n_v (v included and the top nodes included), then the node v represents a cluster of the n_v th generation. The first node before the root is a cluster containing all vertices of G and the root of the tree will not be considered a node and it can conventionally bear the scale label 0: it represents symbolically the value of the graph.

For instance, in Figure 5 a tree θ is drawn: its nodes correspond to clusters whose scale is indicated next to them; in the second part of the drawing, the trivial clusters as well as the clusters of the first generation are enclosed into boxes.

Then consider the next generation clusters, that is, the clusters which only contain clusters of the first generation or trivial ones, and draw boxes enclosing all the graph vertices that can be reached from each of them by descending the tree, etc. Figure 6 represents all boxes (of any generation) corresponding to the nodes of the tree in Figure 5. The representations of the clusters of a graph G by a tree or by hierarchically ordered boxes (see Figures 5 and 6) are completely equivalent provided inside each box not representing a top point of the tree the scale h_v of the corresponding cluster v is marked. For

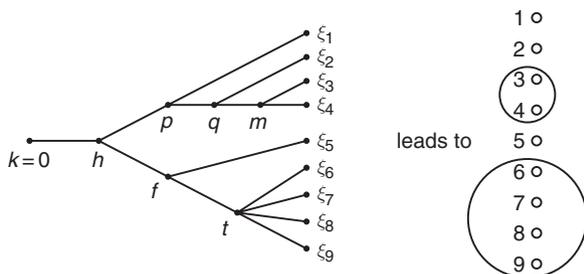


Figure 5 A tree and its clusters of generation 1 and 2.

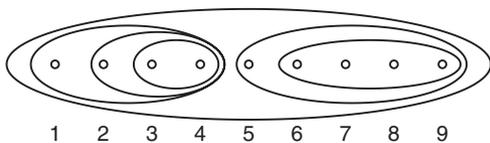


Figure 6 All clusters of any generation for the tree in Figure 5.

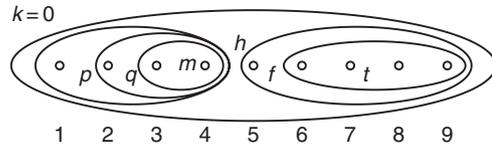


Figure 7 The clusters in Figure 6 after affixing the scale labels.

instance, in the case of Figure 6 one gets Figure 7. By construction, if two top points ξ and η are inside the same box b_v of scale h_v but not in inner boxes, then there is a path of graph lines joining ξ and η all of which have scales $\geq h_v$ and one at least has scale h_v .

Given a graph G , fix one of its points ξ_1 (say) and integrate the absolute value of the graph over the positions of the remaining points. The exponential decay of the propagators implies that if a point η is linked to a point η' by a line of scale h the integration over the position of η' is essentially constrained to extend only over a distance $\gamma^{-h}m^{-1}$. Furthermore, the maximum size of the propagator associated with a line of scale h is bounded proportionally to $\gamma^{(d-2)h}$. Therefore, recalling that $|f_\xi|$ is suppose bounded by 1, the mentioned integral can be immediately bounded by

$$\frac{\lambda^n}{n!r!} C^{n+r} I \stackrel{\text{def}}{=} \frac{\lambda^n C^{n+r}}{n!r!} \prod_{\ell} \gamma^{(d-2)/2h_{\ell}} \prod_v \gamma^{-dh_v(s_v-1)} \quad [16]$$

where, C being a suitable constant, the first product is over the half-lines ℓ composing the graph lines and the second is over the tree nodes (i.e., over the clusters of the graph G), s_v is the number of subclusters contained in the cluster v but not in inner clusters; and in [16] the scale of a half-line ℓ is h_{ℓ} if ℓ is paired with another half-line to form a line ℓ (in the graph G) of scale label h_{ℓ} .

Denoting by v' the cluster immediately containing v in G , by n_v^{inner} the number of half-lines in the cluster v , by n_v, r_v the numbers of graph elements of the first type or of the fourth type in Figure 1 with vertices in the cluster v , and denoting by n_v^e the number of lines which are not in the cluster v but have one extreme on a vertex in v (“lines external to v ”), the identities ($k=0$)

$$\begin{aligned} & \sum_{v>\text{root}} (h_v - k)(s_v - 1) \\ & \equiv \sum_{v>\text{root}} (h_v - h_{v'}) (n_v + r_v - 1) \\ & \sum_{v>\text{root}} (h_v - k) n_v^{\text{inner}} \equiv \sum_{v>\text{root}} (h_v - h_{v'}) \tilde{n}_v^{\text{inner}} \quad [17] \end{aligned}$$

with

$$\tilde{n}_v^{\text{inner}} \stackrel{\text{def}}{=} 4n_v + r_v - n_v^e$$

hold, so that the estimate [16] can be elaborated into

$$I \leq \prod_{v>r} \gamma^{-\rho_v(h_v-h_{v'})} \tag{18}$$

$$\rho_v \stackrel{\text{def}}{=} -d + (4-d)n_v + r_v \frac{d+2}{2} + \frac{d-2}{2} n_v^e$$

where $h_{v'} = k = 0$ if v is the first nontrivial node (i.e., $v' = \text{root}$), and an estimate of the integral of the absolute value of the graphs G with given tree structure but different scale labels is proportional to $\Sigma_{\{h_v\}} I < \infty$ if (and only if) $\rho_v > 0, \forall v$.

But there may be clusters v with only two external lines $n_v^e = 2$ and two graph vertices inside: for which $\rho_v = 0$. However, this can happen only if $d = 3$ and in only one case: namely if the graph G contains a subgraph of the second type in Figure 2 and the three intermediate lines form a cluster v of scale h_v while the other two lines are external to it: hence on scale $h' > h_v$. In this case, one has to remember that the subtraction in the previous section has led to a modification of the contribution of such a subgraph to the value of the graph (integrated over the position labels of the vertices). As discussed in the previous section, the change amounts to replacing the propagator $C_{\eta,\beta}^{(b')}$ by $C_{\eta,\beta}^{(b')} - C_{\xi,\beta}^{(b')}$.

This improves, in [18], the estimate of the contribution of the line joining η to β from being proportional to $\int C_{\xi\eta}^{(\leq h_v)3} C_{\eta\beta}^{(\leq h')}$ $d\eta$ to being proportional to $\int C_{\xi\eta}^{(\leq h_v)3} (C_{\eta\beta}^{(\leq h')} - C_{\xi\beta}^{(\leq h')}) d\eta$; and this changes the contribution of the line $\eta\beta$ from $\gamma^{(d-2)b'}$ to $\int e^{-m\gamma^{h_v}|\xi-\eta|} (\gamma^{b'}|\xi-\eta|)^{1/2} d\eta$ because $C^{(b')}$ is regular on scale $\gamma^{-b'}m^{-1}$, see [10] with $\varepsilon = 1/2$.

Since ξ, η are in a cluster of higher scale h_v this means that the estimate is improved by $\gamma^{-(1/2)(h_v-h')}$. In terms of the final estimate, this means that ρ_v in [18] can be improved to $\bar{\rho}_v = \rho_v + 1/2$ for the clusters for which $\rho_v = 0$. Hence, the integrated value of the graph G (after taking also into account the integration over the initially selected vertex ξ_1 , trivially giving a further factor $|\Lambda|$ by translation invariance), and summed over the possible scale labels is bounded proportionally to $|\Lambda| \Sigma_{\{h_v\}} I < \infty$ once the estimate of I is improved as described.

Note that the graphs contributing to the perturbation series for $(1/|\Lambda|) \log Z_N(\Lambda, f)$ to order λ^n are finitely many because the number r of external vertices is $r \leq 2n + 2$ (since graphs must be connected). Hence, the perturbation series is finite to all orders in λ .

The above is the renormalizability proof of the scalar φ^4 -fields in dimension $d = 2, 3$. The theory is renormalizable even if $d = 4$ as mentioned in the remark at the end of the previous section. The analysis would be very similar to the above: it is just a little more involved power-counting argument.

For more details, the reader is referred to Hepp (1966), Gallavotti (1985), sections 8 and 16.

Asymptotic Freedom ($d = 2, 3$). Heuristic Analysis

Finiteness to all orders of the perturbation expansions is by no means sufficient to prove the existence of the ultraviolet limit for $Z_N(\Lambda, f)$ or for $(1/|\Lambda|) \log Z_N(\Lambda, f)$: and *a priori* it might not even be necessary. For this purpose, the first step is to check uniform (upper and lower) boundedness of $Z_N(\Lambda, f)$ as $N \rightarrow \infty$.

The reason behind the validity of a bound $e^{|\Lambda|E_-(\lambda, f)} \leq Z_N(\Lambda, f) \leq e^{|\Lambda|E_+(\lambda, f)}$ with $E_{\pm}(\lambda, f)$ cutoff independent has been made very clear after the introduction of the renormalization group methods in field theory. The approach studies the integral $Z_N(\Lambda, f)$, recursively, decomposing the field $\varphi_{\xi}^{(\leq N)}$ into its regular components $z_{\xi}^{(b)}$, see [7], and integrating first over $z^{(N)}$, then over $z^{(N-1)}$ and so on.

The idea emerges naturally if the potential V_N in [1] and [4] is written in terms of the “normalized” variables $X_{\xi}^{(N)} \stackrel{\text{def}}{=} \gamma^{-N(d-2)/2} \varphi_{\xi}^{(\leq N)}$, see [6]; here if $d = 2$ the factor $\gamma^{(d-2)/2N}$ is interpreted as $N^{1/2}$.

The key remark is that as far as the integration over the small-scale component $z^{(N)}$ is concerned the field $X_{\xi}^{(N)}$ is a sum of two fields of size of order 1 (statistically),

$$X_{\xi}^{(N)} \equiv z_{\gamma^N \xi}^{(N)} + \gamma^{-(d-2)/2} X_{\xi}^{(N-1)}$$

if $d = 2$ this becomes

$$X_{\xi}^{(N)} \equiv \frac{1}{N^{1/2}} z_{\gamma^N \xi}^{(N)} + \frac{(N-1)^{1/2}}{N^{1/2}} X_{\xi}^{(N-1)}$$

and it can be considered to be smooth on scale $m^{-1}\gamma^{-N}$ (also statistically). Hence, approximately constant and of size of order $O(1)$ on the small cubes Δ of volume $\gamma^{-dN}m^{-d}$ of the pavement \mathcal{Q}_N introduced before [7]; at the same time it can be considered to take (statistically) independent values on different cubes of \mathcal{Q}_N . This is suggested by the inequalities [8]–[10].

Therefore, it is natural to decompose the potential V_N , see [5], as a sum over the small cubes Δ of volume $\gamma^{-dN}m^{-d}$ of the pavement \mathcal{Q}_N as (see [14] for the definition of μ_N, ν_N), taking henceforth $m = 1$,

$$V_N(z^{(N)}) \stackrel{\text{def}}{=} - \sum_{\Delta \in \mathcal{Q}_N} \gamma^{-Nd} \int_{\Delta} \left(\lambda \gamma^{2(d-2)N} X_{\xi}^{(N)4} \right. \\ \left. + \mu_N \gamma^{(d-2)N} X_{\xi}^{(N)2} \right. \\ \left. + \nu_N + f_{\xi} \gamma^{(d-2)/2N} X_{\xi}^{(N)} \right) \frac{d\xi}{|\Delta|} \tag{19}$$

where $\gamma^{(d-2)N}$ is interpreted as N if $d=2$. Hence, if $d=3$ it is

$$V_N(z^{(N)}) \stackrel{\text{def}}{=} - \sum_{\Delta \in \mathcal{Q}_N} \gamma^{-N} \int_{\Delta} (\lambda X_{\xi}^{(N)4} + \bar{\mu}_N X_{\xi}^{(N)2} + \bar{\nu}_N + f_{\xi} \gamma^{-\frac{3}{2}N} X_{\xi}^{(N)}) \frac{d\xi}{|\Delta|} \quad [20]$$

where

$$\begin{aligned} \bar{\mu}_N &\stackrel{\text{def}}{=} (-6\lambda c_N + \lambda^2 N \gamma^{-N} c'_N), \\ \bar{\nu}_N &\stackrel{\text{def}}{=} 3\lambda c_N^2 + \lambda^2 \gamma^{-N} b_N + \lambda^3 N \gamma^{-2N} b'_N \end{aligned}$$

and c_N, c'_N, b_N, b'_N , computable from [15] and [14], admit a limit as $N \rightarrow \infty$. While if $d=2$ it is

$$V_N(z^{(N)}) \stackrel{\text{def}}{=} - \sum_{\Delta \in \mathcal{Q}_N} N^2 \gamma^{-2N} \int_{\Delta} (\lambda X_{\xi}^{(N)4} + \bar{\mu}_N X_{\xi}^{(N)2} + \bar{\nu}_N + f_{\xi} N^{-\frac{3}{2}} X_{\xi}^{(N)}) \frac{d\xi}{|\Delta|} \quad [21]$$

where $\bar{\mu}_N \stackrel{\text{def}}{=} -6\lambda c_N$ and $\bar{\nu}_N = 3\lambda c_N^2$ and c_N , computable from [13], admits a limit as $N \rightarrow \infty$.

The fields $z^{(N)}$ and $X^{(N-1)}$ can be considered constant over boxes $\Delta \in \mathcal{Q}_N$: $z_{\xi}^{(N)} = s_{\Delta}$, $X_{\xi}^{(N-1)} = x_{\Delta}$ for $\xi \in \Delta$ and the s_{Δ} can be considered statistically independent on the scale of the lattice \mathcal{Q}_N .

Therefore, [20] and [21] show that integration over $z^{(N)}$ in the integral defining $Z_N(\Lambda, f)$ is not too different from the computation of a partition function of a lattice continuous spin model in which the “spins” are s_{Δ} and, most important, interact extremely weakly if N is large. In fact, the coupling constants are of order of a power of $|X^{(N-1)}|$ times $O(\gamma^{-N})$ if $d=3$ ($O(N^2 \gamma^{-2N})$ if $d=2$), or of order $O(\gamma^{-N(d+2)/2} \max |f_{\xi}|)$, no matter how large λ and f .

This says that the smallest scale fields are extremely weakly coupled. The fields $X^{(N-1)}$ can be regarded as external fields of size that will be called B_{N-1} , of order 1 or even allowed to grow with a power of N , see [6]. Their presence in V_N does not affect the size of the couplings, as far as the analysis of the integral over $z^{(N)}$ is concerned, because the couplings remain exponentially small in N , see [20] and [21], being at worst multiplied by a power of B_{N-1} , i.e., changed by a factor which is a power of N .

The smallness of the coupling at small scale is a property called “asymptotic freedom.” Once fields and coordinates are “correctly scaled,” the real size of the coupling becomes manifest, that is, it is extremely small and the addends in V_N proportional to the “counter-terms” μ_N, ν_N , which looked

divergent when the fields were not properly scaled, are in fact of the same order or much smaller than the main φ^4 -term.

Therefore, the integration over $z^{(N)}$ can be, heuristically, performed by techniques well established in statistical mechanics (i.e., by straightforward perturbation expansions): at least if the field $X_{\xi}^{(\leq N-1)}$ is smooth and bounded, as prescribed by [6], with $B = B_{N-1}$ growing as a power of N . In this case, denoting symbolically the integration over $z^{(N)}$ by P or by $\langle \dots \rangle$, it can be expected that it should give

$$\int e^{V_N} dP(z^{(N)}) \equiv e^{V_{j;N-1} + \bar{\mathcal{R}}(j,N)|\Lambda|} \quad [22]$$

where $V_{j;N-1}$ is the Taylor expansion of $\log \int e^{V_N} dP(z^{(N)})$ in powers of λ (hence essentially in the very small parameter $\lambda \gamma^{-(4-d)N}$) truncated at order j , that is,

$$\begin{aligned} V_{1;N-1} &= \langle V_N \rangle^{\leq 1} \\ V_{2;N-1} &= \left[\langle V_N \rangle + \frac{\langle (V_N^2) \rangle - \langle V_N \rangle^2}{2!} \right]^{\leq 2} \\ V_{3;N-1} &= \left[\langle V_N \rangle + \frac{\langle (V_N^2) \rangle - \langle V_N \rangle^2}{2!} \right. \\ &\quad \left. + \frac{\langle V_N (\langle V_N^2 \rangle - \langle V_N \rangle^2) \rangle - \langle V_N \rangle (\langle V_N^2 \rangle - \langle V_N \rangle^2)}{3!} \right]^{\leq 3}, \dots \end{aligned} \quad [23]$$

where $[\cdot]^{\leq j}$ denotes truncation to order j in λ , and $\bar{\mathcal{R}}(j, N)$ is a remainder (depending on $\varphi_{\xi}^{(\leq N-1)}$) which can be expected to be estimated, for $d=2, 3$, by

$$\begin{aligned} |\bar{\mathcal{R}}(j, N)| &\leq \mathcal{R}(j, N) \\ &\stackrel{\text{def}}{=} C_j B_N^{4j} (\lambda N^2 \gamma^{-(4-d)N})^{j+1} \gamma^{dN} \end{aligned} \quad [24]$$

for suitable constants C_j , that is, a remainder estimated by the $(j+1)$ th power of the coupling times the number of boxes of scale N in Λ . The relations [22]–[24] result from a naive Taylor expansion (in λ of the $\log \int e^{V_N} dP(z^{(N)})$), taking into account that, in V_N as a function of $z^{(N)}$, the $z^{(N)}$'s appear multiplied by quantities at most of size $\leq \lambda \gamma^{4-d} N^2 B_N^3$, by [20] and [21] if $|X^{(N-1)}| \leq B_{N-1}$. In a statistical mechanics model for a lattice spin system, such a calculation of Z_N would lead to a mean-field equation of state once the remainder was neglected.

The peculiarity of field theory is that a relation like [22] and [24] has to be applied again to $V_{j;N-1}$ to perform the integration over $z^{(N-1)}$ and define $V_{j;N-2}$ and, then, again to $V_{j;N-2} \dots$. Therefore, it will be essential to perform the integral in [22] to an order (in λ) high enough so that the bound $\mathcal{R}(j, N)$ can be

summed over N : this requires (see [24]) an explicit calculation of [23] pushed at least to order $j=1$ if $d=2$ or to order $j=3$ if $d=3$; furthermore it is also necessary to check that the resulting $V_{j;N-1}$ can still be interpreted as low-coupling spin model so that [22] can be iterated with $N-1$ replacing N and then with $N-2$ replacing $N-1, \dots$

The first necessary check towards a proof of the discussed heuristic “expectations” is that, defining recursively $V_{j;b}$ from $V_{j;b+1}$ for $b=N-1, \dots, 1, 0$ by [23] with V_N replaced by $V_{j;b+1}$ and $V_{j;N-1}$ replaced by $V_{j;b}$, the couplings between the variables $z^{(b)}$ do not become “worse” than those discussed in the case $b=N$. Furthermore, the field $\varphi_{\xi}^{(\leq N-1)}$ has a high probability of satisfying [6], but fluctuations are possible: hence the \mathcal{R} -estimate has to be combined with another one dealing with the large fluctuations of $X_{\xi}^{(N-1)}$ which has to be shown to be “not worse.”

For more details, the reader is referred to Gallavotti (1978, 1985) and Benfatto and Gallavotti (1995).

Effective Potentials and Their Scale (In)Dependence

To analyze the first problem mentioned at the end of the previous section, define $V_{j;b}$ by [23] with V_N replaced by $V_{j;b+1}$ for $b=N-1, N-2, \dots, 0$. The quantities $V_{j;b}$, which are called “effective potentials” on scale b (and order j), turn out to be in a natural sense scale independent: this is a consequence of renormalizability, realized by Wilson as a much more general property which can be checked, in the very special cases considered here with $d=2, 3$, at fixed j by induction, and in the super-renormalizable models considered here it requires only an elementary computation of a few Gaussian integrals as the case $j=3$ (or even $j=1$ if $d=2$) is already sufficient for our purposes.

It can also be (more easily) proved for general j by a dimensional argument parallel to the one presented earlier to check finiteness of the renormalized series. The derivation is elementary but it should be stressed that, again, it is possible only because of the special choice of the counter-terms μ_N, ν_N . If $d=3$, the boundedness and smoothness of the fields $\varphi^{(\leq b)}$ and $z^{(b)}$ expressed by the second of [6] and of [10] is essential; while if $d=2$ the smoothness is not necessary.

The structure of $V_{j;b}$ is conveniently expressed in terms of the fields $X_{\xi}^{(b)}$, as a sum of three terms $V_b^{(rel)}$ (standing for “relevant” part), $V_b^{(irr)}$ (standing for “irrelevant” part), and a “field independent” part $E(j, b)|\Lambda|$.

The relevant part in $d=2$ is simply of the form [21] with b replacing N : call it $V_b^{(rel,1)}$. If $d=3$, it is given by [20] with b replacing N plus, for $b < N$, a second “nonlocal” term

$$V_b^{(rel,2)} \stackrel{\text{def}}{=} \frac{4^2 3!}{2! 2!} \lambda^2 \int \left(C_{\eta\eta'}^{(\leq b)3} - C_{\eta\eta'}^{(\leq N)3} \right) \times \left(\varphi_{\eta}^{(\leq b)} - \varphi_{\eta'}^{(\leq b)} \right)^2 d\eta d\eta'$$

which is conveniently expressed in terms of a “nonlocal” field

$$Y_{\eta\eta'}^{(b)} \stackrel{\text{def}}{=} \frac{\varphi_{\eta}^{(\leq b)} - \varphi_{\eta'}^{(\leq b)}}{(\gamma^b |\eta - \eta'|)^{\frac{1}{2}}}$$

as $V_b^{(rel)} = V_b^{(rel,1)} + V_b^{(rel,2)}$ with

$$V_b^{(rel,2)} \stackrel{\text{def}}{=} -\lambda^2 \gamma^{-2b} \sum_{\Delta, \Delta' \in \mathcal{Q}_b} \int_{\Delta \times \Delta'} Y_{\eta\eta'}^{(b)2} A_{\eta\eta'}^{(b)} \times e^{-c' \gamma^b |\eta - \eta'|} \frac{d\eta d\eta'}{|\Delta| |\Delta'|} \tag{25}$$

where

$$0 < a \leq \left(\frac{A_{\eta\eta'}^{(b)}}{(\gamma^b |\eta - \eta'|)^{3-(1/2)}} \right)_N < a'$$

with $a, a', c' > 0$ and the subscript N means that the expression in parenthesis “saturates at scale N ”, i.e., its denominator becomes $\gamma^{(3-(1/2))(b-N)}$ as $|\eta - \eta'| \rightarrow 0$.

The expression [25] is not the full part of the potential $V_{j;b}$ which is of second order in the fields: there are several other contributions which are collected below as “irrelevant.”

It should be stressed that “irrelevant” is a traditional technical term: by no means it should suggest “negligibility.” On the contrary, it could be maintained that the whole purpose of the theory is to study the irrelevant terms. The irrelevant part of the potential can be better designated as the “driven part,” as its behavior is “controlled” by the relevant part: although initially $V_{j;b}$, $b=N$, contains no irrelevant terms, it eventually contains them for $b < N$ and they keep getting generated as b diminishes. Furthermore, the part of the irrelevant terms generated at scale $b_0 \leq N$ becomes very small at scales $b \ll b_0$ so that the irrelevant part of $V_{j;b}$ at small b (e.g., at $b=0$, i.e., on the “physical scale” of the observer) only depend on the relevant terms in a few scales near b .

It also turns out that the Schwinger functions are simply related to the irrelevant terms.

The irrelevant part of the effective potential can be expressed as a finite sum of integrals of

monomials in the fields $X_\xi^{(b)}$ if $d=2$, or in the fields $X_\xi^{(b)}$ and $Y_{\eta\eta'}$ if $d=3$, which can be written as $V_{j;b}^{(irr)}$ given by

$$\int \left(\prod_{k=1}^p X_{\xi_k}^{(b)n_k} \prod_{k'=1}^q Y_{\eta_{k'}\eta'_{k'}}^{(b)n'_{k'}} \right) e^{-\gamma^b c' d(\xi_1, \dots, \eta'_q)} \lambda^n \gamma^{-bt} \times W(\xi_1, \dots, \eta'_q) \prod_{k=1}^p \frac{d\xi_k}{|\Delta_k|} \prod_{k'=1}^q \frac{d\eta_{k'} d\eta'_{k'}}{|\Delta_{k'}^1| |\Delta_{k'}^2|} \quad [26]$$

with the integral extended to products $\Delta_1 \times \dots \times \Delta_p \times \dots \times (\Delta_q^1 \times \Delta_q^2)$ of boxes $\Delta \in \mathcal{Q}_b$, and $d(\xi_1, \dots, \eta'_q)$ is the length of the shortest tree graph that connects all the $p + 2q > 0$ points, the exponents n, t are ≥ 2 , and t is ≥ 3 if $q > 0$; the kernel W depends on all coordinates ξ_1, \dots, η'_q and it is bounded above by $C_j \prod_{k'=1}^q A_{\eta_{k'}\eta'_{k'}}$ for some C_j ; the sums $\sum n_k + \sum n'_{k'}$ cannot exceed $4j$. The test functions f do not appear in [26] because by assumption they are bounded by 1; but W depends on the f 's as well.

The field-independent part is simply the value of $\log Z_N(\Lambda, f)$ computed by the perturbation analysis in the section “**Perturbation theory**” up to order j in λ but using as propagator $(C^{(\leq N)} - C^{(\leq b)})$; thus, $E(j, b)$ is a constant depending on N but uniformly bounded as $N \rightarrow \infty$ (because of the renormalizability proved in the section “**Perturbation theory**”).

If $d=2$, there is no need to introduce the nonlocal fields $Y^{(b)}$ and in [26] one can simply take $q=0$, and the relevant part also can be expressed by omitting the term $V_b^{(rel,2)}$ in [25]; unlike the $d=3$ case, the estimate on the kernels W by an N -independent C_j holds uniformly in b without having to introduce Y . For $d=2$, it will therefore be supposed that $V_b^{(rel,2)} \equiv 0$ in [25] and $q=0$ in [26].

It is not necessary to have more information on the structure of $V_{j;b}$ even though one can find simple graphical rules, closely related to the ones in the section “**Perturbation theory**,” to construct the coefficients W in full detail. The W depend, of course, on b but the uniformity of the bound on W is the only relevant property and in this sense the effective potentials are said to be (almost) “scale independent.”

The above bounds on the irrelevant part can be checked by an elementary direct computation if $j \leq 3$: in spite of its “elementary character,” the uniformity in $b \leq N$ is a result ultimately playing an essential role in the theory together with the dominance of the relevant part over the irrelevant one which, once the fields are properly scaled, is “much smaller” (by a factor of order γ^{-b} , see [26]), at least if b is large.

Remarks

- (i) Checking scale independence for $j=1$ is just checking that $\int P(dz^{(b)}) V_{1;b} = V_{1;b-1}$. Note that

$$V_{1;b} \stackrel{\text{def}}{=} \int_{\Lambda} \lambda \left(\varphi_{\xi}^{(\leq b)4} - 6C_{00}^{(\leq b)} \varphi_{\xi}^{(\leq b)2} + 3C_{00}^{(\leq b)2} \right) d\xi$$

hence, calling $:\varphi_{\xi}^{(\leq b)4}$: the polynomial in the integral (Wick’s monomial of order 4), we have here an elementary Gaussian integral (“martingale property of Wick monomials”). Note the essential role of the counter-terms. For $j > 1$, the computation is similar but it involves higher-order polynomials (up to $4j$) and the distinction between $d=2$ and $d=3$ becomes important.

- (ii) $V_{j;0}$ contains only the field-independent part $E(j, 0)|\Lambda|$ (see above) which is just a number (as there are no fields of scale 0): by the above definitions, it is identical to the perturbative expansion truncated to j th order in λ of $\log Z_N(\Lambda, f)$, well defined as discussed earlier.

Nonperturbative Renormalization: Small Fields

Having introduced the notion of effective potential $V_{j;b}$, of order j and scale b , satisfying the bounds (described after [26]) on the kernels W representing it, the problem is to estimate the remainder in [22] and find its relation with the value [24] given by the heuristic Taylor expansion. Assume $\lambda < 1$ to avoid distinguishing this case from that with $\lambda \geq 1$ which would lead to very similar estimates but to different λ -dependence on some constants.

Define $\chi_B(z^{(b)}) = 1$ if $\|z^{(b)}\|_{\Delta} \leq Bh^2$ for all $\Delta \in \mathcal{Q}_b$, see [8], and 0 otherwise; then the following lemma holds:

Lemma 1 *Let $\|X^{(b)}\|_{\Delta}$ be defined as [8] with z replaced by X and suppose $\|X^{(b)}\|_{\Delta} \leq Bh^4$ for all Δ then, for all $j \geq 1$, it is*

$$\int e^{V_{j;b+1}} \chi_B(z^{(b+1)}) dP(z^{(b+1)}) = e^{V_{j;b} + \mathcal{R}'(j,b+1)|\Lambda|} \quad [27]$$

with, for suitable constants c_-, c'_- ,

$$|\mathcal{R}'_-(j; b+1)| \leq \mathcal{R}_-(j; b+1) \stackrel{\text{def}}{=} \mathcal{R}(j; b+1) + c_- e^{-c'_- B^2(b+1)^2}$$

and $\mathcal{R}(j; b+1)$ given by [24] with $b+1$ in place of N .

Since $Z_N(\Lambda, f) \geq \int e^{V_N} \prod_{b=1}^N \chi_B(z^{(b)}) P(dz^{(b)})$ this immediately gives a lower bound on $E = (1/|\Lambda|) \log Z_N(\Lambda, f)$: in fact if $\chi_B(\|z^{(b)}\|) = 1$ for

$b' = 1, \dots, b$, then $\|X^{(b)}\|_{\Delta} \leq cBb^{4/4}$ for some c so that, by recursive application of Lemma 1, $Z_N(\Lambda, f) \geq e^{V_{i,0} - \sum_{b=1}^N \mathcal{R}_-(j,b)|\Lambda|}$. By the remark at the end of the previous section, given j the lower bound on E just described agrees with the perturbation expansion of $E = (1/|\Lambda|) \log Z_N(\Lambda, f)$ truncated to order j (in λ) up to an error bounded by $\sum_{b=1}^{\infty} \mathcal{R}_-(j, b)$.

Remark The problem solved by Lemma 1 is usually referred to as the small-field problem, to contrast it with the large-field problem discussed later. The proof of the lemma is a simple Taylor expansion in $\lambda\gamma^{-b}$ if $d=3$ or in $\lambda b^2\gamma^{-2b}$ if $d=2$ to order j (in λ). The constraint on $z^{(b+1)}$ makes the integrations over $z^{(b+1)}$, necessary to compute $V_{j;b}$ from $V_{j;b+1}$, not Gaussian. But the tail estimates [9], together with the Markov property of the distribution of $z^{(b)}$ can be used to estimate the difference with respect to the Gaussian unconstrained integrations of $z^{(b+1)}$: and the result is the addition of the small “tail error” changing \mathcal{R} into \mathcal{R}_- in [27]. The estimate of the main part of the remainder \mathcal{R} would be obvious if the fields $z^{(b)}$ were independent on boxes of scale γ^{-b} : they are not independent but they are Markovian and the estimate can be done by taking into account the Markov property.

For more details, the reader is referred to Wilson (1970, 1972), Gallavotti (1978, 1981, 1985), and Benfatto *et al.* (1978).

Nonperturbative Renormalization: Large Fields, Ultraviolet Stability

The small-field estimates are not sufficient to obtain ultraviolet stability: to control the cases in which $|X_{\xi}^{(b)}| > Bb^4$ for some ξ or some b , or $|Y_{\xi\eta}^{(b)}| > Bb^4$ for some $|\xi - \eta| < \gamma^{-b}$, a further idea is necessary and it rests on making use of the assumption that $\lambda > 0$ which, in a sense to be determined, should suppress the contribution to the integral defining $Z_N(\Lambda, f)$ coming from very large values of the field. Assume also $\lambda < 1$ for the same reasons advanced in the section “Effective potentials and their scale (in)dependence.”

Consider first $d=2$. Let \mathcal{D}_N be the “large-field region” where $|X_{\xi}^{(N)}| > BN^4$ and let $V_N(\Lambda/\mathcal{D}_N)$ be the integral defining the potential in [21] extended to the region Λ/\mathcal{D}_N , complement of \mathcal{D}_N . This region is typically very irregular (and random as X itself is random with distribution P_N).

An upper bound on the integral defining $Z_N(\Lambda, f)$ is obtained by simply replacing e^{V_N} by $e^{V_N(\Lambda/\mathcal{D}_N)}$ because in \mathcal{D}_N the first term in the integrand in [21]

is $\leq -\lambda N^2 \gamma^{2N} (BN^4) < 0$ and it overwhelmingly dominates on the remaining terms whose value is bounded by a similar expression with a smaller power of N . Then if $\mathcal{E}^c \stackrel{\text{def}}{=} \Lambda/\mathcal{E}$ denotes the complement in Λ of a set $\mathcal{E} \subset \Lambda$:

Lemma 2 Let $d=2$. Define $V_b(\mathcal{D}_b^c)$ to be given by the expression [22] with the integrals extending over Δ_j/\mathcal{D}_b and define $\mathcal{R}(j, b+1)$ by [24]. Then

$$\int e^{V_{b+1}(\mathcal{D}_{b+1}^c)} dP(z^{(b+1)}) = e^{V_b(\mathcal{D}_b^c) + \overline{\mathcal{R}}_+(j, b+1)|\Lambda|} \quad [28]$$

where $|\overline{\mathcal{R}}_+(j, b+1)| \leq \mathcal{R}_+(j, b+1) \stackrel{\text{def}}{=} \mathcal{R}(j; b+1) + c_+ e^{-c_+ B^2 (b+1)^2}$ with suitable c_+, c'_+ .

Remark Lemma 2 is genuinely not perturbative and making essential use of the positivity of λ . Below the analysis of the proof of the lemma, which consists essentially in its reduction to Lemma 1, is described in detail. It is perhaps the most interesting part and the core of the theory of the proof that truncating the expansion in λ of $(1/|\Lambda|) \log Z_N(\Lambda, f)$ to order j gives as a result an estimate exact to order λ^{j+1} of $(1/|\Lambda|) \log Z_N(\Lambda, f)$.

Let R_N be the cubes $\Delta \in \mathcal{Q}_N$ in which there is at least one point ξ where $|z_{\xi}^{(N)}| \geq BN^2$. By definition, the region $\mathcal{D}_N/\mathcal{D}_{N-1}$ is covered by R_N .

Remark that in the region \mathcal{D}_{N-1}/R_N the field $X^{(N-1)}$ is large but z_N is not large so that $X^{(N)}$ is still very large: this is so because the bounds set to define the regions \mathcal{D} and R are quite different being BN^4 and BN^2 , respectively. Hence, if a point is in \mathcal{D}_{N-1} and not in R_N , then the field $X^{(N)}$ must be of the order $\gg BN^3$. Therefore, by positivity of the $\lambda\varphi_{\xi}^{(\leq N)^4}$ term (which dominates all other terms so that $V^{(N)}(\varphi_{\xi}^{(\leq N)}) < 0$ for $\xi \in \mathcal{D}_N \cup (\mathcal{D}_{N-1}/R_N)$) we can replace $V_N(\mathcal{D}_N^c)$ by $V((\mathcal{D}_N \cup (\mathcal{D}_{N-1}/R_N))^c)$, for the purpose of obtaining an upper bound.

Furthermore, modulo a suitable correction, it is possible to replace $V((\mathcal{D}_N \cup (\mathcal{D}_{N-1}/R_N))^c)$ by $V((\mathcal{D}_{N-1} \cup R_N)^c)$: because the integrand in V_N is bounded below by

$$-b\lambda\gamma^{-2N}N^2$$

if $d=2$ (by $-b\lambda\gamma^{-N}$ if $d=3$), for some b , so that the points in R_N can at most lower $V((\mathcal{D}_N \cup (\mathcal{D}_{N-1}/R_N))^c)$ by $-b\lambda N^2 \gamma^{-(4-d)N} \#(R_N)$ if $\#R_N$ is the number of boxes of \mathcal{Q}_N in R_N and $V(\varphi_{\xi})$ is bounded below by its minimum: thus,

$$V((\mathcal{D}_{N-1} \cup R_N)^c) + b\lambda N^2 \gamma^{(4-d)N} \#(R_N)$$

is an upper bound to $V((\mathcal{D}_N \cup (\mathcal{D}_{N-1}/R_N))^c)$.

In the complement of $\mathcal{D}_{N-1} \cup R_N$, all fields are “small”; if $X^{(N-1)}$ and R_N are fixed this region is not random (as a function of $z^{(N)}$) any more. Therefore,

if $X^{(N-1)}, R_N$ are fixed the integration over $z^{(N)}$, conditioned to having $z^{(N)}$ fixed (and large) in the region R_N , is performed by means of the same argument necessary to prove Lemma 1 (essentially a Taylor expansion in $\lambda\gamma^{-(4-d)N}$). The large size of $z^{(N)}$ in R_N does not affect too much the result because on the boundary of R_N the field $z^{(N)}$ is $\leq BN^2$ (recalling that $z^{(N)}$ is continuous) and since the variable $z^{(N)}$ is Markovian, the boundary effect decays exponentially from the boundary ∂R_N : it adds a quantity that can be shown to be bounded by the number of boxes in R_N on the boundary of R_N , hence by $\#R_N$, times $b'(N-1)^2\gamma^{-(4-d)}(B(N-1)^4)^4$ for some b' .

The result of the integration over $z^{(N)}$ of $e^{V_N((\mathcal{D}_{N-1} \cup \mathcal{D}_{N-1}/R_N)^c)}$ conditioned to the large-field values of $z^{(N)}$ in R_N leads to an upper bound on $\int e^{V_N} P(dz^{(N)})$ as

$$\sum_{R_N} e^{V_{j;N-1}(\mathcal{D}_{N-1}^c) + \mathcal{R}'(j,N)|\Lambda|} \times \prod_{\Delta \in R_N} \left(c e^{-c'(BN^2)^2} e^{+c''\lambda\gamma^{-(4-d)N}N^2(BN^4)^4} \right)^{\#R_N} \quad [29]$$

where c, c', c'' are suitable constants: this is explained as follows.

1. Taylor expansion (in λ) of the integral $e^{V_N((\mathcal{D}_{N-1} \cup R_N)^c) + b\lambda N^2\gamma^{-(4-d)N}\#(R_N)}$ (which, by construction, is an upper bound on $e^{V_N(\mathcal{D}_{N-1}^c)}$ with respect to the field $z^{(N)}$, conditioned to be fixed and large in R_N , would lead to an upper bound as

$$e^{V_{j;N-1}((\mathcal{D}_{N-1} \cup R_N)^c) + \mathcal{R}'(j,N)|\Lambda| + b''\lambda(BN^4)^4\gamma^{(4-d)N}\#(R_N)}$$

with \mathcal{R}' equal to [24] possibly with some C'_j replacing C_j . The second exponential on the RHS of [29] arises partly from the above correction $b''\lambda(BN^4)^4\gamma^{-(4-d)N}\#(R_N)$ and partly from a contribution of similar form explained in (3) below.

2. Integration over the large conditioning fields fixed in R_N is controlled by the second estimate in [9] (the tail estimate): the first factors in parentheses in [29] is the tail estimate just mentioned, i.e., the probability that $z^{(N)}$ is large in the region R_N . The second factor is only partly explained in (1) above.
3. Without further estimates, the bound [29] would contain $V_{j;N-1}((\mathcal{D}_{N-1} \cup R_N)^c)$ rather than $V_{j;N-1}(\mathcal{D}_{N-1}^c)$. Hence, there is the need to change the potential $V_{j;N-1}((\mathcal{D}_{N-1} \cup R_N)^c)$ by “reintroducing” the contribution due to the fields in R_N/\mathcal{D}_{N-1} in order to reconstruct $V_{j;N-1}(\mathcal{D}_{N-1}^c)$. Reintroducing this part of the potential costs a

quantity like $b'\lambda N^2\gamma^{(4-d)N}(BN^4)^4\#(R_N)$ (because the reintroduction occurs in the region R_N/\mathcal{D}_{N-1} which is covered by R_N and in such points the field $X_{\xi}^{(N-1)}$ is *not large*, being bounded by $B(N-1)^4$); so that their contribution to the effective potential is still dominated by the φ^4 -term and therefore by $\gamma^{-(4-d)N}$ times a power of BN^4 times the volume of R_N (in units γ^{-N} , i.e., $\#R_N$). All this is taken care of by suitably fixing c'' .

Note that the sum over R_N of [29] is

$$(1 + c e^{-c'B^2N^4} e^{+c''\lambda\gamma^{-(4-d)N}N^2(BN^4)^4})^{\gamma^{dN}|\Lambda|}$$

(because Λ contains $|\Lambda|\gamma^{dN}$ cubes of \mathcal{Q}_N); hence, it is bounded above by $e^{c_+ e^{-c'_+ B^2N^2}}$ for suitably defined c_+, c'_+ .

The same argument can be repeated for $V_{j;b}(\mathcal{D}_b^c)$ with any b if $V_{j;b}(\mathcal{D}_b^c)$ is defined by the sum over Δ 's in \mathcal{Q}_b of the same integrals as those in [25] and [26] with Δ_j/\mathcal{D}_b replacing Δ_j in the integration domains.

Applying Lemma 1 recursively with $j \geq 1$ (if $d=3$ it would become necessary to take $j \geq 3$), it follows that there exist N -independent upper and lower bounds $E_{\pm}|\Lambda|$ on $\log Z(\Lambda, f)$ of the form $V_{j;0} \pm \sum_{b=1}^{\infty} (\mathcal{R}(j, b) + c_{\pm} e^{-c'_{\pm} B^2 b^2})|\Lambda|$ for $c_{\pm}, c'_{\pm} > 0$ suitably chosen and λ -independent for $\lambda < 1$. By the remark at the end of Sec.6, given j , the bounds just described agree with the perturbation expansion $E(j, 0)|\Lambda| \equiv V_{j;0}$ of $\log Z(\Lambda, f)$ truncated to order j (in λ) up to the remainders $\pm \sum_{b=1}^{\infty} \mathcal{R}_{\pm}(j, b)$. Hence, if B is chosen proportional to $\log_+ \lambda^{-1} \stackrel{\text{def}}{=} \log(e + \lambda^{-1})$, the upper and lower bounds coincide to order j in λ with the value obtained by truncating to order j the perturbative series.

The latter remark is important as it implies not only that the bounds are finite (by the section “Perturbation theory”) but also that the function $(1/|\Lambda|)\log Z(\Lambda, f)$ is not quadratic in f : already to order 1 in λ it is quartic in f (containing a term equal to $-\lambda(\int C_{\xi,0} f_{\xi} d\xi)^4$).

The latter property is important as it excludes that the result is a “Gaussian” generating function. Thus, the outline of the proof of Lemma 2, which together with Lemma 1 forms the core of the analysis of the ultraviolet stability for $d=2$, is completed.

If $d=3$, more care is needed because (very mild) smoothness, like the considered Hölder continuity with exponent 1/4, of z, X is necessary to obtain the key scale independence property discussed in earlier: therefore, the natural measure of the size of $z^{(b)}$ and $X^{(b)}$ in a box $\Delta \in \mathcal{Q}_b$ is no longer the maximum of $|z_{\xi}^{(b)}|$ or of $|X_{\xi}^{(b)}|$. The region \mathcal{D}_b becomes more

involved as it has to consist of the points ξ where $|X_\xi^{(b)}| > Bh^4$ and of the pairs η, η' where

$$|Y_{\eta, \eta'}| \equiv \frac{|X_\eta^{(b)} - X_{\eta'}^{(b)}|}{(\gamma^b |\eta - \eta'|)^{\frac{1}{2}}} > Bh^4$$

i.e., it is not just a subset of Λ .

However, if $d = 3$, the relevant part also contains the negative term $V^{\text{(rel,2)}}$, see [25]: and since it dominates over all other terms which contain a Y -field (because their couplings [25] are smaller by about γ^{-b}), the argument given for $d = 2$ can be adapted to the new situation. Two regions $\mathcal{D}_b^1, \mathcal{D}_b^2$ will be defined: the first consists of all the points ξ where $|X_\xi^{(b)}| > Bh^4$ and the second of all the pairs η, η' where $|Y_{\eta, \eta'}^{(b)}| > Bh^4$. The region R_b will be the collection of all $\Delta \in \mathcal{Q}_b$, where $\|z^{(b)}\|_\Delta > Bh^2$, see [8] with $\tau = 0$. Then $V(\mathcal{D}_b^c)$ will be defined as the sum of the integrals in [25] and [26] with the integrals over ξ_i further restricted to $\xi_i \notin \mathcal{D}_b^1$ and those over the pairs η_i, η'_i are further restricted to $(\eta_i, \eta'_i) \notin \mathcal{D}_b^2$. With the new settings, Lemma 2 can be proved also for $d = 3$ along the same lines as in the $d = 2$ case.

For more details, the reader is referred to Wilson (1970, 1972), Benfatto *et al.* (1978), and Gallavotti (1981).

Ultraviolet Limit, Infrared Behavior, and Other Applications

The results on the ultraviolet stability are nonperturbative, as no assumption is made on the size of λ (the assumption $\lambda < 1$ has been imposed in the last two sections only to obtain simpler expressions for the λ -dependence of various constants): nevertheless the multiscale analysis has allowed us to use perturbative techniques (i.e., the Taylor expansion in Lemmata 1, 2) to find the solution. The latter procedure is the essence of the renormalization group methods: they aim at reducing a difficult multiscale problem to a sequence of simple single-scale problems. Of course, in most cases, it is difficult to implement the approach and the scalar quantum fields in dimensions 2, 3 are among the simplest examples. The analysis of the beta function and of the running couplings, which appear in essentially all renormalization group applications, does not play a role here (or, better, their role is so inessential that it has even been possible to avoid mentioning them). This makes the models somewhat special from the renormalization group viewpoint: the running couplings at length scale b , if introduced, would tend exponentially to 0 as $b \rightarrow \infty$; unlike what happens in the most interesting

renormalization group applications in which they either tend to zero only as powers of b or do not tend to zero at all.

The multiscale analysis method, i.e., the renormalization group method, in a form close to the one discussed here has been applied very often since its introduction in physics and it has led to the solution of several important problems. The following is not an exhaustive list and includes a few open questions.

1. The arguments just discussed imply, with minor extra work that $Z_N(\Lambda, f)$ as $N \rightarrow \infty$ not only admit uniform upper and lower bounds but also that the limit as $N \rightarrow \infty$ actually exists and it is a C^∞ function of λ, f . Its λ and f -derivatives at $\lambda = 0$ and $f = 0$ are given by the formal perturbation calculation. In some cases, it is even possible to show that the formal series for $Z_N(\Lambda, f)$ in powers of λ is Borel summable.
2. The problem of removing the infrared cutoff (i.e., $\Lambda \rightarrow \infty$) is in a sense more a problem of statistical mechanics. In fact, it can be solved for $d = 2, 3$ by a typical technique used in statistical mechanics, the “cluster expansion.” This is not intended to mean that it is technically an easy task: understanding its connection with the low-density expansions and the possibility of using such techniques has been a major achievement that is not discussed here.
3. The third problem mentioned in the introduction, that is, checking the axioms so that the theory could be interpreted as a quantum field theory is a difficult problem which required important efforts to control and which is not analyzed here. An introduction to it can be its analysis in the $d = 2$ case.
4. Also the problem of keeping the ultraviolet cutoff and removing the infrared cutoff while the parameter m^2 in the propagator approaches 0 is a very interesting problem related to many questions in statistical mechanics at the critical point.
5. Field theory methods can be applied to various statistical mechanics problems away from criticality: particularly interesting is the theory of the neutral Coulomb gas and of the dipole gas in two dimensions.
6. The methods can be applied to Fermi systems in field theory as well as in equilibrium statistical mechanics. The understanding of the ground state in not exactly soluble models of spinless fermions in one dimension at small coupling is one of the results. And via the transfer matrix theory it has led to the understanding of nontrivial critical behavior in two-dimensional models that are not exactly soluble (like Ising next-nearest-neighbor or Ashkin–Teller model). Fermi systems are of particular interest also because in their analysis the large-fields problem is absent, but this great

technical advantage is somewhat offset by the anticommutation properties of the fermionic fields, which do not allow us to employ probabilistic techniques in the estimates.

7. An outstanding open problem is whether the scalar φ^4 -theory is possible and nontrivial in dimension $d=4$: this is a case of a renormalizable not asymptotically free theory. The conjecture that many support is that the theory is necessarily trivial (i.e., the function $Z_N(\Lambda, f)$ becomes necessarily a Gaussian in the limit $N \rightarrow \infty$). One of the main problems is the choice of the ultraviolet cut-off; unlike the $d=2, 3$ cases in which the choice is a matter of convenience it does not seem that the issue of triviality can be settled without a careful analysis of the choice and of the role of the ultraviolet cut-off.
8. Very interesting problems can be found in the study of highly symmetric quantum fields: gauge invariance presents serious difficulties to be studied (rigorously or even heuristically) because in its naive forms it is incompatible with regularizations. Rigorous treatments have been in some cases possible and in few cases it has been shown that the naive treatment is not only not rigorous but it leads to incorrect results.
9. In connection with item (8) an outstanding problem is to understand relativistic pure gauge Higgs fields in dimension $d=4$: the latter have been shown to be ultraviolet stable but the result has not been followed by the study of the infrared limit.
10. The classical gauge theory problem is quantum electrodynamics, QED, in dimension 4: it is a renormalizable theory (taking into account gauge invariance) and its perturbative series truncated after the first few orders give results that can be directly confronted with experience, giving very accurate predictions. Nevertheless, the model is widely believed to be incomplete: in the sense that, if treated rigorously, the result would be a field describing free noninteracting assemblies of photons and electrons. It is believed that QED can make sense only if embedded in a model with more fields, representing other particles (e.g., the standard model), which would influence the behavior of the electromagnetic field by providing an effective ultraviolet cutoff high enough for not altering the predictions on the observations on the time and energy scales on which present (and, possibly, future over a long time span) experiments are performed. In dimension $d=3$, QED is super-renormalizable, once the gauge symmetry is properly taken into account, and it can be studied with the techniques described above for the scalar fields in the corresponding dimension.

In general, constructive quantum field theory seems to be in a deep crisis: the few solutions that have been found concern very special problems and are very demanding technically; the results obtained have often not been considered to contribute appreciably to any “progress.” And many consider that the work dedicated to the subject is not worth the results that one can even hope to obtain. Therefore, in recent years, attempts have been made to follow other paths: an attitude that in the past usually did not lead, in general to great achievements but that is always tempting and worth pursuing because the rare major progresses made in physics resulted precisely by such changes of attitude, leaving aside developments requiring work which was too technical and possibly hopeless: just to mention an important case, one can recall quantum mechanics which disposed of all attempts at understanding the observed atomic levels quantization on the basis of refined developments of classical electromagnetism.

For more details, the reader is referred to Nelson (1966), Guerra (1972), Glimm *et al.* (1973), Glimm and Jaffe (1981), Simon (1974), Benfatto *et al.* (1978, 2003), Aizenman (1982), Gawedzky and Kupiainen (1983, 1985a, b), Balaban (1983), and Giuliani and Mastropietro (2005).

See also: Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Euclidean Field Theory; Integrability and Quantum Field Theory; Perturbation Theory and its Techniques; Quantum Field Theory: A Brief Introduction; Scattering, Asymptotic Completeness and Bound States.

Further Reading

- Aizenman M (1982) Geometric analysis of φ^4 -fields and Ising models. *Communications in Mathematical Physics* 86: 1–48.
- Balaban T (1983) (Higgs) $_{3,2}$ quantum fields in a finite volume. III. Renormalization. *Communications in Mathematical Physics* 88: 411–445.
- Benfatto G, Cassandro M, Gallavotti G *et al.* (1978) Some probabilistic techniques in field theory. *Communications in Mathematical Physics* 59: 143–166.
- Benfatto G, Cassandro M, Gallavotti G *et al.* (1980) Ultraviolet stability in Euclidean scalar field theories. *Communications in Mathematical Physics* 71: 95–130.
- Benfatto G and Gallavotti G (1995) *Renormalization Group*, pp. 1–143. Princeton: Princeton University Press.
- Benfatto G, Giuliani A, and Mastropietro V (2003) Low temperature analysis of two dimensional Fermi systems with symmetric Fermi surface. *Annales Henry Poincaré* 4: 137–193.
- De Calan C and Rivasseau V (1981) Local existence of the Borel transform in euclidean ϕ_4^4 . *Communications in Mathematical Physics* 82: 69–100.
- Fröhlich J (1982) On the triviality of $\lambda\phi_4^4$ theories and the approach to the critical point in $d \geq 4$ dimensions. *Nuclear Physics B* 200: 281–296.

- Gallavotti G (1978) Some aspects of renormalization problems in statistical mechanics. *Memorie dell' Accademia dei Lincei* 15: 23–59.
- Gallavotti G (1981) Elliptic operators and Gaussian processes. In: *Aspects Statistiques et Aspects Physiques des Processus Gaussiens*, pp. 349–360. Colloques Internat. C.N.R.S, St. Flour. Publications du CNRS, Paris.
- Gallavotti G (1985) Renormalization theory and ultraviolet stability via renormalization group methods. *Reviews of Modern Physics* 57: 471–569.
- Gawedzky K and Kupiainen A (1983) Block spin renormalization group for dipole gas and $(\partial\phi)^4$. *Annals of Physics* 147: 198–243.
- Gawedzky K and Kupiainen A (1985a) Gross–Neveu model through convergent perturbation expansion. *Communications in Mathematical Physics* 102: 1–30.
- Gawedzky K and Kupiainen A (1985b) Massless lattice ϕ_4^4 theory: rigorous control of a renormalizable asymptotically free model. *Communications in Mathematical Physics* 99: 197–252.
- Giuliani A and Mastropietro V (2005) Anomalous universality in the anisotropic Ashkin–Teller model. *Communications in Mathematical Physics* 256: 681–735.
- Glimm J, Jaffe A, and Spencer T (1973) Velo G and Wightman A (eds.) *Constructive Field theory*, Lecture Notes in Physics, vol. 25, pp. 132–242. New York: Springer.
- Glimm J and Jaffe A (1981) *Quantum Physics*. Springer.
- Guerra F (1972) Uniqueness of the vacuum energy density and Van Hove phenomena in the infinite volume limit for two-dimensional self-coupled Bose fields. *Physical Review Letters* 28: 1213–1215.
- Hepp K (1966) *Théorie de la renormalization*. Lecture Notes in Physics, vol. 2. Heidelberg: Springer.
- Nelson E (1966) A quartic interaction in two dimensions. In: Goodman R and Segal I (eds.) *Mathematical Theory of Elementary Particles*, pp. 69–73. Cambridge: M.I.T.
- Osterwalder K and Schrader R (1973) Axioms for Euclidean Green's functions. *Communications in Mathematical Physics* 31: 83–112.
- Simon B (1974) *The $P(\varphi)_2$ Euclidean (Quantum) Field Theory*. Princeton: Princeton University Press.
- Streater RF and Wightman AS (1964) *PCT, Spin, Statistics and All That*. Benjamin-Cummings (reprinted Princeton University Press, 2000).
- Wightman AS and Gårding L (1965) Fields as operator-valued distributions in relativistic quantum theory. *Arkiv för Fysik* 28: 129–189.
- Wilson KG (1970) Model of coupling constant renormalization. *Physical Review D* 2: 1438–1472.
- Wilson KG (1972) Renormalization of a scalar field in strong coupling. *Physical Review D* 6: 419–426.

Contact Manifolds

J B Etnyre, University of Pennsylvania, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Contact geometry has been seen to underly many physical phenomena and is related to many other mathematical structures. Contact structures first appeared in the work of Sophus Lie on partial differential equations. They reappeared in Gibbs' work on thermodynamics, Huygens' work on geometric optics, and in Hamiltonian dynamics. More recently, contact structures have been seen to have relations with fluid mechanics, Riemannian geometry, and low-dimensional topology, and these structures provide an interesting class of subelliptic operators.

After summarizing the basic definitions, examples, and facts concerning contact geometry, this article discusses the connections between contact geometry and symplectic geometry, Riemannian geometry, complex geometry, analysis, and dynamics. The article ends by discussing two of the most-studied connections with physics: Hamiltonian dynamics and geometric optics. References for other important topics in contact geometry

(e.g., thermodynamics, fluid dynamics, holomorphic curves, and open book decompositions) are provided in the “Further reading” section.

Basic Definitions and Examples

A hyperplane field ξ on a manifold M is a codimension-1 sub-bundle of the tangent bundle TM . Locally, a hyperplane field can always be described as the kernel of a 1-form. In other words, for every point in M there is a neighborhood U and a 1-form α defined on U such that the kernel of the linear map $\alpha_x: T_x M \rightarrow \mathbb{R}$ is ξ_x for all x in U . The form α is called a local defining form for ξ . A contact structure on a $(2n+1)$ -dimensional manifold M is a “maximally nonintegrable hyperplane field” ξ . The hyperplane field ξ is maximally nonintegrable if for any (and hence every) locally defining 1-form α for ξ the following equation holds:

$$\alpha \wedge (d\alpha)^n \neq 0 \quad [1]$$

(this means that the form is, pointwise, never equal to 0). Geometrically, the nonintegrability of ξ means that no hypersurface in M can be tangent to ξ along an open subset of the hypersurface. Intuitively, this means that the hyperplanes “twist too much” to be tangent to hypersurfaces (Figure 1). The pair (M, ξ)

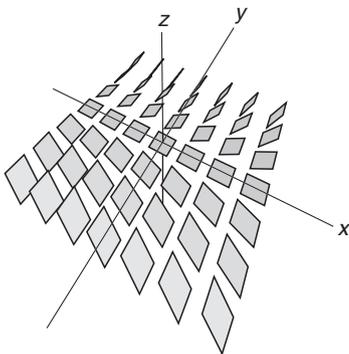


Figure 1 The standard contact structure on \mathbb{R}^3 given as the kernel of $dz - y dx$. Courtesy of Stephan Schönenberger.

is called a contact manifold and any locally defining form α for ξ is called a contact form for ξ .

Example 1 The most basic example of a contact structure can be seen on \mathbb{R}^{2n+1} as the kernel of the 1-form $\alpha = dz - \sum_{i=1}^n y_i dx_i$, where the coordinates on \mathbb{R}^{2n+1} are $(x_1, y_1, \dots, x_n, y_n, z)$. This example is shown in **Figure 1** when $n = 1$.

Example 2 Recall that on the cotangent space of any n -manifold M , there is a canonical 1-form λ , called the Liouville form. If (q_1, \dots, q_n) are local coordinates on M , then any 1-form can be expressed as $\sum_{i=1}^n p_i dq_i$, so $(q_1, p_1, \dots, q_n, p_n)$ are local coordinates on T^*M . In these coordinates,

$$\lambda = \sum_{i=1}^n p_i \pi^* dq_i \tag{2}$$

where $\pi: T^*M \rightarrow M$ is the natural projection map. The 1-jet space of M is the manifold $J^1(M) = T^*M \times \mathbb{R}$ and can be considered as a bundle over M . The 1-jet space has a natural contact structure given as the kernel of $\alpha = dz - \lambda$, where z is the coordinate on \mathbb{R} . Note that if $M = \mathbb{R}^n$ then we recover the previous example.

Example 3 The (oriented) projectivized cotangent space of a manifold M is the set P^*M of nonzero covectors in T^*M where two covectors are identified if they differ by a positive real number, that is,

$$P^*M = (T^*M \setminus \{0\})/\mathbb{R}_+ \tag{3}$$

where $\{0\}$ is the zero section of T^*M and \mathbb{R}_+ denotes the positive real numbers. If M has a metric then P^*M can be easily identified with the space of unit covectors. Considering P^*M as unit covectors, we can restrict the canonical 1-form λ to P^*M to get a 1-form α whose kernel defines a contact structure ξ on P^*M . (Although there is no canonical contact form on P^*M , the contact structure ξ is still well defined.) Note that if

M is compact then so is P^*M ; so this gives examples of contact structures on compact manifolds.

If α and α' are two locally defining 1-forms for ξ , then there is a nonzero function f such that $\alpha' = f\alpha$. Thus, $\alpha' \wedge (d\alpha')^n = f^{n+1} \alpha \wedge (d\alpha)^n$ is a nonzero top dimensional form on M and if n is odd then the orientation defined by the local defining form is independent of the actual form. Hence, when n is odd, a contact structure defines an orientation on M (this is independent of whether or not ξ is orientable!). If M had a preassigned orientation (and n is odd), then the contact structure is called “positive” if it induces the given orientation and “negative” otherwise. One should be careful when reading the literature, as some authors build positive into their definition of contact structure, especially when $n = 1$. If there is a globally defined 1-form α whose kernel defines ξ , then ξ is called transversally orientable or co-orientable. This is equivalent to the bundle ξ being orientable when n is odd or when n is even and M is orientable. In this article the discussion is restricted to transversely orientable contact structures.

Suppose that α is a contact form for ξ , then **eqn [1]** implies that $d\alpha|_\xi$ is a symplectic form on ξ . This is one sense in which a contact structure is like an odd-dimensional analog of a symplectic structure.

A submanifold L of a contact manifold (M, ξ) is called Legendrian if $\dim M = 2 \dim L + 1$ and $T_p L \subset \xi_p$.

Example 4 A fiber in the unit cotangent bundle with the contact structure from **Example 3** is a Legendrian sphere.

Example 5 Let $f: M \rightarrow \mathbb{R}$ be a function. Then $j_1(f)(q) = (q, df_q, f(q))$ is a section of the 1-jet space $J^1(M)$ of M ; it is called the 1-jet of f . If s is any section of the 1-jet space, then it is Legendrian if and only if it is the 1-jet of a function.

This observation is the basis for Lie’s study of partial differential equations. More specifically, a first-order partial differential equation on M can be considered as giving an algebraic equation on $J^1(M)$. Then, a section of $J^1(M)$ satisfying this algebraic equation corresponds to the 1-jet of a solution to the original partial differential equation if and only if it is Legendrian.

Recently, Legendrian submanifolds have been much studied. There are various classification results in three dimensions and several striking existence results in higher dimensions.

Local Theory

The natural equivalence between contact structures is contactomorphism. Two contact structures ξ_0 and

ξ_1 on manifolds M_0 and M_1 , respectively, are contactomorphic if there is a diffeomorphism $f: M_0 \rightarrow M_1$ such that $f_*(\xi_0) = \xi_1$. All contact structures are locally contactomorphic. In particular, we have the following theorem.

Theorem 1 (Darboux’s Theorem). *Suppose ξ_i is a contact structure on the manifold $M_i, i = 0, 1$, and M_0 and M_1 have the same dimension. Given any points p_0 and p_1 in M_0 and M_1 , respectively, there are neighborhoods N_i of p_i in M_i and a contactomorphism from $(N_0, \xi_0|_{N_0})$ to $(N_1, \xi_1|_{N_1})$. Moreover, if α_i is a contact form for ξ_i near p_i , then the contactomorphism can be chosen to pull α_1 back to α_0 .*

Thus, locally all contact structures (and contact forms!) look like the one given in Example 1 above.

Furthermore, contact structures are “local in time.” That is, compact deformations of contact structures do not produce new contact structures.

Theorem 2 (Gray’s theorem). *Let M be an oriented $(2n + 1)$ -dimensional manifold and $\xi_t, t \in (0, 1)$, a family of contact structures on M that agree off of some compact subset of M . Then there is a family of diffeomorphisms $\phi_t: M \rightarrow M$ such that $(\phi_t)_*\xi_t = \xi_0$.*

In particular, on a compact manifold, all deformations of contact structures come from diffeomorphisms of the underlying manifold. The theorem is not true if the contact structures do not agree off of a compact set. For example, there is a one-parameter family of noncontactomorphic contact structures on $S^1 \times \mathbb{R}^2$.

Existence and Classification

The existence of contact structures on closed odd-dimensional manifolds is quite difficult. However, Gromov has shown that contact structures on open manifolds obey an h-principle. To explain this, we note that if (M^{2n+1}, ξ) is a co-oriented contact manifold then the tangent bundle of M can be written as $\xi \oplus \mathbb{R}$ and thus the structure group of TM can be reduced to $U(n)$ (since ξ has a conformal symplectic structure on it). Such a reduction of the structure group is called an almost contact structure on M . Clearly, a contact structure on M induces an almost contact structure. If M is an open manifold, Gromov proved that the inclusion of the space of co-oriented contact structures on M into the space of almost contact structures on M is a weak homotopy equivalence. In particular, if an open manifold meets the necessary algebraic condition for the existence of an almost contact structure, then the manifold has a co-oriented contact structure.

Lutz and Martinet proved a similar, but weaker, result for oriented closed 3-manifolds. More specifically, every closed oriented 3-manifold admits a co-oriented contact structure and in fact has at least one for every homotopy class of plane field. There has been much progress on classifying contact structures on 3-manifolds and here an interesting dichotomy has appeared. Contact structures break into one of two types: tight or overtwisted. Overtwisted contact structures obey an h-principle and are in general easy to understand. Tight contact structures have a more subtle, geometric nature. In higher dimensions there is much less known about the existence (or classification) of contact structures.

Relations with Symplectic Geometry

Let (X, ω) be a symplectic manifold. A vector field v satisfying

$$L_v\omega = \omega \tag{4}$$

(where $L_v\omega$ is the Lie derivative of ω in the direction of v) is called a symplectic dilation. A compact hypersurface M in (X, ω) is said to have “contact type” if there exists a symplectic dilation v in a neighborhood of M that is transverse to M . Given a hypersurface M in (X, ω) , the characteristic line field LM in the tangent bundle of M is the symplectic complement of TM in TX . (Since M is codimension 1, it is coisotropic; thus, the symplectic complement lies in TM and is one dimensional.)

Theorem 3 *Let M be a compact hypersurface in a symplectic manifold (X, ω) and denote the inclusion map $i: M \rightarrow X$. Then M has contact type if and only if there exists a 1-form α on M such that $d\alpha = i^*\omega$ and the form α is never zero on the characteristic line field.*

If M is a hypersurface of contact type, then the 1-form α is obtained by contracting the symplectic dilation v into the symplectic form: $\alpha = \iota_v\omega$. It is easy to verify that the 1-form α is a contact form on M . Thus, a hypersurface of contact type in a symplectic manifold inherits a co-oriented contact structure.

Given a co-orientable contact manifold (M, ξ) , its symplectization $\text{Symp}(M, \xi) = (X, \omega)$ is constructed as follows. The manifold $X = M \times (0, \infty)$, and given a global contact form α for ξ the symplectic form is $\omega = d(t\alpha)$, where t is the coordinate on \mathbb{R} . (The symplectization is also equivalently defined as $(M \times \mathbb{R}, d(e^t\alpha))$.)

Example 6 The symplectization of the standard contact structure on the unit cotangent bundle

(see Example 3) is the standard symplectic structure on the complement of the zero section in the cotangent bundle.

The symplectization is independent of the choice of contact form α . To see this, fix a co-orientation for ξ and note the manifold X which can be identified (in many ways) with the sub-bundle of T^*M whose fiber over $x \in M$ is

$$\{\beta \in T_x^*M : \beta(\xi_x) = 0 \text{ and } \beta > 0 \text{ on vectors positively transverse to } \xi_x\} \quad [5]$$

and restricting $d\lambda$ to this subspace yields a symplectic form ω , where λ is the Liouville form on T^*M defined in Example 2. A choice of contact form α fixes an identification of X with the sub-bundle of T^*M under which $d(t\alpha)$ is taken to $d\lambda$.

The vector field $v = \partial/\partial t$ on (X, ω) is a symplectic dilation that is transverse to $M \times \{1\} \subset X$. Clearly, $\iota_v \omega|_{M \times \{1\}} = \alpha$. Thus, we see that any co-orientable contact manifold can be realized as a hypersurface of contact type in a symplectic manifold. In summary, we have the following theorem.

Theorem 4 *If (M, ξ) is a co-oriented contact manifold, then there is a symplectic manifold $\text{Symp}(M, \xi)$ in which M sits as a hypersurface of contact type. Moreover, any contact form α for ξ gives an embedding of M into $\text{Symp}(M, \xi)$ that realizes M as a hypersurface of contact type.*

We also note that all the hypersurfaces of contact type in (X, ω) look locally, in X , like a contact manifold sitting inside its symplectification.

Theorem 5 *Given a compact hypersurface M of contact type in a symplectic manifold (X, ω) with the symplectic dilation given by v , there is a neighborhood of M in X symplectomorphic to a neighborhood of $M \times \{1\}$ in $\text{Symp}(M, \xi)$ where the symplectization is identified with $M \times (0, \infty)$ using the contact form $\alpha = \iota_v \omega|_M$ and $\xi = \ker \alpha$.*

The Reeb Vector Field and Riemannian Geometry

Let (M, ξ) be a contact manifold. Associated to a contact form α for ξ is the Reeb vector field v_α . This is the unique vector field satisfying

$$\iota_{v_\alpha} \alpha = 1 \quad \text{and} \quad \iota_{v_\alpha} d\alpha = 0 \quad [6]$$

One may readily check that v_α is transverse to the contact hyperplanes and the flow of v_α preserves ξ (in fact, it preserves α). These two conditions characterize Reeb vector fields; that is, a vector field v is the Reeb vector field for some contact form

for ξ if and only if it is transverse to ξ and its flow preserves ξ .

The fundamental question concerning Reeb vector fields asks if its flow has a (contractible) periodic orbit. A paraphrasing of the Weinstein conjecture asserts a positive answer to this question. Most progress on this conjecture has been made in dimension 3 where H Hofer has proved the existence of periodic orbits for all Reeb fields on S^3 and on 3-manifolds with essential spheres (i.e., embedded S^2 's that do not bound a 3-ball in the manifold). Relations with Hamiltonian dynamics are discussed below.

Recall, from Example 3, that a Riemannian metric g on a manifold M provides an identification of the (oriented) projectivized cotangent bundle P^*M with the unit cotangent bundle. Considered as a subset of T^*M , P^*M inherits not only a contact structure but also a contact form α (by restricting the Liouville form). Let v_α be the associated Reeb vector field. The metric g also provides an identification of the tangent and cotangent bundles of M . Thus, P^*M may be considered as the unit tangent bundle of M . Let w_g be the vector field on the unit tangent bundle generating the geodesic flow on M .

Theorem 6 *The Reeb vector field v_α is identified with geodesic flow field w_g when P^*M is identified with the unit tangent space using the metric g .*

Relations with Complex Geometry and Analysis

Let X be a complex manifold with boundary and denote the induced complex structure on TX by J . The complex tangencies ξ to $M = \partial X$ are described by the equation $d\phi \circ J = 0$, where ϕ is a function defined in a neighborhood of the boundary such that 0 is a regular value and $\phi^{-1}(0) = M$. The form $L(v, w) = -d(d\phi \circ J)(v, Jw)$, for $v, w \in \xi$, is called the Levi form, and when $L(v, w)$ is positive (negative) definite, then X is said to have strictly pseudoconvex (pseudoconcave) boundary. The hyperplane field ξ will be a contact structure if and only if $d(d\phi \circ J)$ is a nondegenerate 2-form on ξ (if and only if $L(v, w)$ is definite). A well-studied source of examples comes from Stein manifolds.

Example 7 Let X be a complex manifold and again let J denote the induced complex structure on TX . From a function $\phi : X \rightarrow \mathbb{R}$, we can define a 2-form $\omega = -d(d\phi \circ J)$ and a symmetric form $g(v, w) = \omega(v, Jw)$. If this symmetric form is positive definite, the function ϕ is called “strictly plurisubharmonic.” The manifold X is a Stein manifold if X

admits a proper strictly plurisubharmonic function $\phi : X \rightarrow \mathbb{R}$. An important result says that X is Stein if and only if it can be realized as a closed complex submanifold of C^n . Clearly any noncritical level set of ϕ gives a contact manifold.

Contact manifolds also give rise to an interesting class of differential operators. Specifically, a contact structure ξ on M defines a symbol-filtered algebra of pseudodifferential operators $\Psi_\xi^*(M)$, called the “Heisenberg calculus.” Operators in this algebra are modeled on smooth families of convolution operators on the Heisenberg group. An important class of operators of this type are the “sum-of-squares” operators. Locally, the highest-order part of such an operator takes the form

$$L = \sum_{j=1}^{2n} v_j^2 + ia v_\alpha \tag{7}$$

where $\{v_1, \dots, v_{2n}\}$ is a local framing for the contact field and v_α is a Reeb vector field. This operator belongs to $\Psi_\xi^2(M)$ and is subelliptic for a outside a discrete set.

Hamiltonian Dynamics

Given a symplectic manifold (X, ω) , a function $H : X \rightarrow \mathbb{R}$ will be called a Hamiltonian. (Only autonomous Hamiltonians are discussed here.) The unique vector field satisfying

$$\iota_{v_H} \omega = -dH$$

is called the Hamiltonian vector field associated to H . Many problems in classical mechanics can be formulated in terms of studying the flow of v_H for various H .

Example 8 If $(X, \omega) = (\mathbb{R}^{2n}, d\lambda)$, where λ is from Example 2, then the flow of the Hamiltonian vector field is given by

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}$$

A standard fact says that the flow of v_H preserves the level sets of H .

Theorem 7 *If M is a level set of H corresponding to a regular value and M is a hypersurface of contact type, then the trajectories of v_H and of the Reeb vector field (associated to M in Theorem 3) agree.*

Thus under suitable hypothesis, Hamiltonian dynamics is a reparametrization of Reeb dynamics. In particular, searching for periodic orbits in such a Hamiltonian system is equivalent to searching for periodic orbits in a Reeb flow. Thus in this context,

Weinstein’s conjecture asserts a positive answer to the questions: Does the Hamiltonian flow along a regular level set of contact type have a periodic orbit? Viterbo proved that the answer was yes if the hypersurface is compact and in $(\mathbb{R}^{2n}, \omega = d\alpha)$. Other progress has been made by studying Reeb dynamics.

Geometric Optics

In this section, we study the propagation of light (or various other disturbances) in a medium (for the moment, we do not specify the properties of this medium). The medium will be given by a three-dimensional manifold M . Given a point p in M and $t > 0$, let $I_p(t)$ be the set of all points to which light can travel in time $\leq t$. The wave front of p at time t is the boundary of this set and is denoted as $\Phi_p(t) = \partial I_p(t)$.

Theorem 8 (Huygens’ principle). *$\Phi_p(t + t')$ is the envelope of the wave fronts $\Phi_q(t')$ for all $q \in \Phi_p(t)$.*

This is best understood in terms of contact geometry. Let $\pi : (T^*M \setminus \{0\}) \rightarrow P^*M$ be the natural projection (see Example 3) and let S be any smooth sub-bundle of $T^*M \setminus \{0\}$ that is transverse to the radial vector field in each fiber and for which $\pi|_S : S \rightarrow P^*M$ is a diffeomorphism. The restriction of the Liouville form to S gives a contact form α and a corresponding Reeb vector field v . Given a subset F of M with a well-defined tangent space at every point set

$$L_F = \{p \in S : \pi(p) \in F \text{ and } p(w) = 0 \text{ for all } w \in T_{\pi(p)}F\} \tag{8}$$

The set L_F is a Legendrian submanifold of S and is called the “Legendrian lift” of F . If L is a generic Legendrian submanifold in S , then $\pi(L)$ is called the front projection of L and $L_{\pi(L)} = L$. Given a Legendrian submanifold L , let $\Psi_t(L)$ be the Legendrian submanifold obtained from L by flowing along v for time t .

Example 9 Given a metric g on M , Fermat’s principle says that light travels along geodesics. Thus, if S is the unit cotangent bundle, then using g to identify the geodesic flow with the Reeb flow one sees that light will travel along trajectories of the Reeb vector field. Given a point p in M , the Legendrian submanifold L_p is a sphere sitting in T_p^*M . The Huygens principle follows from the observation that $\Phi_p(t) = \pi(\Psi_t(L_p))$.

Using the more general S discussed above, one can generalize this example to light traveling in a medium that is nonhomogeneous (i.e., the speed differs from point to point in M) and anisotropic (i.e., the speed differs depending on the direction of travel).

See also: Hamiltonian Fluid Dynamics; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Minimax Principle in the Calculus of Variations.

Further Reading

- Aebischer B, Borer M, Kälin M, Leuenberger Ch, and Reimann HM (1994) *Symplectic Geometry*, Progress in Mathematics, vol. 124. Basel: Birkhäuser.
- Arnol'd VI (1989) *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics, vol. 60, xvi+516, pp. 163–179. New York: Springer.
- Arnol'd VI (1990) *Contact Geometry: The Geometrical Method of Gibbs's Thermodynamics*, Proceedings of the Gibbs Symposium. (New Haven, CT, 1989), pp. 163–179. Providence, RI: American Mathematical Society.
- Beals R and Greiner P (1988) Calculus on Heisenberg manifolds. *Annals of Mathematics Studies* 119.
- Eliashberg Y, Givental A, and Hofer H (2000) *Introduction to Symplectic Field Theory*, GAFA 2000 (Tel Aviv, 1999), *Geom. Funct. Anal.* 2000, Special Volume, Part II, pp. 560–673.
- Etnyre J. Legendrian and transversal knots. *Handbook of Knot Theory* (in press).
- Etnyre J (1998) *Symplectic Convexity in Low-Dimensional Topology*, Symplectic, Contact and Low-Dimensional Topology (Athens, GA, 1996), *Topology Appl.*, vol. 88, No. 1–2, pp. 3–25.
- Etnyre J and Ng L (2003) *Problems in Low Dimensional Contact Topology*, Topology and Geometry of Manifolds (Athens, GA, 2001), pp. 337–357, *Proc. Sympos. Pure Math.*, vol. 71. Providence, RI: American Mathematical Society.
- Geiges H Contact geometry. *Handbook of Differential Geometry*, vol. 2 (in press).
- Geiges H (2001a) *Contact Topology in Dimension Greater than Three*, European Congress of Mathematics, vol. II (Barcelona, 2000), Progress in Mathematics, vol. 202, pp. 535–545. Basel: Birkhäuser.
- Geiges H (2001b) A brief history of contact geometry and topology. *Expositiones Mathematicae* 19(1): 25–53.
- Ghrist R and Komendarczyk R (2001) Topological features of inviscid flows. *An Introduction to the Geometry and Topology of Fluid Flows* (Cambridge, 2000), 183–201, NATO Sci. Ser. II Math. Phys. Chem., vol. 47. Dordrecht: Kluwer Academic.
- Giroux E (2002) *Géométrie de contact: de la dimension trois vers les dimensions supérieures*, Proceedings of the International Congress of Mathematicians, vol. II (Beijing, 2002), pp. 405–414. Beijing: Higher Ed. Press.
- Hofer H and Zehnder E (1994) *Symplectic Invariants and Hamiltonian Dynamics*, Birkhäuser Advanced Texts: Basler Lehrbücher, pp. xiv+341. Basel: Birkhäuser.
- Taylor ME (1984) *Noncommutative Microlocal Analysis, Part I*, *Mem Amer. Math. Soc.*, 52, no. 313. American Mathematical Society.

Control Problems in Mathematical Physics

B Piccoli, Istituto per le Applicazioni del Calcolo, Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Control Theory is an interdisciplinary research area, bridging mathematics and engineering, dealing with physical systems which can be “controlled,” that is, whose evolution can be influenced by some external agent. A general model can be written as

$$y(t) = A(t, y(0), u(\cdot)) \quad [1]$$

where y describes the state variables, $y(0)$ the initial condition, and $u(\cdot)$ the control function. Thus, eqn [1] means that the state at time t depends on the initial condition but also on some parameters u which can be chosen as function of time. To be precise, there are some control problems which are not of evolutionary type; however, in this presentation we restrict ourselves to this case.

One has to distinguish among the control set U where the control function can take values: $u(t) \in U$, and the space of control functions, \mathcal{U} , to which each control function should belong: $u(\cdot) \in \mathcal{U}$. Thus, for example, we may have $U = R^m$ and $\mathcal{U} = L^\infty([0, T], R^m)$.

There are various problems one can formulate regarding systems of type [1], among which:

Controllability Given any two states y_0 and y_1 determine a control function $u(\cdot)$ such that for some time $t > 0$ we have $y_1 = A(t, y_0, u(\cdot))$.

Optimal control Consider a cost function $J(y(\cdot), u(\cdot))$ depending both on the evolutions of y and u and determine a control function $\tilde{u}(\cdot)$ and a trajectory $\tilde{y}(t) = A(t, y_0, \tilde{u}(\cdot))$ such that $\tilde{y}(\cdot)$ steers the system from y_0 to y_1 , as before, and the cost J is minimized (or maximized).

Stabilization We say that \bar{y} is an equilibrium if there exists $\bar{u} \in U$ such that $A(t, \bar{y}, \bar{u}) = \bar{y}$ for every $t > 0$ (here \bar{u} indicates also the constant in time control function). Determine the control u as function of the state y so that \bar{y} is a (Lyapunov) stable equilibrium for the uncontrolled dynamical system $y(t) = A(t, y(0), u(y(\cdot)))$.

Observability Assume that we can observe not the state y , but a function $\phi(y)$ of the state. Determine conditions on ϕ so that the state y can be reconstructed from the evolution of $\phi(y)$ choosing $u(\cdot)$ suitably.

For the sake of simplicity, we restrict ourselves mainly to the first two problems and just mention

some facts about the others. Also, we focus on two cases:

Control of ordinary differential equations (ODEs) In this case $t \in \mathbb{R}, y \in \mathbb{R}^n$, U is a set, typically $U \subset \mathbb{R}^m$, and A is determined by a controlled ODE

$$\dot{y} = f(t, y, u) \tag{2}$$

A typical example in mathematical physics is the control of mechanical systems (Bloch 2003, Bullo and Lewis 2005).

Control of partial differential equations (PDEs) In this case $t \in \mathbb{R}, x \in \mathbb{R}^n, y(x)$ belongs to a Banach functional space, for example, $H^s(\mathbb{R}^{n+1}, \mathbb{R})$, U is a functional space, and A is determined by a controlled PDE,

$$F(t, x, y, y_t, y_{x_1}, \dots, y_{x_n}, y_t, \dots, u) = 0 \tag{3}$$

A typical example in mathematical physics is the control of wave equation using boundary conditions, see below.

There are various other possible situations we do not treat here: “stochastic control,” when y is a random variable and A defined by a (controlled) stochastic differential equation; “discrete time control,” where $t \in \mathbb{N}$; “hybrid control,” where t and y may have both discrete and continuous components, and so on.

As shown above, the control law can be assigned in (at least) two basically different ways. In open-loop form, as a function of time: $t \rightarrow u(t)$, and in closed-loop form or feedback, as a function of the state: $y \rightarrow u(y)$. For example, in optimal control we look for a control $\tilde{u}(t)$ in open-loop form, while in stabilization we search for a feedback control $u(y)$. The open-loop control depends on $y(0)$, while a feedback control can stabilize regardless of the initial condition.

Example 1 A point with unit mass moves along a straight line; if a controller is able to apply an external force u , then, calling $y_1(t), y_2(t)$, respectively, the position and the velocity of the point at time t , the motion is described by the control system

$$(\dot{y}_1, \dot{y}_2) = (y_2, u) \tag{4}$$

It is easy to check that the feedback control $u(y_1, y_2) = -y_1 - y_2$ stabilizes the system asymptotically to the origin, that is, for every initial data (\bar{y}_1, \bar{y}_2) , the solution of the corresponding Cauchy problem satisfies $\lim_{t \rightarrow \infty} (y_1, y_2)(t) = (0, 0)$.

Another simple problem consists in driving the point to the origin with zero velocity in minimum time from given initial data. It is quite easy to see that the optimal strategy is to accelerate towards the

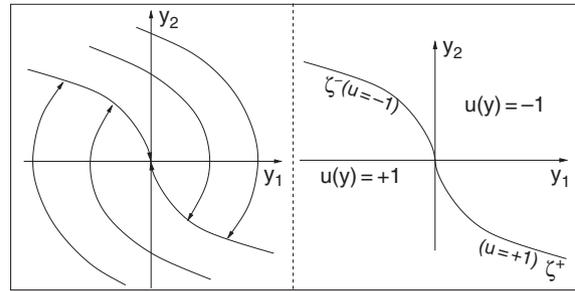


Figure 1 Example 1. The simplest example of (a) optimal synthesis and (b) corresponding feedback.

origin with maximum force on some interval $[0, t]$ and then to decelerate with maximum force to reach the origin at velocity zero. The set of optimal trajectories is depicted in Figure 1a: they can be obtained using the following discontinuous feedback, see Figure 1b. Define the curves $\zeta^\pm = \{(y_1, y_2) : \mp y_2 > 0, y_1 = \pm y_2^2\}$ and let ζ be defined as the union $\zeta^+ \cup \{0\}$. We define A^+ to be the region below ζ and A^- the one above. Then the feedback is given by

$$u(x) = \begin{cases} +1 & \text{if } (y_1, y_2) \in A^+ \cup \zeta^+ \\ -1 & \text{if } (y_1, y_2) \in A^- \cup \zeta^- \\ 0 & \text{if } (y_1, y_2) = (0, 0) \end{cases}$$

Example 2 Consider a (one-dimensional) vibrating string of unitary length with a fixed endpoint. The model for the motion of the displacement of the string with respect to the rest position is given by

$$y_{tt} + \Delta y = 0, \quad y(t, 0) = 0 \tag{5}$$

with initial data

$$y(0, \cdot) = y_0, \quad y_t(0, \cdot) = y_1 \tag{6}$$

Assume that we can control the position of the second endpoint; then,

$$y(t, 1) = u(t) \tag{7}$$

for some control function $u(\cdot) \in \mathbb{R}$.

Let us introduce another key concept: the reachable set at time t from \bar{y} is the set

$$R(t; \bar{y}) = \{A(t, \bar{y}, u(\cdot)) : u(\cdot) \in \mathcal{U}\}$$

Various problems can be formulated in terms of reachable sets, for example, controllability requires that for every \bar{y} the union of all $R(t; \bar{y})$ as $t \rightarrow \infty$ includes the entire space. The dependence of $R(t; \bar{y})$ on time t and on the set of controls \mathcal{U} is also a subject of investigation: one may ask whether the same points in $R(t; \bar{y})$ can be reached by using controls which are piecewise constant, or take values within some subsets of U .

Control of ODEs

For most proofs we refer to Agrachev and Sachkov (2004) and Sontag (1998).

Controllability

Consider first the case of a linear system:

$$\dot{y} = Ay + Bu, \quad u \in U, \quad y(0) = y_0 \quad [8]$$

where $y, y_0 \in R^n$, $U \subset R^m$, A is an $n \times n$ matrix and B an $n \times m$ matrix. We have the following property of reachable sets:

Theorem 1 *If U is compact convex then the reachable set $R(t)$ for [8] is compact and convex.*

A control system [8] is controllable if taking $U = R^m$ we have $R(t) = R^n$ for every $t > 0$. By linearity, this is equivalent to requiring the reachable set to be a neighborhood of the origin in case of bounded controls. Define the controllability matrix to be the $n \times nm$ matrix

$$C(A, B) = (B, AB, \dots, A^{n-1}B)$$

Controllability is characterized by the following:

Theorem 2 (Kalman controllability theorem). *The linear system [8] is controllable if and only if $\text{rank}(C(A, B)) = n$.*

For linear systems, there exists a duality between controllability and observability in the sense of the following theorem:

Theorem 3 *Consider the linear control system [8] and assume to observe the variable $z(y) = Cy$ for some $p \times n$ matrix C . Then, observability holds if and only if the linear system $\dot{y} = A^t y + C^t v$ is controllable.*

There exists no characterization of controllability for nonlinear systems as for linear ones, but we have the linearization result:

Theorem 4 *A nonlinear system is locally controllable if its linearization is. The converse is false.*

There are many results for the important class of control-affine systems

$$\dot{y} = f_0(y) + \sum_{i=1}^m f_i(y)u_i \quad [9]$$

where f_0, \dots, f_m are smooth vector fields on R^n and $U = R^m$. In general, there exists no explicit representation for the trajectories of [9], in terms of integrals of the control as it happens for linear systems. Still, a rich mathematical theory has been developed applying techniques and ideas from differential geometry:

the so-called geometric control theory. The main idea is that controllability (and properties of optimal trajectories) is determined by the Lie algebra generated by vector fields f_i . For example:

Theorem 5 (Lie-algebraic rank condition). *Let \mathcal{L} be the Lie algebra generated by the vector fields $f_i, i = 1, \dots, m$, and assume $f_0 = 0$. If $\mathcal{L}(y)$ is of dimension n at every point y then the system is controllable.*

We refer to Agrachev and Sachkov (2004) and Jurdjevic (1997) for general presentation of geometric control theory and give a simple example to show how Lie brackets characterize reachable directions.

Example 3 Consider the Brockett integrator

$$\dot{y}_1 = u_1, \quad \dot{y}_2 = u_2, \quad \dot{y}_3 = u_1 y_2 - u_2 y_1$$

Starting from the origin, using constant controls, we can move along curves tangent to the $y_1 y_2$ plane. However, let $f_1 = (1, 0, y_2)$ and $f_2 = (0, 1, -y_1)$ (fields corresponding to constant controls); then their Lie bracket is given by

$$[f_1, f_2](0) = (Df_2 \cdot f_1 - Df_1 \cdot f_2)(0) = (0, 0, -2)$$

Moving for time t first along the integral curve of f_1 , then of f_2 , then of $-f_1$, and finally of $-f_2$, we reach a point $t^2[f_1, f_2](0) + o(t^2)$ along the vertical direction y_3 . This corresponds to say that the system satisfies LARC.

Optimal Control

The theory of optimal control has developed in three main directions:

Existence of optimal controls, under various assumptions on L, f, U . When the sets $F(t, y)$ are convex, optimal solutions can be constructed following the direct method of Tonelli for the calculus of variations, that is, as limits of minimizing sequences: the two main ingredients are compactness and lower-semicontinuity. If convexity does not hold, existence is not granted in general but for special cases.

Necessary conditions for the optimality of a control $u(\cdot)$. The major result in this direction is the celebrated ‘‘Pontryagin maximum principle’’ (PMP) which extends the Euler–Lagrange equation to control systems, and the Weierstrass necessary conditions for a strong local minimum in the calculus of variations. Various extensions and other necessary conditions are now available (Agrachev and Sachkov 2004).

Sufficient conditions for optimality. The standard procedure resorts to embedding the optimal control problem in a family of problems, obtained by

varying the initial conditions. One defines the value function V by

$$V(t, \bar{y}) = \inf J(y(\cdot), u(\cdot))$$

where the inf is taken over the set of trajectories and controls satisfying $y(t) = \bar{y}$. Under suitable assumptions, V is the solution to a first-order Hamilton–Jacobi PDE. The lack of regularity of the value function V has long provided a major obstacle to a rigorous mathematical analysis, solved by the theory of viscosity solutions (Bardi and Capuzzo Dolcetta 1997). Another method consists in building an optimal synthesis, that is, a collection of trajectory–control pairs.

Pontryagin maximum principle Consider a general autonomous control system:

$$\dot{y} = f(y, u) \tag{10}$$

where $y \in R^n$ and $u \in U$ compact subset of R^m . We assume to have regularity of f guaranteeing existence and uniqueness of trajectories for every $u(\cdot) \in \mathcal{U}$. For a fixed $T > 0$, an optimal control problem in Mayer form is given by

$$\min_{u(\cdot) \in \mathcal{U}} \psi(y(T, u)), \quad y(0) = \bar{y} \tag{11}$$

where ψ is the final cost and \bar{y} the initial condition. More generally, one can consider also the Lagrangian cost $\int L(y, u)dt$ and reduce to this case by adding a variable $y_0(0) = 0$ and $\dot{y}_0 = L$.

The well-known PMP provides, under suitable assumptions, a necessary condition for optimality in terms of a lift of the candidate optimal trajectory to the cotangent bundle. For problems as [11], PMP can be stated as follows:

Theorem 6 *Let $u^*(\cdot)$ be a (bounded) admissible control whose corresponding trajectory $y^*(\cdot) = y(\cdot, u^*)$ is optimal. Call $p: [0, T] \mapsto R^n$ the solution of the adjoint linear equation*

$$\begin{aligned} \dot{p}(t) &= -p(t) \cdot D_y f(y^*(t), u^*(t)) \\ p(T) &= \nabla \psi(y^*(T)) \end{aligned} \tag{12}$$

Then the maximality condition

$$p(t) \cdot f(y^*(t), u^*(t)) = \max_{\omega \in U} p(t) \cdot f(y^*(t), \omega) \tag{13}$$

holds for almost every time $t \in [0, T]$.

Notice that the conclusion of the theorem can be interpreted by saying that the pair (y, p) satisfies the system:

$$\dot{y} = \frac{\partial H(y^*, p, u^*)}{\partial p}, \quad \dot{p} = - \frac{\partial H(y^*, p, u^*)}{\partial y}$$

where $H(y, p, u) = \langle p, f(y, u) \rangle$. This is a pseudo–Hamiltonian system, since H also depends on u^* .

Alternatively, one can define the maximized Hamiltonian

$$\mathcal{H}(y, p) = \max_u \langle p, f(y, u) \rangle$$

but \mathcal{H} may fail to be smooth. Another difficulty lies in the fact that an initial condition is given for y and a final condition is given for λ .

The proof of PMP relies on a special type of variations, called needle variations, of a reference trajectory. Given a candidate optimal control u^* and corresponding trajectory y^* , a time τ of approximate continuity for $f(y^*(\cdot), u^*(\cdot))$ and $\omega \in U$, a needle variation is a family of controls u_ε obtained by replacing u^* with ω on the interval $[\tau - \varepsilon, \tau]$. A needle variation gives rise to a variation v of the trajectory satisfying the variational equation

$$\dot{v}(t) = D_y f(y^*(t), u^*(t)) \cdot v(t) \tag{14}$$

in classical sense only after time τ . Recently Piccoli and Sussmann (2000) introduced a setting in which needle and other variations happen to be differentiable.

One may also consider some final (or initial) constraint:

$$(T, y(T)) \in S \tag{15}$$

where $S \subset R \times R^n$ (and T not fixed). In this case, the final condition for p is more complicated as well as the proof of PMP. It is interesting to note the many connections between PMP and classical mechanics framework well illustrated by Bloch (2003) and Jurdjevic (1997).

Value function and HJB equation In this section we consider the minimization problem

$$\inf_{u \in \mathcal{U}} \psi(T, y(T, u)) \tag{16}$$

for the control system

$$\dot{y} = f(t, y, u), \quad u(t) \in U \quad \text{a.e.} \tag{17}$$

subject to the terminal constraints [15], where $S \subset R^{n+1}$ is a closed target set.

Theorem 7 (PDE of dynamic programming). *Assume that the value function V , for [15]–[17], is C^1 on some open set $\Omega \subseteq R \times R^n$, not intersecting the target set S . Then V satisfies the Hamilton–Jacobi equation*

$$\begin{aligned} V_s(s, y) + \min_{\omega \in U} \{ V_y(s, y) \cdot f(s, y, \omega) \} &= 0 \\ \forall (s, y) \in \Omega \end{aligned} \tag{18}$$

Equation [18] is called the Hamilton–Jacobi–Bellman (HJB) equation, after Richard Bellman. In general,

however, V fails to be differentiable: this is the case for Example 1 along the lines ζ^\pm . To isolate V as the unique solution of the HJB equation, one has to resort to the concept of viscosity solution. The dynamic programming and HJB equation apparatus applies also to stochastic problems for which the equation happens to be parabolic, because of the Ito formula.

Optimal syntheses Roughly speaking, an optimal synthesis is a collection of optimal trajectories, one for each initial condition \bar{y} . Geometric techniques provide a systematic method to construct syntheses:

Step 1 Study the properties of optimal trajectories via PMP and other necessary conditions.

Step 2 Determine a (finite-dimensional) sufficient family for optimality, that is, a class of trajectories (satisfying PMP) containing all possible optimal ones.

Step 3 Construct a synthesis selecting one trajectory for every initial condition in such a way as to cover the state space in a regular fashion.

Step 4 Prove that the synthesis of Step 3 is indeed optimal.

One of the main problems in step 2 is the possible presence of optimal controls with an infinite number of discontinuities, known as Fuller phenomenon. The key concept of regular synthesis, of step 3, was introduced by Boltianskii and recently refined by Piccoli and Sussmann (2000) to include Fuller phenomena. The above strategy works only in some special cases, for example for two-dimensional minimum-time problems (Boscaïn and Piccoli 2004); we report below an example.

Example 4 Consider the problem of orienting in minimum time a satellite with two orthogonal rotors: the speed of one rotor is controlled, while the second rotor has constant speed. This problem is modelled by a left-invariant control system on $SO(3)$:

$$\dot{y} = y(F + uG), \quad y \in SO(3), \quad |u| \leq 1$$

where F and G are two matrices of $\mathfrak{so}(3)$, the Lie algebra of $SO(3)$. Using the isomorphism of Lie algebras $(SO(3), [\cdot, \cdot]) \sim (R^3, \times)$, the condition that the rotors are orthogonal reads: $\text{trace}(F \cdot G) = 0$. If we are interested to orient only a fixed semi-axis then we project the system on the sphere S^2 :

$$\dot{y} = y(F + uG), \quad y \in S^2, \quad |u| \leq 1$$

In this case, $F + G$ and $F - G$ are rotations around two fixed axes and, if the angle between these two axes is less than $\pi/2$, every optimal trajectory is a finite concatenation of arcs corresponding to constant control $+1$ or -1 . The “optimal synthesis” can be obtained by the feedback shown in Figure 2.

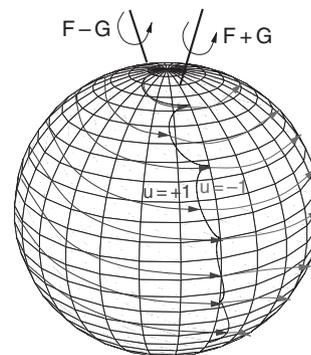


Figure 2 Optimal feedback for Example 4.

Control of PDEs

The theory for control of models governed by PDEs is, as expected, much more ramified and much less complete. An exhaustive resume of the available results is not possible in short space, thus we focus on Example 2 and few others to illustrate some techniques to treat control problems and give various references (see also Fursikov and Imanuvilov (1996), Komornik (1994), and Lasiecka and Triggiani (2000), and references therein).

Besides the variety of control problems illustrated in the Introduction, for PDE models one can consider different ways of applying the control, for example:

Boundary control One consider the system [3] (with F independent of u) and impose the condition $y(t, x) = u(t, x)$ to hold for every time t and every x in some region. Usually, we assume $y(t)$ to be defined bounded region Ω and the control acts on some set $\Gamma \subset \partial\Omega$. Obviously, also Neumann conditions are natural as $\partial_\nu y = u$ where ν is the exterior normal to Ω .

Internal control One consider the system [3] with F depending on u . Thus, the control acts on the equation directly.

Other controls There are various other control problems one may consider as Galerkin-type approximation and control of some finite family of modes. An interesting example is given by Coron (2002), where the position of a tank is controlled to regulate the water level inside.

Control of a Vibrating String

We consider Example 2, but various results hold for hyperbolic linear systems in general. First consider the uncontrolled system

$$z_{tt} = \Delta z, \quad z(0, t) = z(1, t) = 0 \quad [19]$$

A first integral is the energy given by

$$E(t) = \frac{1}{2} \int \left[|z_x|^2 + |z_t|^2 \right] dx$$

Then we say that the system [19] is observable at time T if there exists $C(T)$ such that

$$E(0) \leq C(T) \int_0^T |z_x(1, t)|^2 dt$$

which means that if we observe zero displacement on the right end for time T then the solution has zero energy and hence vanishes. In this case, the system is observable for every time $T \geq 2$: this is precisely the time taken by a wave to travel from the right end point to the left one and backward.

Thanks to a duality as for the finite-dimensional case, observability of [19] is equivalent to null controllability for [5]–[7], that is, to the property that for every initial conditions y_0, y_1 there exists a control $u(\cdot)$ such that the corresponding solution verifies $y(x, T) = y_t(x, T) = 0$. More precisely, the desired control is given by $u(t) = \tilde{z}_x(1, t)$, where \tilde{z} is the solution of [19] minimizing the functional (over $L^2 \times H^{-1}$)

$$\begin{aligned} J(z(\cdot, 0), z_t(\cdot, 0)) \\ = \frac{1}{2} \int_0^T |z_x(1, t)|^2 dt + \int y_0 z_t(\cdot, 0) dx - \int y_1 z(\cdot, 0) dx \end{aligned}$$

One can check that this functional is continuous and convex, and the coercivity is granted by the observability of [19]; thus, a minimum exists by the direct method of Tonelli. This is an example of the method known as Hilbert's uniqueness method introduced by Lions (1988).

In the multidimensional case, controllability can be characterized by imposing a condition on the region $\Gamma \subset \partial\Omega$ on which the control acts. More precisely, rays of geometric optics in Ω should intersect Γ (Zuazua 2005).

If we consider infinite-time horizon $T = +\infty$ and introduce the functional

$$J = \int_0^{+\infty} \|y\|^2 dt + N \int u^2 dt dx$$

then the optimal control is determined as follows. If (y, p) is a solution of the optimality system: [5]–[6] with $y = 0$ outside Γ and

$$\begin{aligned} p_{tt} - \Delta p + y &= 0, \quad \partial_\nu p + Ny = 0 \quad \text{on } \Gamma \\ p &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

then $u = y$ on Γ (Lions 1988, Zuazua 2005).

Controllability via Return Method of Coron

As we saw in Theorem 4, a nonlinear system may be controllable even if its linearization is not. In this case, controllability can be proved by the return

method of Coron, which consists in finding a trajectory y such that the following hold:

1. $y(0) = y(T) = 0$;
2. the linearized system around y is controllable.

Then by implicit-function theorem, local controllability is granted, that is, there exists $\varepsilon > 0$ such that for every data y_0, y_1 of norm less than ε , there exists a control steering the system from y_0 to y_1 in time T .

This method does not give many advantages in the finite-dimensional case, but permits to obtain excellent results for PDE systems such as Euler, Navier–Stokes, Saint–Venant, and others (Coron 2002).

Control of Schrödinger Equation

Consider the issue of designing an efficient transfer of population between different atomic or molecular levels using laser pulses. The mathematical description consists in controlling the Schrödinger equation. Many results are available in the finite-dimensional case. Finite-dimensional closed quantum systems are in fact left-invariant control systems on $SU(n)$, or on the corresponding Hilbert sphere $S^{2n-1} \subset \mathbb{C}^n$, where n is the number of atomic or molecular levels, and powerful techniques of geometric control are available both for what concerns controllability and optimal control (Agrachev and Sachkov 2004, Boscain and Piccoli 2004, Jurdjevic 1997).

Recent papers consider the minimum-time problem with unbounded controls as well as minimization of the energy of transition. Boscain *et al.* (2002) have applied the techniques of sub-Riemannian geometry on Lie groups and of optimal synthesis on two-dimensional manifolds to the population transfer problem in a three-level quantum system driven by two external fields of arbitrary shape and frequency.

Although many results are available for finite-dimensional systems, only few controllability properties have been proved for the Schrödinger equation as a PDE, and in particular no satisfactory global controllability results are available at the moment.

Further Reading

- Agrachev A and Sachkov Y (2004) *Control from a Geometric Perspective*. Springer.
- Bardi M and Capuzzo Dolcetta I (1997) *Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations*. Boston: Birkhauser.
- Bloch AM (2003) *Nonholonomic Mechanics and Control*, with the collaboration of J. Baillieul, P. Crouch and J. Marsden, with scientific input from P. S. Krishnaprasad, R. M. Murray and D. Zenkov. New York: Springer.
- Boscain U and Piccoli B (2004) *Optimal Synthesis for Control Systems on 2-D Manifolds*. Springer SMAI, vol. 43. Heidelberg: Springer.

- Boscain U, Chambrion T, and Gauthier J-P (2002) On the $K + P$ problem for a three-level quantum system: optimality implies resonance. *Journal of Dynamical and Control Systems* 8: 547–572.
- Bullo F and Lewis AD (2005) *Geometric Control of Mechanical Systems*. New York: Springer.
- Coron JM (2002) Return method: some application to flow control. *Mathematical Control Theory, Part 1, 2 (Trieste, 2001)*. In: Agrachev A (ed.) *ICTP Lecture Notes*, vol. VIII. Trieste: Abdus Salam Int. Cent. Theoret. Phys.
- Fursikov AV and Imanuvilov O Yu (1996) *Controllability of Evolution Equations*. Lecture Notes Series, vol. 34. Seoul: Seoul National University.
- Jurdjevic V (1997) *Geometric Control Theory*. Cambridge: Cambridge University Press.
- Komornik V (1994) *Exact Controllability and Stabilization. The Multiplier Method*. Chichester: Wiley.
- Lasiecka I and Triggiani R (2000) *Control theory for Partial Differential Equations: Continuous and Approximation Theories*. Cambridge: Cambridge University Press.
- Lions JL (1988) Exact controllability, stabilization and perturbations for distributed systems. *SIAM Review* 30: 1–68.
- Piccoli B and Sussmann HJ (2000) Regular synthesis and sufficiency conditions for optimality. *SIAM Journal of Control Optimization* 39: 359–410.
- Sontag ED (1998) *Mathematical Control Theory*. New York: Springer.
- Zuazua E (2005) Propagation, observation and control of wave approximatem by finite difference methods. *SIAM Review* 47: 197–243.

Convex Analysis and Duality Methods

G Bouchitté, Université de Toulon et du Var,
La Garde, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Convexity is an important notion in nonlinear optimization theory as well as in infinite-dimensional functional analysis. As will be seen below, very simple and powerful tools will be derived from elementary duality arguments (which are by-products of the Moreau–Fenchel transform and Hahn–Banach theorem). We will emphasize on applications to a large range of variational problems. Some arguments of measure theory will be skipped.

Basic Convex Analysis

In the following, we denote by X a normed vector space, and by X^* the topological dual of X . If a topology different from the normed topology is used on X , we will denote it by τ . For every $x \in X$ and $A \subset X$, \mathcal{V}_x denotes the open neighborhoods of x and $\text{int} A$, $\text{cl} A$, respectively, the interior and the closure of A . We deal with extended real-valued functions $f: X \rightarrow \mathbb{R} \cup \{+\infty\}$. We denote by $\text{dom} f = f^{-1}(\mathbb{R})$ and by $\text{epi} f = \{(x, \alpha) \in X \times \mathbb{R}: f(x) \leq \alpha\}$ the domain and the epigraph of f , respectively. We say that f is proper if $\text{dom} f \neq \emptyset$. Recall that f is convex if for every $(x, y) \in X^2$ and $t \in [0, 1]$, there holds

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

(by convention $\infty + a = +\infty$)

The notion of convexity for a subset $A \subset X$

is recovered by saying that χ_A is convex, where its indicator function χ_A is defined by setting

$$\chi_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise} \end{cases}$$

Continuity and Lower-Semicontinuity

A first consequence of the convexity is the continuity on the topological interior of the domain. We refer for instance to [Borwein and Lewis \(2000\)](#) for a proof of

Theorem 1 *Let $f: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex and proper. Assume that $\sup_U f < +\infty$, where U is a suitable open subset of X . Then f is continuous and locally Lipschitzian on all $\text{int}(\text{dom} f)$.*

As an immediate corollary, a convex function on a normed space is continuous provided it is majorized by a locally bounded function. In the finite-dimensional case, it is easily deduced that a finite-valued convex function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is locally Lipschitz. Furthermore, by Aleksandrov's theorem, f is almost everywhere twice differentiable and the non-negative Hessian matrix $\nabla^2 f$ coincides with the absolutely continuous part of the distributional Hessian matrix $D^2 f$ (it is a Radon measure taking values in the non-negative symmetric matrices).

However, in infinite-dimensional spaces, for ensuring compactness properties (as, e.g., in condition (ii) of [Theorem 4](#) below), we need to use weak topologies and the situation is not so simple. A major idea consists in substituting the continuity property with lower-semicontinuity.

Definition 2 A function $f: X \rightarrow \mathbb{R} \cup \{+\infty\}$ is τ -l.s.c. at $x_0 \in X$ if for all $\alpha \in \mathbb{R}$, there exists $U \in \mathcal{V}_{x_0}$ such that $f > \alpha$ on U . In particular, f will be l.s.c. on all X provided $f^{-1}((r, +\infty))$ is open for every $r \in \mathbb{R}$.

Remark 3

- (i) The following sequential notion can be also used: f is τ -sequentially l.s.c. at x_0 if

$$\forall (x_n) \subset X \ x_n \xrightarrow{\tau} x_0 \implies \liminf_{n \rightarrow +\infty} f(x_n) \geq f(x_0)$$

It turns out that this notion (weaker in general) is equivalent to the previous one provided x_0 admits a countable basis of neighborhoods.

- (ii) A well-known consequence of Hahn–Banach theorem is that, for convex functions, the lower-semicontinuity property with respect to the normed topology of X is equivalent to the weak (or weak sequential) lower-semicontinuity.

Theorem 4 (Existence). *Let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, such that*

- (i) f is τ -l.s.c.,
- (ii) $\forall r \in \mathbb{R}, f^{-1}((-\infty, r])$ is τ -relatively compact.

Then there is $\bar{x} \in X$ such that $f(\bar{x}) = \inf f$ and $\operatorname{argmin} f := \{x \in X | f(x) = \inf f\}$ is τ -compact.

In practice, the choice of the topology τ is ruled by the condition (ii) above. For example, if X is a reflexive infinite-dimensional Banach space and if f is coercive (i.e., $\lim_{\|x\| \rightarrow \infty} f(x) = +\infty$), we may take for τ the weak topology (but never the normed topology). This restriction implies in practice that the first condition in **Theorem 4** may fail. In this case, it is often useful to substitute f with its lower-semicontinuous (l.s.c.) envelope.

Definition 5 Given a topology τ , the relaxed function $\bar{f} (= \bar{f}^\tau)$ is defined as

$$\bar{f}(x) = \sup \{g(x) | g : X \rightarrow \mathbb{R} \cup \{+\infty\}, \\ g \text{ is } \tau\text{-l.s.c.}, g \leq f\}$$

It is easy to check that f is τ -l.s.c. at x_0 if and only if $\bar{f}(x_0) = f(x_0)$. Furthermore,

$$\bar{f}(x) = \sup_{U \in \mathcal{V}_x} \inf_U f, \quad \operatorname{epi} \bar{f} = \operatorname{cl}_{(X \times \mathbb{R})} (\operatorname{epi} f)$$

We can now state the relaxed version of Theorem 1.4.

Theorem 6 (Relaxation). *Let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$, then: $\inf f = \inf \bar{f}$. Assume further that, for all real r , $f^{-1}((-\infty, r])$ is \mathcal{T} -relatively compact; then f attains its minimum and $\operatorname{argmin} f = \operatorname{argmin} \bar{f} \cap \{x \in X | f(x) = \bar{f}(x)\}$.*

Moreau–Fenchel Conjugate

The duality between X and X^* will be denoted by the symbol $\langle \cdot | \cdot \rangle$. If X is a Euclidian space, we identify X^* with X via the scalar product denoted $(\cdot | \cdot)$.

Definition 7 Let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$. The Moreau–Fenchel conjugate $f^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ of f is defined by setting, for every $x^* \in X^*$:

$$f^*(x^*) = \sup \{ \langle x | x^* \rangle - f(x) | x \in X \}$$

In a symmetric way, if f^* is proper on X^* , we define the biconjugate $f^{**} : X \rightarrow \mathbb{R} \cup \{+\infty\}$ by setting

$$f^{**}(x) = \sup \{ \langle x | x^* \rangle - f^*(x^*) | x^* \in X^* \}$$

As a consequence, the so-called Fenchel inequality holds:

$$\langle x | x^* \rangle \leq f(x) + f^*(x^*), \quad (x, x^*) \in X \times X^*$$

Notice that f does not need to be convex. However, if f is convex, then f^* agrees with the Legendre–Fenchel transform.

Definition 8 Let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$. The sub-differential of f at x is the possibly void subset of $\partial f(x) \subset X^*$ defined by

$$\partial f(x) := \{x^* \in X^* : f(x) + f^*(x^*) = \langle x, x^* \rangle\}$$

It is easy to check that $\partial f(x)$ is convex and weak-star closed. Moreover, if f is convex and has a differential (or Gateaux derivative) $f'(x)$ at x , then $\partial f(x) = \{f'(x)\}$. After summarizing some elementary properties of the Fenchel transform, we give examples in \mathbb{R}^d or in infinite-dimensional spaces.

Lemma 9

- (i) f^* is convex, l.s.c. with respect to the weak star topology of X^* .
- (ii) $f^*(0) = -\inf f$ and $f \geq g \implies f^* \leq g^*$.
- (iii) $(\inf_i f_i)^* = \sup_i f_i^*$, for every family $\{f_i\}$.
- (iv) $f^{**}(x) = \sup \{g(x) : g \text{ affine continuous on } X \text{ and } g \leq f\}$ (by convention, the supremum is identically $-\infty$ if no such g exists).

Proof (i) This assertion is a direct consequence of the fact that f^* can be written as the supremum of functions g_x , where $g_x := \langle x | \cdot \rangle - f(x)$. Clearly, these functions are affine and weakly star-continuous on X^* . The assertions (ii), (iii) are trivial. To obtain (iv), it is enough to observe that an affine function g of the form $g(x) = \langle x, x^* \rangle - \beta$ satisfies $g \leq f$ iff $f^*(x^*) \leq \beta$. \square

Example 1 Let $f : X \rightarrow \mathbb{R}$, be defined by

$$f(x) = \frac{1}{p} \|x\|_X^p, \quad 1 < p < +\infty$$

then,

$$f^*(x^*) = \frac{1}{p'} \|x^*\|_{X^*}^{p'}, \quad \text{with } \frac{1}{p} + \frac{1}{p'} = 1$$

whereas, for $p=1$, we find $f^* = \chi_{B^*}$, where $B^* = \{\|x^*\| \leq 1\}$.

Example 2 Let $A \in \mathbb{R}_{\text{sym}}^{d^2}$ be a symmetric positive-definite matrix and let $\tilde{f}(x) := (1/2)(Ax | x)$ ($x \in \mathbb{R}^d$). Then, for all $y \in \mathbb{R}^d$, we have $f^*(y) = (1/2)(A^{-1}y | y)$. Notice that if A has a negative eigenvalue, then $f^* \equiv +\infty$.

Particular examples on \mathbb{R}^d are also very popular. For instance:

Minimal surfaces

$$f(x) = \sqrt{1 + |x|^2}$$

$$f^*(y) = \begin{cases} -\sqrt{1 - |y|^2} & \text{if } |y| \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

Entropy

$$f(x) = \begin{cases} x \log x & \text{if } x \in \mathbb{R}_+ \\ +\infty & \text{otherwise} \end{cases}, \quad f^*(y) = \exp(y - 1)$$

Example 3 Let $C \subset X$ be convex, and let $f = \chi_C$. Then,

$$f^*(x^*) = \sigma_C(c^*) = \sup_{x \in C} \langle x | x^* \rangle$$

(support function of C)

Notice that if M is a subspace of X , then $(\chi_M)^* = \chi_{M^\perp}$. We specify now a particular case of interest.

Let Ω be a bounded open subset of \mathbb{R}^n . Take $X = C_0(\bar{\Omega}; \mathbb{R}^d)$ to be the Banach space of continuous functions on the compact $\bar{\Omega}$ with values in \mathbb{R}^d . As usual, we identify the dual X^* with the space $\mathcal{M}_b(\bar{\Omega}; \mathbb{R}^d)$ of \mathbb{R}^d -valued Borel measures on $\bar{\Omega}$ with finite total variation. Let K be a closed convex of \mathbb{R}^d such that $0 \in K$. Then $\rho_K^0(\xi) := \sup\{(\xi | z) : z \in K\}$ is a non-negative convex l.s.c. and positively 1-homogeneous function on \mathbb{R}^d (e.g., ρ_K is the Euclidean norm if K is the unit ball of \mathbb{R}^d). Let us define $C := \{\varphi \in X : \varphi(x) \in K, \forall x \in \Omega\}$. Then, we have

$$(\chi_C)^*(\lambda) = \int_{\Omega} \rho_K^0(\lambda)$$

$$:= \int_{\Omega} \rho_K^0\left(\frac{d\lambda}{d\theta}\right) \theta(dx) \quad [1]$$

where θ is any non-negative Radon measure such that $\lambda \ll \theta$ (the choice of θ is indifferent). In the case where K is the unit ball, we recover the total variation of λ .

Example 4 (Integral functionals). Given $1 \leq p < +\infty$, $(\Omega, \mu, \mathcal{T})$ a measured space and $\varphi : \Omega \times$

$\mathbb{R}^d \rightarrow [0, +\infty]$ a $\mathcal{T} \otimes B_{\mathbb{R}^d}$ -measurable integrand. Then the partial conjugate $\varphi^*(x, z^*) := \sup\{\langle z | z^* \rangle - \varphi(x, z) : z \in \mathbb{R}^d\}$ is a convex measurable integrand. Let us define

$$I_\varphi : u \in (L^p_\mu)^d \rightarrow \int_{\Omega} \varphi(x, u(x)) d\mu \in \mathbb{R} \cup \{+\infty\}$$

and assume that I_φ is proper. Then there holds $(I_\varphi)^* = I_{\varphi^*}$, where

$$(I_\varphi)^* : v \in (L^{p'}_\mu)^d \rightarrow \int_{\Omega} \varphi^*(x, v(x)) d\mu$$

Duality Arguments

Two Key Results

The first result related to the biconjugate f^{**} is a consequence of the Hahn–Banach theorem. Recalling the assertion (v) of Lemma 9, we notice that the existence of an affine minorant for f is equivalent to the properness of f^* (i.e., $\exists x_0^* \in X^* : f^*(x_0^*) < +\infty$).

Theorem 10 Let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex and proper. Then

- (i) f is l.s.c. at x_0 if and only if f^* is proper and $f^{**}(x_0) = f(x_0)$. In particular, the lower-semicontinuity of f on all X is equivalent to the identity $f \equiv f^{**}$.
- (ii) If f^* is proper, then $f^{**} = \bar{f}$.

Proof We notice that by Lemma 9, $f^{**} \leq f$ and f^{**} is l.s.c (even for the weak topology). Therefore, $f^{**} \leq \bar{f}$ and, moreover, f is l.s.c. at x_0 if $f^{**}(x_0) \geq f(x_0)$. Conversely, if f is l.s.c. at x_0 , for every $\alpha_0 < f(x_0)$, there exists a neighborhood V of x_0 such that $V \times (-\infty, \alpha_0) \cap \text{epi } \bar{f} = \emptyset$. It follows that $\text{epi } \bar{f}$ is a proper closed convex subset of $X \times \mathbb{R}$ which does not intersect the compact singleton $\{(x_0, \alpha_0)\}$. By applying the Hahn–Banach strict separation theorem, there exists $(x_0^*, \beta_0) \in X^* \times \mathbb{R}$ such that

$$\langle x_0, x_0^* \rangle + \alpha_0 \beta_0 < \langle x, x_0^* \rangle + \alpha \beta_0$$

for all $(x, \alpha) \in \text{epi } \bar{f}$

Taking $\alpha \rightarrow \infty$ and $x \in \text{dom } f$, we find $\beta_0 \geq 0$. In fact, $\beta_0 > 0$ as the strict inequality above would be violated for $x = x_0$. Eventually, we obtain that f is minorized by the affine continuous function $g(x) = -\langle x - x_0, x_0^* / \beta \rangle + \alpha_0$. Thus, we conclude that f^* is proper and that $f^{**}(x_0) \geq \alpha_0$.

The assertion (ii) is a direct consequence of the equivalence in (i). □

Theorem 11 Let X be a normed space and let $f: X \rightarrow [0, +\infty]$ be a convex and proper function; assume that f is continuous at 0, then

- (i) f^* achieves its minimum on X^*
- (ii) $f(0) = f^{**}(0) = -\inf f^*$

Proof

- (i) Let M be an upper bound of f on the ball $\{\|x\| \leq R\}$. Then

$$f^*(x^*) \geq \sup\{\langle x, x^* \rangle - f(x) : \|x\| \leq R\} \geq R\|x^*\|_{X^*} - M$$

Hence, for every r , the set $\{x^* \in X^* : f^*(x^*) \leq r\}$ is bounded, thus τ -relatively compact, where τ is the weak-star topology on X^* . By assertion (i) of Lemma 9, f^* is τ -l.s.c. and Theorem 4 applies.

- (ii) By Theorem 10, since f is convex proper and l.s.c. at $x_0 = 0$, we have $f(0) = f^{**}(0) = -\inf f^*$. □

Some Useful Consequences

Proposition 12 (Conjugate of a sum). Let $f, g: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex such that

$$\exists x_0 \in X : f \text{ is continuous at } x_0 \text{ and } g(x_0) < +\infty \quad [2]$$

Then

- (i) $(f + g)^*(x^*) = \inf_{x_1^* + x_2^* = x^*} \{f^*(x_1^*) + g^*(x_2^*)\}$

(the equality holds in $\bar{\mathbb{R}}$).

- (ii) If both sides of the equality in (i) are finite, then the infimum in the right-hand side is achieved.

Proof Without any loss of generality, we may assume that $x^* = 0$ (we reduce to this case by substituting g with $g - \langle \cdot, x^* \rangle$). We let

$$h(p) = \inf\{f(x + p) + g(x) | x \in X\}$$

Noticing that $(p, x) \mapsto f(x + p) + g(x)$ is convex, we infer that $h(p)$ is convex as well. As h is majorized by the function $p \mapsto f(x_0 + p) + g(x_0)$, which by [2] continuous at 0, we deduce from Theorems 1 and 11 that $h(0) = h^{**}(0)$ and that h^* achieves its infimum. Now $h(0) = \inf(f + g) = -(f + g)^*(0)$ and

$$\begin{aligned} h^*(p^*) &= \sup\{\langle p, p^* \rangle - h(p) : p \in X\} \\ &= \sup\{\langle p, p^* \rangle - f(x + p) - g(x) : x \in X, p \in X\} \\ &= g^*(-p^*) + f^*(p^*) \end{aligned}$$

The assertions (i), (ii) follow since $-h^{**}(0) = \min h^* = \min \{g^*(-p^*) + f^*(p^*)\}$. □

Proposition 13 (Composition). Let X, Y be two Banach spaces and $A: X \mapsto Y$ a linear operator with dense domain $D(A)$. Let $\Psi: Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be a

convex l.s.c. function and let $F: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be the convex functional defined by

$$F(u) = \begin{cases} \Psi(Au) & \text{if } u \in D(A) \\ +\infty & \text{otherwise} \end{cases}$$

Assume that there exists $u_0 \in D(A)$ such that Ψ is continuous at Au_0 . Then

- (i) The Fenchel conjugate of F is given by

$$\forall f \in X^*, \quad F^*(f) = \inf\{\Psi^*(\sigma) : \sigma \in Y^*, A^*\sigma = f\}$$

where, if both sides of the equality are finite, the infimum on the right-hand side is achieved.

- (ii) If, in addition, Y is reflexive and Ψ is l.s.c. coercive, we have

$$\bar{F}(u) = F^{**}(u) = \inf\{\Psi(p) : (u, p) \in \overline{G(A)}\} \quad [3]$$

where $G(A)$ denotes the graph of A .

Proof

- (i) Define $H, K: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ by

$$H(u, p) = \chi_{G(A)}(u, p), \quad K(u, p) = \Psi(p)$$

Then we have the identity $F^*(f) = (H + K)^*(f, 0)$, where the conjugate of $H + K$ is taken with respect to the duality $(X \times Y, X^* \times Y^*)$. From the assumption, K is continuous at $(u_0, Au_0) \in \text{dom } H$. By Proposition 12, we obtain

$$\begin{aligned} (H + K)^*(f, 0) &= \inf_{(g, \sigma) \in X^* \times Y^*} \{K^*(f - g, \sigma) + H^*(g, -\sigma)\} \end{aligned}$$

After a simple computation, it is easy to check that

$$\begin{aligned} H^*(g, -\sigma) &= \begin{cases} 0 & \text{if } A^*\sigma = f \\ +\infty & \text{otherwise} \end{cases} \\ K^*(f - g, \sigma) &= \begin{cases} \Psi^*(\sigma) & \text{if } g = f \\ +\infty & \text{otherwise} \end{cases} \end{aligned}$$

- (ii) Let $J(u) := \inf\{\Psi(p) : (u, p) \in \overline{G(A)}\}$. As observed for F^* in the proof of (i), we have the identity $J^*(f) = (H + K)^*(f, 0)$. Therefore, in view of Theorem 10, $\bar{F} = F^{**} = J^{**}$ and it is enough to prove that J is convex l.s.c. proper. Let us consider a sequence (u_n) in X converging to some $u \in X$. Without any loss of generality, we may assume that $\liminf J(u_n) = \lim J(u_n) < +\infty$. Then there is a sequence (p_n) such that, for every n , $(u_n, p_n) \in \bar{G}(A)$ and $J(u_n) \geq \psi(u_n) - 1/n$. As ψ is coercive, $\{p_n\}$ is bounded in the reflexive space Y and possibly passing to a subsequence,

we may assume that p_n converges weakly to some p . Since $\overline{G(A)}$ is a (weakly) closed subspace of $X \times Y$, we infer that (u, p) as the limit of (u_n, p_n) still belongs to $\overline{G(A)}$. Thus, we conclude, thanks to the (weak) lower-semicontinuity of Ψ

$$\liminf_n J(u_n) = \lim_n \Psi(p_n) \geq \Psi(p) \geq J(u) \quad \square$$

An immediate consequence of Propositions 12 and 13 is the following variant:

Proposition 14 *Under the same notation as in Proposition 13, let $\Phi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function and assume that there exists $u_0 \in D(A)$ such that $F(u_0) < +\infty$ and Ψ is continuous at Au_0 . Then we have*

$$\inf_{u \in X} \{\phi(u) + \Psi(Au)\} = \sup_{\sigma \in Y^*} \{-\phi^*(-A^*\sigma) - \Psi^*(\sigma)\}$$

where the supremum on the right-hand side is achieved. Furthermore, a pair $(\bar{u}, \bar{\sigma})$ is optimal if and only if it satisfies the relations: $\bar{\sigma} \in \partial\Psi(A\bar{u})$ and $-A^*\bar{\sigma} \in \partial\phi(\bar{u})$.

Remark 15 From the assertion (ii) of Proposition 13, we may conclude that F is l.s.c. whenever the operator A is closed. If now A is merely closable (with closure denoted by \bar{A}), we obtain

$$\bar{F}(u) = \begin{cases} G(\bar{A}u) & \text{if } u \in \text{dom } \bar{A} \\ +\infty & \text{otherwise} \end{cases}$$

This is the typical situation when F is an integral functional defined on smooth functions of the kind

$$F(u) = \int_{\Omega} f(x, \nabla u) \, dx$$

where Ω is an bounded open subset of \mathbb{R}^n , $f: \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex integrand with quadratic growth (i.e., $c|z|^2 \leq f(x, z) \leq C(1 + |z|^2)$ for suitable $C \geq c > 0$). Then $X = L^2(\Omega)$, $Y = L^2(\Omega; \mathbb{R}^n)$,

$$G(v) = \int_{\Omega} f(x, v(x)) \, dx$$

and $A: u \in C^1(\Omega) \mapsto \nabla u \in L^2(\Omega; \mathbb{R}^n)$. It turns out that A is closable and that the domain of \bar{A} characterizes the Sobolev space $W^{1,2}(\Omega)$ on which \bar{A} coincides with the distributional gradient operator.

The situation is more involved if we consider

$$F(u) = \int_{\Omega} f(x, \nabla u) \, d\mu$$

μ is a possibly concentrated Radon measure supported on Ω . In general, the operator $A: u \in C^1(\Omega) \subset L^2_{\mu}(\Omega) \mapsto \nabla u \in L^2_{\mu}(\Omega; \mathbb{R}^n)$ is not closable and we need to come back to the general formula [3]. The general structure of $\overline{G(A)}$ has been given in Bouchitté *et al.* (1997) and Bouchitté and Fragalà (2002, 2003), namely

$$(u, \xi) \in \overline{G(A)} \iff u \in W_{\mu}^{1,2}, \exists \eta \in L^2_{\mu}(\Omega; \mathbb{R}^n): \xi = \nabla_{\mu} u + \eta, \eta(x) \in T_{\mu}(x)^{\perp}$$

where $T_{\mu}(x), \nabla_{\mu}(x)$ are suitable notions of tangent space and tangential gradient with respect to μ , and $W_{\mu}^{1,2}$ denotes the domain of the extended tangential gradient operator.

Remark 16 The assertion (ii) of Proposition 13 is not valid in the nonreflexive case. In particular, for

$$F(u) = \int_{\Omega} f(x, \nabla u) \, dx$$

where $f(x, \cdot)$ has a linear growth at infinity, we need to take Y as the space of \mathbb{R}^n -values vector measures on Ω and the relaxed functional F^{**} needs to be indentified on the space $BV(\Omega)$ of integrable functions with bounded variations. The computation of F^{**} is a delicate problem for which we refer to Bouchitté and Dal Maso (1993) and Bouchitté and Valadier (1998).

Remark 17 By duality techniques, it is possible also to handle variational integrals of the kind

$$F(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) \, dx$$

even if the dependence of $f(x, u, z)$ with respect to u is nonconvex. The idea consists in embedding the space $BV(\Omega)$ in the larger space $BV(\Omega \times \mathbb{R})$ through the map $u \mapsto 1_u$, where 1_u is the characteristic function defined on $\Omega \times \mathbb{R}$ by setting

$$1_u(x, t) := \begin{cases} 1 & \text{if } u(x) > t \\ 0 & \text{otherwise} \end{cases}$$

Then it is possible to show, under suitable conditions on the integrand f , that there exists a convex l.s.c., 1-homogeneous functional $G: BV(\Omega \times \mathbb{R}) \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $\bar{F}(u) = G(1_u)$. This functional G is constructed as in the Example 3 taking C to be a suitable convex subset of $C^0(\Omega \times \mathbb{R})$. This nice new idea has been the key tool of the calibration method developed recently (Alberti *et al.* 2003).

Convex Variational Problems in Duality

Finite-Dimensional Case

We sketch the duality scheme in two cases.

Linear programming Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m$ and A an $m \times n$ matrix. We denote by A^T the transpose matrix. We consider the linear program

$$(P) \quad \inf\{(c|x): x \geq 0, Ax \leq b\}$$

and its perturbed version ($p \in \mathbb{R}^m$)

$$h(p) := \inf\{(c|x): x \geq 0, Ax + p \leq b\}$$

An easy computation gives

$$\forall y \in \mathbb{R}^m, \quad h^*(y) = \begin{cases} -(b|y) & \text{if } A^T y + c \leq 0, y \geq 0 \\ +\infty & \text{otherwise} \end{cases} \quad [4]$$

Lemma 18 Assume that $\inf(P)$ is finite. Then:

- (i) h is convex proper and l.s.c. at 0.
- (ii) (P) has at least one solution.

Proof We introduce the $(n+m) \times (m+1)$ matrix B defined by

$$B := \begin{pmatrix} c^T & 0 \\ A & I_m \end{pmatrix}$$

(I_m is the m -dimensional identity matrix). Denote $\{b_1, b_2, \dots, b_{n+m}\} \subset \mathbb{R}^{m+1}$ the columns of B and K the convex cone $K := \{\sum_{j=1}^{n+m} \lambda_j b_j: \lambda_j \geq 0\}$. By Farkas lemma, this cone K is closed.

- (i) Let $\alpha := \liminf \{h(p): p \rightarrow 0\}$. We have to prove that $\alpha \geq h(0) = \inf P$. Let $\{p_\varepsilon\}$ be a sequence in \mathbb{R}^m such that $p_\varepsilon \rightarrow 0$ and $h(p_\varepsilon) \rightarrow \alpha$. By the definition of h , we may choose $x_\varepsilon \geq 0$ such that $Ax_\varepsilon \leq b$ and $(c|x_\varepsilon) \rightarrow \alpha$. Then we see that the column vector \tilde{x}_ε associated with $(x_\varepsilon, b - Ax_\varepsilon) \in \mathbb{R}^{n+m}$ satisfies: $B\tilde{x}_\varepsilon \in K$ and

$$B\tilde{x}_\varepsilon \rightarrow \begin{pmatrix} \alpha \\ b \end{pmatrix}$$

Therefore,

$$\begin{pmatrix} \alpha \\ b \end{pmatrix} \in K$$

and there exists $\tilde{x} = (x, x')$ such that $x \geq 0, x' \geq 0$, $(c|x) = \alpha$ and $Ax + x' = b$. It follows that x is admissible for (P) and then $(c|x) = \alpha \geq h(0)$.

- (ii) We repeat the proof of (i) choosing $p_\varepsilon = 0$ so that $\alpha = \inf(P)$. \square

Thanks to the assertion (i) in Lemma 18, we deduce from Theorem 10 that $\inf(P) = h(0) = h^{**}(0) =$

$\sup -b^*$. Recalling [4], we therefore consider the dual problem:

$$(P^*) \quad \sup\{-b \cdot y: y \geq 0, A^T y + c \geq 0\}$$

Theorem 19 The following assertions are equivalent:

- (i) (P) has a solution.
- (ii) (P^*) has a solution.
- (iii) There exists $(x_0, y_0) \in \mathbb{R}_+^n \times \mathbb{R}_+^m$ such that $Ax_0 \leq b, A^T y_0 + c \geq 0$.

In this case, we have $\min(P) = \max(P^*)$ and an admissible pair (\bar{x}, \bar{y}) is optimal if and only if $c \cdot \bar{x} = -b \cdot \bar{y}$ or, equivalently, satisfies the complementarity relations: $(A\bar{x} - b) \cdot \bar{y} = (A^T \bar{y} + c) \cdot \bar{x} = 0$.

Convex programming Let $f, g_1, \dots, g_m: X \rightarrow \mathbb{R}$ be convex l.s.c. functions and the optimization problem

$$(P) \quad \inf\{f(x): g_j(x) \leq 0, j = 1, 2, \dots, m\}$$

Here $X = \mathbb{R}^n$ or any Banach space. As before, we introduce the value function

$$p \in \mathbb{R}^m, \quad h(p) := \inf\{f(x): g_j(x) + p_j \leq 0, j = 1, 2, \dots, m\}$$

and compute its Fenchel conjugate:

$$\lambda \in \mathbb{R}^m, \quad h^*(\lambda) = \begin{cases} \inf_{x \in X} \{L(x, \lambda)\} & \text{if } \lambda \geq 0 \\ +\infty & \text{otherwise} \end{cases}$$

where $L(x, \lambda) := f(x) + \sum \lambda_j g_j(x)$ is the so-called Lagrangian. We notice that h is convex and that the equality $h(0) = h^{**}(0)$ is equivalent to the zero-duality gap relation

$$\inf_x \sup_\lambda L(x, \lambda) = \sup_\lambda \inf_x L(x, \lambda)$$

This condition is fulfilled, in particular, if we make the following qualification assumption (ensuring that h is continuous at 0 and Theorem 11 applies):

$$\exists x_0 \in X: f \text{ continuous at } x_0, g_j(x_0) < 0, \forall j \quad [5]$$

Theorem 20 Assume that [5] holds. Then \bar{x} is optimal for (P) if and only if there exist Lagrangian multipliers $\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_m$ in \mathbb{R}_+ such that

$$\bar{x} \in \operatorname{argmin}_X \left(f + \sum_j \bar{\lambda}_j g_j \right), \quad \bar{\lambda}_j g_j(\bar{x}) = 0, \quad \forall j$$

Notice that the existence of such a solution \bar{x} is ensured if, for example, $X = \mathbb{R}^n$ and if, for some $k > 0$, the function $f + k \sum_j g_j$ is coercive.

Primal–Dual Formulations in Mechanics

We present here the example of elasticity which motivated the pioneering work by J J Moreau on convex duality techniques. Further examples can be found in Ekeland and Temam (1976). An elastic body is placed in a bounded domain $\Omega \subset \mathbb{R}^n$ whose boundary Γ consists of two disjoint parts $\Gamma = \Gamma_0 \cup \Gamma_1$. The unknown $u : \Omega \rightarrow \mathbb{R}^n$ (deformation) satisfies a Dirichlet condition $u = 0$ on Γ_0 , where the body is clamped. The system is subjected to a surface load $g \in L^2(\Gamma_1; \mathbb{R}^n)$ and to a volumic load $f \in L^2(\Omega; \mathbb{R}^n)$. The static equilibrium problem has the following variational formulation:

$$(\mathcal{P}) \quad \inf_{u=0 \text{ on } \Gamma_0} \left\{ \int_{\Omega} j(x, e(u)) \, dx - \int_{\Omega} f \cdot u \, dx - \int_{\Gamma_1} g \cdot u \, d\mathcal{H}^{n-1} \right\}$$

where $e(u) := (1/2)(u_{i,j} + u_{j,i})$ denotes the symmetric strain tensor and $j : (x, z) \in \Omega \times \mathbb{R}_{\text{sym}}^{n^2} \rightarrow \mathbb{R}_+$ is a convex integrand representing the local elastic behavior of the material. We assume a quadratic growth as in Remark 15 (in the case of linear elasticity, an isotropic homogeneous material is characterized by the quadratic form

$$j(x, z) = \frac{\lambda}{2} |\text{tr}(z)|^2 + \mu |z|^2$$

λ, μ being the Lamé constants).

We apply Proposition 14 with $X = W^{1,2}(\Omega; \mathbb{R}^n)$, $Y = L^2(\Omega; \mathbb{R}_{\text{sym}}^{n^2})$, $Au = e(u)$ and where we set

$$\Phi(u) = \begin{cases} - \int_{\Omega} f \cdot u \, dx - \int_{\Gamma_1} g \cdot u \, d\mathcal{H}^{n-1} & \text{if } u = 0 \text{ on } \Gamma_0 \\ +\infty & \text{otherwise} \end{cases}$$

$$\Psi(v) = \int_{\Omega} j(x, v) \, dx$$

After some computations, we may write the supremum appearing in Proposition 14 as our dual problem

$$(\mathcal{P}^*) \quad \sup \left\{ - \int_{\Omega} j^*(x, \sigma) \, dx : \sigma \in L^2(\Omega; \mathbb{R}_{\text{sym}}^{n^2}), -\text{div } \sigma = f \text{ on } \Omega, \sigma \cdot n = g \text{ on } \Gamma_1 \right\}$$

where j^* is the Moreau–Fenchel conjugate with respect to the second argument and $n(x)$ denotes the exterior unit normal on Γ . The matrix-valued map σ is called the stress tensor and j^* the stress potential. Note that the boundary conditions for σn have to be understood in the sense of traces.

Theorem 21 *The problems (\mathcal{P}) and (\mathcal{P}^*) have solutions and we have the equality: $\inf(\mathcal{P}) = \sup(\mathcal{P}^*)$.*

Furthermore, a pair $(\bar{u}, \bar{\sigma})$ is optimal if and only if it satisfies the following system:

$$\begin{aligned} -\text{div } \bar{\sigma} &= f && \text{on } \Omega && (\text{equilibrium}) \\ \bar{\sigma}(x) &\in \partial j(x, e(\bar{u})) && \text{a.e. on } \Omega && (\text{constitutive law}) \\ \bar{u} &= 0 && \text{a.e. on } \Gamma_0 \\ \bar{\sigma} n &= g && \text{on } \Gamma_1 \end{aligned}$$

Duality in Mass Transport Problems

General Cost Functions

Let X, Y be a compact metric space and $c : X \times Y \rightarrow [0, +\infty)$ a continuous cost function. We denote by $\mathcal{P}(X), \mathcal{P}(X \times Y)$ the sets of probability measures on X and $X \times Y$, respectively. Given two elements $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, we denote by $\Gamma(\mu, \nu)$ the subset of probability measures in $\mathcal{P}(X \times Y)$ whose marginals are, respectively, μ and ν . Identified as a subset of $(C^0(X \times Y))^*$ (the space of signed Radon measures on $X \times Y$), it is convex and weakly-star compact. The Monge–Kantorovich formulation of the mass transport problem reads as follows:

$$T_c(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x, y) \gamma(dx dy) : \gamma \in \Gamma(\mu, \nu) \right\} \quad [6]$$

This formulation, where the infimum is achieved (as we minimize an l.s.c. functional on a compact set for the weak star topology), is already a relaxation of the initial Monge mass transport problem,

$$\inf_T \left\{ \int_X c(x, Tx) \mu(dx) : T^\#(\mu) = \nu \right\}$$

where the infimum is searched among all transports maps $T : X \mapsto Y$ pushing forward μ on ν (i.e., such that $\mu(T^{-1}(B)) = \nu(B)$ for all Borel subset $B \subset Y$). This is equivalent to restricting the infimum in [6] to the subclass $\{\gamma_T\} \subset \Gamma(\mu, \nu)$, where

$$\langle \gamma_T, \phi(x, y) \rangle := \int_X \phi(x, Tx) \mu(dx)$$

In order to find a dual problem for [6], we fix $\nu \in \mathcal{P}(Y)$ and consider the functional $F : \mathcal{M}_b(X) \rightarrow [0, +\infty)$ defined by

$$F(\mu) = \begin{cases} T_c(\mu, \nu) & \text{if } \mu \geq 0, \mu(X) = 1 \\ +\infty & \text{otherwise} \end{cases}$$

($\mathcal{M}_b(X)$ denote the Banach space of (bounded) signed Radon measures on X).

Lemma 22 *F is convex, weakly-star l.s.c. and proper. Its Moreau–Fenchel conjugate is given by*

$$\forall \varphi \in C^0(X), \quad F^*(\varphi) = - \int_Y \varphi^c(y) \nu(dy)$$

where

$$\varphi^c(y) := \inf\{c(x, y) - \varphi(x) : x \in X\}$$

Proof The convexity property is obvious and the properness follows from the fact that

$$F(\mu) \leq \int_{X \times Y} c(x, y) \mu \otimes \nu(dx dy)$$

Let μ_n be such that $\mu_n \rightharpoonup \mu$ (weakly star). We may assume that $\liminf_n F(\mu_n) = \lim_n F(\mu_n) := \alpha$ is finite. Then μ_n and the associated optimal γ_n are probability measures on X and on $X \times Y$, respectively. As X and Y are compact, possibly passing to a subsequence, we may assume that $\gamma_n \rightharpoonup \gamma$, and clearly we have $\gamma \in \Gamma(\mu, \nu)$. Since $c(x, y)$ is l.s.c. non-negative, we conclude that

$$\begin{aligned} \liminf_n F(\mu_n) &= \liminf_n \int_{X \times Y} c(x, y) \gamma_n(dx dy) \\ &\geq \int_{X \times Y} c(x, y) \gamma(dx dy) \\ &= F(\mu) \end{aligned}$$

Let us compute now $F^*(\varphi)$. We have

$$\begin{aligned} -F^*(\varphi) &= \inf \left\{ \int_{X \times Y} c(x, y) \gamma(dx dy) \right. \\ &\quad \left. - \int_X \varphi d\mu : \mu \in \mathcal{P}(X), \gamma \in \Gamma(\mu, \nu) \right\} \\ &= \inf \left\{ \int_{X \times Y} (c(x, y) - \varphi(x)) \gamma(dx dy) : \right. \\ &\quad \left. \gamma \in \Gamma(\mu, \nu) \right\} \\ &\geq \int_Y \varphi^c(y) \nu(dy) \end{aligned}$$

To prove that the last inequality is actually an equality, we observe that, for every $y \in Y$ and $\varphi \in C^0(X)$, the minimum of the l.s.c. function $c(\cdot, y) - \varphi$ is attained on the compact set X and there exists a Borel selection map $S(y)$ such that $\varphi^c(y) = c(S(y), y) - \varphi(S(y))$ for all $y \in Y$. We obtain the desired equality by choosing γ defined, for every test ψ , by

$$\int_{X \times Y} \psi(x, y) \gamma(dx dy) := \int_Y \psi(S(y), y) \nu(dy)$$

□

We observe that, for every $\varphi \in C^0(X)$, the function φ^c introduced in Lemma 22 is continuous (use the uniform continuity of c) and therefore the pair (φ, φ^c) belong to the class

$$\mathcal{F}_c := \{(\varphi, \Psi) \in C^0(X) \times C^0(Y) : \varphi(x) + \psi(y) \leq c(x, y)\}$$

Let us introduce the dual problem of [6]:

$$\sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu : (\varphi, \psi) \in \mathcal{F}_c \right\} \quad [7]$$

We will say that $(\varphi, \psi) \in \mathcal{F}_c$ is a pair of c -concave conjugate functions if $\psi = \varphi^c$ and $\varphi = \psi^c$ (where symmetrically $\psi^c(x) := \inf\{c(x, y) - \psi(y) : y \in Y\}$). Checking the latter condition amounts to verifying that φ enjoys the so-called c -concavity property $\varphi^{cc} = \varphi$ (in general, we have only $\varphi^{cc} \geq \varphi$, whereas $\varphi^{ccc} = \varphi^c$). We refer for instance to Villani (2003) for further details about this c -duality.

Now, by exploiting Theorem 10 and Lemma 22, we obtain a very simple proof of Kantorovich duality theorem:

Theorem 23 *The following duality formula holds:*

$$T_c(\mu, \nu) = \sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu : (\varphi, \psi) \in \mathcal{F}_c \right\}$$

Moreover, the supremum in the right-hand side member is achieved by a pair $(\bar{\varphi}, \bar{\psi})$ of conjugate c -concave functions such that, for any optimal $\bar{\gamma}$ in [6], there holds $\bar{\varphi}(x) + \bar{\psi}(y) = c(x, y)$, $\bar{\gamma}$ -a.e.

Proof By Theorem 10 and Lemma 22, we have

$$\begin{aligned} T_c(\mu, \nu) &= F^{**}(\mu) \\ &= \sup \left\{ \int_X \varphi d\mu + \int_Y \varphi^c d\nu : \varphi \in C^0(X) \right\} \\ &\leq \sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu : (\varphi, \psi) \in \mathcal{F}_c \right\} \\ &\leq T_c(\mu, \nu) \end{aligned}$$

where the last inequality follows from the definition of \mathcal{F}_c . Therefore, $\inf [6] = \sup [7]$. Furthermore, on the right-hand side of first equality, we increase the supremum by substituting φ with φ^{cc} (recall that $\varphi^{ccc} = \varphi^c$). Thus,

$$\begin{aligned} \sup [7] &= \sup \left\{ \int_X \varphi d\mu + \int_Y \varphi^c d\nu : \varphi \in C^0(X), \right. \\ &\quad \left. \varphi \text{ } c\text{-concave} \right\} \end{aligned}$$

Take a maximizing sequence (φ_n, φ_n^c) of c -concave conjugate functions. It is easy to check that $\{f_n\}$ is equicontinuous on X : this follows from the c -concavity property and from the uniform continuity of c (observe that $\varphi_n(x_1) - \varphi_n(x_2) = \varphi_n^{cc}(x_1) - \varphi_n^{cc}(x_2) \leq \sup_Y \{c(x_1, \cdot) - c(x_2, \cdot)\}$). Then, by Ascoli's theorem, possibly passing to subsequences, we may assume that: $\varphi_n - c_n$ converges uniformly to some continuous function $\bar{\varphi}$ where $\{c_n\}$ is a suitable sequence of reals. Then, one checks that $\bar{\varphi}$ is still c -concave and that $(\varphi_n - c_n)^c = \varphi_n^c + c_n$ converges uniformly to

$\bar{\varphi}^c$. Thus, recalling that $\mu(X) = \nu(Y) = 1$, we deduce that

$$\begin{aligned} \sup[7] &= \lim_n \left(\int_X \varphi_n d\mu + \int_Y \varphi_n^c d\nu \right) \\ &= \lim_n \left[\int_X (\varphi_n - c_n) d\mu + \int_Y (\varphi_n^c + c_n) d\nu \right] \\ &= \int_X \bar{\varphi} d\mu + \int_Y \bar{\varphi}^c d\nu \end{aligned}$$

The last assertion is a consequence of the extremality relation:

$$\begin{aligned} 0 &= \inf[6] - \sup[7] \\ &= \int_{X \times Y} (c(x, y) - \bar{\varphi}(x) - \bar{\psi}(y)) \bar{\gamma}(dx dy) \end{aligned}$$

□

Remark 24

- (i) In their discrete version (i.e., μ, ν are atomic measures), problems [6] and [7] can be seen as particular linear programming problems (see the section “Finite-dimensional case”).
- (ii) The case $X = Y \subset \mathbb{R}^n$ and $c(x, y) = (1/2)|x - y|^2$ is important. In this case, the notion of c -concavity is linked to convexity and the Fenchel transform since, for every $\varphi \in C^0(X)$, one has

$$\frac{|\cdot|^2}{2} - \varphi^c = \left(\frac{|\cdot|^2}{2} - \varphi \right)^*$$

Then if $(\bar{\varphi}, \bar{\varphi}^c)$ is a solution of [7], we find that

$$\varphi_0(x) := \frac{|x|^2}{2} - \bar{\varphi}(x)$$

is convex continuous and that the extremality condition: $\bar{\varphi}(x) + \bar{\varphi}^c(y) = c(x, y)$ is equivalent to Fenchel equality $\varphi_0(x) + \varphi_0^*(y) = (x|y)$. Therefore, any optimal $\bar{\gamma}$ is supported in the graph of the subdifferential map $\partial\varphi_0$. In the case where μ is absolutely continuous with respect to the Lebesgue measure, it is then easy to deduce that the optimal $\bar{\gamma}$ is unique and that $\bar{\gamma} = \gamma_{T_0}$, where $T_0 = \nabla\varphi_0$ is the unique gradient (a.e. defined) of a convex function such that $\nabla\varphi_0^\#(\mu) = \nu$. This is a celebrated result by Y Brenier (see, e.g., the monographs by Evans (1997) and Villani (2003)).

The Distance Case

In the following, we assume that $X = Y$ and that $c(x, y)$ is a semidistance. As an immediate

consequence of the triangular inequality, we have the following equivalence:

$$\begin{aligned} \varphi \text{ } c\text{-concave} &\Leftrightarrow \varphi(x) - \varphi(y) \leq c(x, y), \quad \forall(x, y) \\ &\Leftrightarrow \varphi^c = -\varphi \end{aligned}$$

Let us denote $\text{Lip}_1(X) := \{u \in C^0(X) : u(x) - u(y) \leq c(x, y)\}$. The first assertion of Theorem 23 becomes the Kantorovich–Rubintein duality formula:

$$T_c(\mu, \nu) = \max \left\{ \int_X u d(\mu - \nu) : u \in \text{Lip}_1(X) \right\} \quad [8]$$

As it appears, $T_c(\mu, \nu)$ depends only on the difference $f = \mu - \nu$, which belongs to the space $\mathcal{M}_0(X)$ of signed measure on X with zero average. Defining $N(f) := T_c(f^+, f^-)$ provides a seminorm (Kantorovich norm) on $\mathcal{M}_0(X)$ (it turns out that $\mathcal{M}_0(X)$ is not complete and that in general its completion is a strict subspace of the dual of $\text{Lip}(X)$).

We will now specialize to the case where X is a compact manifold equipped with a geodesic distance. This will allow us to link the original problem to another primal–dual formulation closer to that considered in the section “Primal–dual formulation in mechanics” and yielding to a connection with partial differential equations. As a model example, let us assume that $K = \bar{\Omega}$, where Ω is a bounded connected open subset of \mathbb{R}^n with a Lipschitz boundary. Let $\Sigma \subset \bar{\Omega}$ be a compact subset (on which the transport will have zero cost) and define

$$\begin{aligned} c(x, y) &:= \inf \{ \mathcal{H}^1(S \setminus \Sigma) : \\ &\quad S \text{ Lipschitz curve joining } x \text{ to } y, S \subset \bar{\Omega} \} \quad [9] \end{aligned}$$

where \mathcal{H}^1 denotes the one-dimensional Hausdorff measure (length). It is easy to check that

$$c(x, y) = \min \{ \delta_\Omega(x, y), \delta_\Omega(x, \Sigma) + \delta_\Omega(y, \Sigma) \}$$

where $\delta_\Omega(x, y)$ is the geodesic distance on Ω (induced by the Euclidean norm). Furthermore, the following characterization holds:

$$\begin{aligned} u \in \text{Lip}_1(X) &\Leftrightarrow u \in W^{1,\infty}(\Omega), \\ |\nabla u| &\leq 1 \text{ a.e. in } \Omega, \quad u = \text{cte on } \Sigma \quad [10] \end{aligned}$$

Since $f := \mu - \nu$ is balanced, the value of the constant on Σ in [10] is irrelevant and can be set to 0. Thus we may rewrite the right hand side member of [8] in a equivalent way as

$$\begin{aligned} \max \left\{ \int_\Omega u df : u \in W^{1,\infty}(\Omega), \right. \\ \left. |\nabla u| \leq 1 \text{ a.e. on } \Omega, \quad u = 0 \text{ on } \Sigma \right\} \quad [11] \end{aligned}$$

We will now derive a new dual problem for [11] by using Proposition 14. To this aim, we consider

$X = C^1(\bar{\Omega})$ (as a closed subspace of $W^{1,\infty}(\Omega)$), $Y = C^0(\bar{\Omega}; \mathbb{R}^n)$, $Y^* = \mathcal{M}_b(\bar{\Omega}; \mathbb{R}^n)$ and the operator $A : u \in X \mapsto \nabla u \in Y$.

Theorem 25 *Let $\mu, \nu \in \mathcal{P}(\bar{\Omega})$, $f = \mu - \nu$ and c defined by [9]. Then,*

$$T_c(\mu, \nu) = \min \left\{ \int_{\bar{\Omega}} |\lambda| : \lambda \in \mathcal{M}_b(\bar{\Omega}; \mathbb{R}^n), \right. \\ \left. -\operatorname{div} \lambda = f \text{ on } \bar{\Omega} \setminus \Sigma \right\} \quad [12]$$

where the divergence condition is intended in the sense that

$$\int_{\bar{\Omega}} \lambda \cdot \nabla \varphi = \int_{\bar{\Omega}} \varphi \, df$$

for all $\varphi \in C^\infty$ compactly supported in $\mathbb{R}^n \setminus \Sigma$.

Proof (sketch) We apply Proposition 14 with $\phi(u) = -\int_{\bar{\Omega}} u \, df$ if $u = 0$ on Σ ($+\infty$ otherwise), $A = \nabla$, and $\psi(v) = 0$ if $|v| \leq 1$ on $\bar{\Omega}$ ($+\infty$ otherwise). We obtain that the minimum α in [12] is reached and that $\alpha = \beta$, where

$$-\beta := \inf \left\{ -\int_{\bar{\Omega}} u \, df : u \in C^1(\bar{\Omega}), \right. \\ \left. |\nabla u| \leq 1 \text{ on } \Omega \ u = 0 \text{ on } \Sigma \right\}$$

To prove that $\beta = T_c(\mu, \nu) = \sup$ (11), we consider a maximizer \bar{u} in [11] and prove that it can be approximated uniformly by a sequence $\{u_n\}$ of functions in $C^1(\bar{\Omega})$ which satisfy the same constraints. This technical part is done by truncation and convolution arguments (we refer to Bouchitté et al. (2003) for details). \square

Remark 26 By localizing the integral identity associated with [12], it is possible to deduce the optimality conditions which characterize optimal pairs $(\bar{u}, \bar{\lambda})$ for [11], [12] (without requiring any regularity). This is done by using a weak notion of tangential gradient with respect to a measure (see Bouchitté et al. (1997) and Bouchitté and Fragalà (2002)). If $\bar{\lambda} = \bar{\sigma} \, dx$ where $\bar{\sigma} \in L^1(\Omega; \mathbb{R}^n)$ and if $\Sigma \subset \partial\Omega$, then we find that $\bar{\sigma} = a \nabla \bar{u}$, where the pair (\bar{u}, a) solves the following system:

$$\begin{aligned} -\operatorname{div}(a \nabla \bar{u}) &= f \text{ on } \Omega && \text{(diffusion equation)} \\ |\nabla \bar{u}| &= 1 \text{ a.e. on } \{a > 0\} && \text{(eikonal equation)} \\ u &= 0 \text{ a.e. on } \Sigma \\ \frac{\partial u}{\partial n} &= 0 \text{ on } \Sigma \end{aligned}$$

Remark 27 Given a solution $\bar{\gamma}$ for [6], we can construct a solution $\bar{\lambda}$ for [12] by selecting for every $(x, y) \in \operatorname{spt}(\bar{\gamma})$ a geodesic curve S_{xy} joining x and y (possibly passing through the free-cost zone Σ) and by setting, for every test ϕ :

$$\langle \bar{\lambda}, \phi \rangle := \int_{\bar{\Omega} \times \bar{\Omega}} \left(\int_{S_{xy}} \phi \cdot \tau_{S_{xy}} \, d\mathcal{H}^1 \right) \bar{\lambda}(dx dy)$$

where $\tau_{S_{xy}}$ denote the unit oriented tangent vector (see Bouchitté and Buttazzo (2001)). It is also possible to show (see Ambrosio (2003)) that any solution $\bar{\lambda}$ can be represented as before through a particular solution $\bar{\gamma}$. As a consequence, the support of any solution $\bar{\gamma}$ of [12] is supported in the geodesic envelope of the set $\operatorname{spt}(\mu) \cup \operatorname{spt}(\nu) \cup \Sigma$. However, we stress the fact that, in general, there is no uniqueness at all of the optimal triple $(\bar{\gamma}, \bar{u}, \bar{\lambda})$ for [6], [11] and [12].

Remark 28 An approximation procedure for particular solutions of problems [11], [12] can be obtained by solving a p -Laplace equation and then by sending p to infinity. Precisely, consider the solution $u_p \in W^{1,p}(\Omega)$ of

$$\begin{aligned} -\operatorname{div}(|\nabla u|^{p-2} \nabla u) &= f \text{ on } \bar{\Omega} \setminus \Sigma \\ u &= 0 \text{ on } \Sigma \end{aligned}$$

which, for $p > n$, exists (due to the compact embedding $W^{1,p}(\Omega) \subset C^0(\bar{\Omega})$) and is unique. In Bouchitté et al. (2003) it is proved that the sequence $\{(u_p, \sigma_p)\}$, where $\sigma_p = |\nabla u_p|^{p-2} \nabla u_p$, is relatively compact in $\mathcal{M}_b(\bar{\Omega}; \mathbb{R}^n) \times C^0(\bar{\Omega})$ (weakly star with respect to the first component) and that every cluster point $(\bar{u}, \bar{\lambda})$ solves [11], [12]. It is an open problem to know whether or not such a cluster point is unique. If the answer is “yes,” the process described above would select one optimal pair among all possible solutions. As far as problem [11] is concerned, this problem is connected with the theory of viscosity solutions for the infinite Laplacian (see Evans (1997)) although this theory does not provide an answer as it erases the role of the source term f . On the other hand, a new entropy selection principle should be found for the solutions of dual problem [12]. In fact, the following partial result holds: let $E : \mathcal{M}_b(\bar{\Omega}; \mathbb{R}^n) \rightarrow \mathbb{R} \cup \{+\infty\}$ be the functional defined by

$$E(\lambda) := \begin{cases} \int_{\bar{\Omega}} |\sigma| \log(|\sigma|) \, dx & \text{if } \lambda \ll dx \text{ and } \sigma = \frac{d\lambda}{d|\lambda|} \\ +\infty & \text{otherwise} \end{cases}$$

Assume that [12] admits at least one solution λ_0 such that $E(\lambda_0) < +\infty$. Then it can be shown that

the sequence $\{\sigma_p\}$ does converge weakly-star to $\bar{\lambda}$, the unique minimizer of the problem

$$\inf\{E(\lambda): \lambda \text{ solution of [12]}\}$$

The general case, in particular when all optimal measures are singular, is open.

Remark 29 Variational problems [11], [12] have important counterparts in the theory of elasticity and in optimal design problems (see Bouchitté and Buttazzo (2001)). They read, respectively, as

$$\begin{aligned} \max \left\{ \int_{\Omega} u \cdot df: u \in \cap_{p>1} W^{1,p}(\Omega; \mathbb{R}^n), \right. \\ \left. \nabla u(x) \in K \text{ a.e. on } \Omega, u = 0 \text{ on } \Sigma \right\} \\ \min \left\{ \int_{\bar{\Omega}} \rho_K^0(\lambda): \lambda \in \mathcal{M}_b(\bar{\Omega}; \mathbb{R}_{\text{sym}}^{n^2}), \right. \\ \left. -\operatorname{div} \lambda = f \text{ on } \bar{\Omega} \setminus \Sigma \right\} \end{aligned}$$

where $K \subset \mathbb{R}_{\text{sym}}^{n^2}$ is a convex compact subset of symmetric second-order tensors associated with the elastic material, $\rho_K^0(\xi) = \sup\{\xi \cdot z: z \in K\}$ is convex positively 1-homogeneous and the functional on measures $\int_{\bar{\Omega}} \rho_K^0(\lambda)$ is intended in the sense given in [1]. A celebrated example is given by Michell's problem (Michell 1904) where $n=2$ and $K := \{z \in \mathbb{R}_{\text{sym}}^{n^2}, |\rho(z)| \leq 1\}$, $\rho(z)$ being the largest singular value of z . The potential ρ_K^0 is given by the nondifferentiable convex function $\rho_K^0(\xi) = \tau_1(\xi) + \tau_2(\xi)$, where the $\tau_i(\xi)$'s are the singular values of ξ .

Unfortunately, it is not known if the vector variational problem above can be linked to an optimal transportation problem of the type [6], even if the analogous of equivalence [10] does exist in the Michell's case, namely (for Ω convex):

$$\begin{aligned} \rho(e(u)) \leq 1 \quad \text{on } \Omega \\ \iff |(u(x) - u(y)) \cdot (x - y)| \leq |x - y|^2, \quad \forall (x, y) \end{aligned}$$

Further Reading

- Alberti G, Bouchitté G, and Dal Maso G (2003) The calibration method for the Mumford–Shah functional and free-discontinuity problems. *Calculus of Variations and Partial Differential Equations* 16(3): 299–333.
- Ambrosio L (2003) Lecture notes on optimal transport problems. In: *Mathematical Aspects of Evolving Interfaces (Funchal 2000)*, Lecture Notes in Mathematics, vol. 1812, pp. 1–52. Berlin: Springer.
- Borwein M and Lewis SA (2000) *Convex Analysis and Nonlinear Optimization. Theory and Examples*, CMS Series. Berlin: Springer.
- Bouchitté G and Buttazzo G (2001) Characterization of optimal shapes and masses through Monge–Kantorovich equations. *Journal of the European Mathematical Society* 3: 139–168.
- Bouchitté G, Buttazzo G, and De Pascale L (2003) A p-Laplacian approximation for some mass optimization problems. *Journal of Optimization Theory and Applications* 118: 1–25.
- Bouchitté G, Buttazzo G, and Seppecher P (1997) Energies with respect to a measure and applications to low dimensional structures. *Calculus of Variations and Partial Differential Equations* 5: 37–54.
- Bouchitté G and Dal Maso G (1993) Integral representation and relaxation of convex local functionals on $BV(\Omega)$. *Annali della Scuola Superiore di Pisa* 20(4): 483–533.
- Bouchitté G and Fragalà I (2002) Variational theory of weak geometric structures: the measure method and its applications. *Variational Methods for Discontinuous Structures, Ser. PNLDE*, vol. 51, pp. 19–40. Basel: Birkhäuser.
- Bouchitté G and Fragalà I (2003) Second order energies on thin structures: variational theory and non-local effects. *Journal of Functional Analysis* 204(1): 228–267.
- Bouchitté G and Valadier M (1988) Integral representation of convex functionals on a space of measures. *Journal of Functional Analysis* 80: 398–420.
- Ekeland I and Temam R (1976) *Analyse convexe et problèmes variationnels*. Paris: Dunod-Gauthier Villars.
- Evans LC (1997) Partial differential equations and Monge–Kantorovich mass transfer. In: Bott R, Jaffe A, Jerison D, Lutsztig G, Singer I, and Yau JT (eds.) *Current Developments in Mathematics*, pp. 65–126. Cambridge.
- Michell AGM (1904) The limits of economy of material in frame structures. *Philosophical Magazine and Journal of Science* 6: 589–597.
- Rockafellar RT (1970) *Convex Analysis*. Princeton: Princeton University Press.
- Villani C (2003) *Topics in Optimal transportation*, Graduate studies in Mathematics, vol. 58. Providence, RI: AMS.

Cosmology: Mathematical Aspects

G F R Ellis, University of Cape Town,
Cape Town, South Africa

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Mathematical cosmology focuses on the geometrical and mathematical aspects of the study of the universe as a whole. Because the structure of spacetime (with metric tensor $g_{ab}(x^i)$) is governed by gravity, with matter and energy causing spacetime curvature according to the nonlinear gravitational field equations of the theory of general relativity, it has its roots in differential geometry. It is to be distinguished from the three other major aspects of modern cosmology, namely astrophysical cosmology, high-energy physics cosmology, and observational cosmology; see [Peacock \(1999\)](#) for these aspects.

The Einstein field equations (EFEs) are

$$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda g_{ab} = \kappa T_{ab} \quad [1]$$

where R_{ab} is the Ricci tensor, R the Ricci scalar, T_{ab} the matter tensor, Λ the cosmological constant, and κ the gravitational constant. Cosmological models differ from generic solutions of these equations in that they have preferred world lines in spacetime associated with the motion of matter and distribution of radiation ([Ellis 1971](#)). This is a classic case of a broken symmetry: the underlying equations [1] are locally Lorentz invariant but their solutions are not. These preferred world lines, characterized by a unit 4-velocity vector u^a , are associated at late times with “fundamental observers,” and a key aspect of cosmological modeling is determining the observational relations such observers would determine through astronomical observations.

The dynamics of cosmological models is determined by their matter content. This is usually represented in simplified form, often using the “perfect-fluid” approximation to represent the effect of matter or radiation; that is,

$$T_{ab} = (\rho + p)u_a u_b + pg_{ab} \quad [2]$$

where ρ is the energy density and p the pressure, and the matter 4-velocity u_b is the preferred cosmological 4-velocity. This description can include a scalar field ϕ with dynamics governed by the Klein–Gordon equation, provided u_a is normal to spacelike surfaces $\{\phi = \text{const}\}$. Suitable equations of state describe the nature of the matter envisaged (e.g., $p=0$ for baryons, whereas $p=\rho/3$ for

radiation); in the case of a scalar field with potential $V(\phi)$ and spacelike surfaces $\{\phi = \text{const}\}$, on choosing u^a orthogonal to these surfaces, the stress tensor has a perfect-fluid form with $\rho = (1/2)\dot{\phi}^2 + V(\phi)$, $p = (1/2)\dot{\phi}^2 - V(\phi)$. A cosmological constant Λ can be represented as a perfect fluid with $\rho + p = 0$, $\Lambda = p$. More general matter may involve a momentum flux density q_a and anisotropic pressures π_{ab} ([Ehlers 1961](#)). Whatever the nature of the matter, it will usually be required to satisfy energy conditions ([Hawking and Ellis 1973](#)). All realistic matter has a positive inertial mass density:

$$\rho + p > 0 \quad [3]$$

(note that realistic cosmological models are non-empty), whereas all ordinary matter has a positive gravitational mass density:

$$\rho + 3p > 0 \quad [4]$$

but this is not necessarily true for a scalar field or effective cosmological constant.

Mathematical cosmology ([Ellis and van Elst 1999](#)) studies (1) generic properties of solutions with a preferred 4-velocity field and matter content as indicated above, (2) the standard FLRW models, (3) approximate FLRW solutions, and (4) other exact and approximate cosmological solutions. The ultimate underlying issue is (5) the origin of the universe. We look at these in turn. We aim to use covariant methods as far as possible, to avoid being misled by coordinate effects, and to obtain exact solutions and exact results as far as possible, because approximate methods can be misleading in the case of these nonlinear field equations.

Exact Properties

We can split the equations into spacelike and timelike parts relative to the 4-velocity u^a , obtaining the (1 + 3) covariant dynamical equations and identities in terms of the fluid shear σ_{ab} , vorticity ω_{ab} , expansion $\Theta = u^a{}_{;a}$, and acceleration $a^b = u^a{}_{;b}u^b$ ([Ehlers 1961](#), [Ellis 1971](#), [Ellis and van Elst 1999](#)). The energy density of a perfect fluid obeys the conservation equation

$$\dot{\rho} = -3(\rho + p)\frac{\dot{\Sigma}}{\Sigma} \quad [5]$$

with extra terms occurring in the case of more complex matter. From the momentum equations, pressure-free solutions are geodesic ($a^b = 0$). The crucial Raychaudhuri–Ehlers equation for the

time derivative of the expansion (Ehlers 1961) can be written as

$$3\frac{\ddot{S}}{S} = 2(\omega^2 - \sigma^2) + a_{;b}^b - \frac{\kappa}{2}(\rho + 3p) + \Lambda \quad [6]$$

where the representative length scale S is defined by $\Theta = 3\dot{S}/S$. This is the basis of the “fundamental singularity theorem”: if in an expanding universe $\omega = 0 = a^b$ and the combined matter present satisfies [4], with $\Lambda \leq 0$, then there was a singularity where $S \rightarrow 0$ a finite time $t_0 < 1/H_0$ ago, $H_0 = (\dot{S}/S)_0$ being the present value of the Hubble constant. The energy density will diverge there, so this is a spacetime singularity: an origin of physics, matter, and spacetime itself. However, the deduction does not follow if there is rotation or acceleration, which could conceivably avoid the singularity, so this result is by itself inconclusive for realistic cosmologies.

The vorticity obeys conservation laws analogous to those in Newtonian theory (Ehlers 1961). Vorticity-free solutions ($\omega = 0$) occur whenever the fluid flow lines are hypersurface-orthogonal in spacetime, that is, there exists a cosmic time function for the comoving observers, which will measure proper time along the flow lines if additionally the fluid flow is geodesic. The rate of change of shear is related to the conformal curvature (Weyl) tensor, which represents the free gravitational field, and which splits into an electric part E_{ab} and a magnetic part H_{ab} in close analogy with electromagnetic theory. Shear-free solutions ($\sigma = 0$) are very special because they strongly constrain the Weyl tensor; indeed if the flow is shear free and geodesic, then it either does not expand ($\Theta = 0$), or does not rotate ($\omega = 0$) (Ellis 1967). The set of cosmological observations associated with generic cosmological models has been characterized in power series form by Kristian and Sachs (1966), and that result has been extended to general models by Ellis *et al.* (1985).

The local regularity of the theory is expressed in existence and uniqueness theorems for the EFEs, provided the matter behavior is well defined through prescription of suitable equations of state (Hawking and Ellis 1973). However, in general the theory breaks down in the large, and this feature is specified by the Hawking–Penrose singularity theorems, predicting the existence of a geodesic incompleteness of spacetime under conditions applicable to realistic cosmological models satisfying the energy conditions given by eqns [3] and [4] (Hawking and Ellis 1973, Tipler *et al.* 1980). However, the conclusion does not follow if the energy conditions are not satisfied. Furthermore, the deduction follows

only if the gravitational field equations remain valid to arbitrarily early times; but we would in fact expect that, at high enough energy densities, quantum gravity would take over from classical gravity, so whether or not there was indeed a singularity would depend on the nature of the as yet unknown theory of quantum gravity. The cash value of the singularity theorems then is the implication that, when the energy conditions are satisfied, one would indeed be involved in such a quantum gravity realm in the very early universe.

The Standard Friedmann–Lemaître Models

The standard models of cosmology are the Friedmann–Lemaître (FL) models with Robertson–Walker (RW) geometry: that is, they are exactly spatially homogeneous and locally isotropic, invariant under a G_6 of isometries (Robertson 1933, Ehlers 1961). They have a unique cosmic time function t , with space sections $\{t = \text{const.}\}$ of constant spatial curvature orthogonal to the uniquely preferred 4-velocity u^a . The fluid acceleration, vorticity, and shear all vanish, and all physical quantities depend only on the time coordinate t . They can be represented by a metric with scale factor $S(t)$:

$$\begin{aligned} ds^2 &\equiv g_{ab} dx^a dx^b \\ &= -dt^2 + S^2(t) \{ dr^2 + f^2(r) (d\theta^2 + \sin^2 \theta d\phi^2) \} \end{aligned} \quad [7]$$

in comoving coordinates $(x^a) = (t, r, \theta, \phi)$, where $f(r) = \{\sin r, r, \sinh r\}$ if $\{k = +1, 0, -1\}$, and the matter is a perfect fluid with 4-velocity vector $u^a = dx^a/ds = \delta_0^a$. The curvature of the space sections $\{t = \text{const.}\}$ is $K = k/S^2$; these 3-spaces are necessarily closed (compact) if they are positively curved ($k = +1$), but may be open or closed in the flat ($k = 0$) and negatively curved ($k = -1$) cases, depending on their topology (Lachieze-Rey and Luminet 1995).

Matter obeys the conservation equation [5], whose outcome depends on the equation of state; for baryons $\rho = M/S^3$, whereas for radiation $\rho = M/S^4$, where M is a constant. The dynamics of the models is governed by the Raychaudhuri equation

$$3\frac{\ddot{S}}{S} = -\frac{\kappa}{2}(\rho + 3p) + \Lambda \quad [8]$$

which has the Friedmann equation

$$\frac{3\dot{S}^2}{S^2} = \kappa\rho + \Lambda - \frac{3k}{S^2} \quad [9]$$

as a first integral whenever $\dot{S} \neq 0$. Depending on the matter components present, one can qualitatively

characterize the dynamical behavior of these models (Robertson 1933) and find exact and approximate solutions to these equations as well as phase planes representing the relation of the different models to each other; for example, Ehlers and Rindler (1989) give the phase planes for models with noninteracting matter and radiation and an arbitrary cosmological constant. Universes with maxima or minima in $S(t)$ can only occur if $k = +1$; when $\Lambda = 0$, the universe recollapses in the future iff $k = +1$. Static solutions are possible only if $k = +1$ and (assuming [4]) $\Lambda > 0$. The simplest expanding solutions are the Einstein–de Sitter universes with $k = 0 = \Lambda$.

Equation [8] is a special case of [6], with corresponding implications: if the combined matter present satisfies [4], with $\Lambda \leq 0$, then there must have been an initial singularity, or at least the universe must have emerged from a quantum gravity domain. The temperature would have been arbitrarily high in the past, so there was a hot big bang era in the early universe where matter and radiation were in equilibrium with each other at very high temperatures that rapidly fell as the universe expanded. Many physical processes took place then, in particular nucleosynthesis of light elements took place at $\sim 10^9$ K. Decoupling of matter and radiation took place at a temperature of ~ 4000 K, followed by formation of stars and galaxies (see Peacock (1999) for a discussion of these physical processes). The black-body radiation emitted by the surface of last scattering at 4000 K is observed by us today as cosmic black-body radiation (CBR) at a temperature of 2.75 K.

One can determine observational relations for these models such as the magnitude–redshift relation for “standard candles” at recent times from the EFEs (Sandage 1961). The aim of observations is to determine the Hubble constant H_0 , dimensionless deceleration parameter $q_0 = -(3/H_0^2)(\dot{S}/S)_0$, and normalized density parameters $\Omega_{0i} = \kappa_i \rho_{0i}/3H_0^2$ for each component of matter present. The spatial curvature and the cosmological constant then follow from [6] and [9]; also the present scale factor S_0 is determined if $k \neq 0$. The universe is of positive spatial curvature ($k = +1$) iff $\Omega_0 \equiv \Omega_m + \Omega_\Lambda > 1$, where $\Omega_m \equiv \sum_i \Omega_{0i}$, $\Omega_\Lambda = \Lambda/3H_0^2$. Current observations indicate $\Omega_m \simeq 0.3, \Omega_\Lambda \simeq 0.7, \Omega_0 \simeq 1.02 \pm 0.02$. Because the nucleosynthesis results limit the baryon density to a very low value ($\Omega_{0b} \simeq 0.02$), which is about the same as the density of luminous matter, this indicates the dominant presence of both nonbaryonic dark matter and a repulsive force corresponding to either a cosmological constant or varying scalar field (dark energy).

Crucial causal limitations occur because of the existence of particle horizons (Rindler 1956), the

nature of which is most clear when represented in conformal diagrams (Hawking and Ellis 1973, Tipler *et al.* 1980). These result from the fact that light can only proceed a finite distance in the finite time since the origin of the universe, and imply that for a standard radiation-dominated hot-big-bang early universe, regions of larger than $\sim 1^\circ$ angular size on the surface of last scattering, which emits the CBR, are causally disconnected: hence, no causal process since the start of the universe can account for the extreme isotropy of the CBR ($\Delta T/T \simeq 10^{-5}$ over the whole sky, once a dipole anisotropy $\Delta T/T \simeq 10^{-3}$ due to our local velocity relative to the cosmological rest frame is allowed for). This is the “horizon problem,” one of the driving forces behind the theory of “inflation” (Guth 1981): the idea that, in the very early universe, a slow-rolling scalar field led to a brief exponential expansion through at least 50 e-folds (during which time the spacetime was approximately de Sitter), thus smoothing the universe and solving the horizon problem (Guth 1981, Peacock 1999). This is possible because a scalar field can violate the energy condition [3] and so allows acceleration: $\dot{S} > 0$. Consequently, there are now many studies of the dynamics of FLRW solutions driven by scalar fields and the subsequent decay of these scalar fields into radiation. One interesting point is that one can obtain exact solutions of this kind for arbitrarily chosen evolutions $S(t)$, provided they satisfy a restriction on the magnitude of \dot{S}^2 , by running the field equations backwards to determine the needed potential $V(\phi)$ (Ellis and Madsen 1991). The inflationary paradigm is dominant in present-day theoretical cosmology, but suffers from the problem that it is not in fact a well-defined theory, for there is no single accepted proposal for the physical nature of the effective scalar field underlying the supposed exponential expansion; rather there are numerous competing proposals. As the inflaton has not yet been identified, this theory is not yet soundly linked to well-established physics.

Approximate FL Solutions

The real universe is, of course, not exactly FL, and studies of structure formation depend on studies of solutions that are approximately FL models – they are realistic (“lumpy”) universe models. These enable detailed studies of observable properties such as CBR anisotropies and gravitational lensing induced by matter inhomogeneities, and of the development of those inhomogeneities from quantum fluctuations in the very early universe that then get expanded to very large scales by inflation.

The key problem here is that apart from the standard coordinate freedom allowed in general relativity, there is a serious gauge issue: the background FL model is not uniquely determined by the realistic universe model; however, the magnitudes of many perturbed quantities depend on how it is fitted into the lumpy model. For example, the density perturbation $\delta\rho$ is determined pointwise by the equation

$$\delta\rho(x^i) \equiv \rho(x^i) - \bar{\rho}(x^i)$$

where $\bar{\rho}(x^i)$ is the background density. But by altering the correspondence between the background and realistic models (specifically, by the choice of surfaces $\bar{\rho}(x^i) = \text{const.}$ in the realistic model) one can assign that quantity any value, including zero (if one chooses $\bar{\rho}(x^i) = \rho(x^i)$). This is the “gauge problem.”

One can handle it by using standard variables and keeping close track of the gauge freedom at all times. However, one then ends up with higher-order equations than necessary because some of the perturbation modes present are pure gauge modes with no physical significance. Alternatively, one can fix the gauge by some unique specification of how the background model is fitted into the realistic model, but there is no agreement on a unique way to do this, and different choices give different answers. The preferable resolution is to use gauge-invariant variables, either coordinate based (Bardeen 1980) or covariant, based on the (1+3) covariant decomposition of spacetime quantities mentioned above (Ellis and Bruni 1989), in either case resulting in perturbation equations without gauge freedom and of order corresponding to the physical degrees of freedom. The key point in the latter approach is to choose covariant variables that vanish in the background spacetime; they are then automatically gauge invariant. Realistic structure formation studies carry out this process for a mixture of matter components with different average velocities, and extend to a kinetic theory description of the background radiation (see Ellis and van Elst (1999) and references therein). The outcome is a prediction of the CBR anisotropy power spectrum, determined by the inhomogeneities in the gravitational field and the motions of the matter components at decoupling (Sachs and Wolfe 1967). This spectrum can then be compared with observations and used in determining the values of the cosmological parameters mentioned above (see Peacock 1999).

One crucial issue is why it is reasonable to use a perturbed FL model for the observable region of the universe. The key argument is that this is plausible because of the high isotropy of all observations around us when averaged on a sufficiently large spatial scale, and particularly the very low anisotropy

of the CBR. The Ehlers–Geren–Sachs (EGS) theorem (Ehlers *et al.* 1968) provides a sound basis for this argument: it shows that if freely propagating CBR (obeying the Liouville equation) is exactly isotropic in an expanding universe domain \mathcal{U} , then the universe is exactly FL in that domain (i.e., it has exactly the RW spatially homogenous and isotropic geometry there), the point being that any inhomogeneities in the matter distribution between us and the surface of last scattering will produce anisotropies in the CBR temperature we measure. But that result does not apply to the real universe, because the CBR is not exactly isotropic. The “almost EGS” theorem (Stoeger *et al.* 1995) shows that this result is stable: almost isotropic CBR in the domain \mathcal{U} implies that the universe is almost-FL in that domain. The application to the real universe comes by making a weak Copernican assumption: “we assume we are not special, so all observers in \mathcal{U} (taken to be the visible part of the universe) will also see almost isotropic CBR, just as we do.” The result then follows. A further argument for homogeneity of the universe comes from postulating “uniform thermal histories” (Bonnor and Ellis 1986), but that argument is yet to be completed and applied in a practical way.

Anisotropic and Inhomogeneous Models

The FL universes are geometrically extremely special. We wish further to understand the full range of possible universe models, their dynamical behaviors, and which of them might, at some epoch, realistically represent the real universe. This enables us to see how the approximate FL models fit into this wider set of possibilities, and under what circumstances they are attractors in this set of cosmologies.

Exact solutions are characterized by their space-time symmetries. Symmetries are characterized by the dimension s of the surfaces of homogeneity and the dimension q of the isotropy group at a general point, together giving the dimension $r = s + t$ of the group of isometries G_r (at special points, such as a center of symmetry, s can decrease and q increase but always so that r stays unchanged). In the case of a cosmological model, because the 4-velocity u^a is invariant under isotropies, the only possible dimensions for the isotropy group are $q = 3, 1, 0$; whereas the dimension t of the surfaces of homogeneity can take any value from 4 to 0. This gives the basis for a classification of cosmological spacetimes (Ellis 1967, Ellis and van Elst 1999).

When $q = 3$, we have isotropic solutions – there are no preferred spatial directions – and it is then a theorem that they must be spatially homogenous FL universes (Ehlers 1961). When $q = 1$, we

have locally rotationally symmetric (LRS) solutions, with precisely one preferred spacelike direction at a generic point (Ellis 1967). When $q=0$, the solutions are anisotropic in that there can be no continuous group of rotations leaving the solution invariant; however, there can be discrete isotropies in some special cases.

When $t=4$, we have spacetime homogeneous solutions, with all physical quantities constant; they cannot expand (by [5] and [3]). Nevertheless, two cases are of interest. For $q=1$ ($r=5$) we find the Gödel universe, rotating everywhere with constant vorticity, which illustrates important causal anomalies (Gödel 1949, Hawking and Ellis 1973). For $q=3$ ($r=6$), we find the Einstein “static universe” (Einstein 1917), the unique nonexpanding FL model with $k=1$ and $\Lambda > 0$. It is of interest because it could possibly represent the asymptotic initial state of nonsingular inflationary universe models (Ellis *et al.* 2003). The higher-symmetry models (de Sitter and anti-de Sitter universes with higher-dimensional isotropy groups) are not included here because they do not obey the energy condition [3] – they are empty universes, which can be interesting asymptotic states but are not by themselves good cosmological models.

When $t=3$, we have spatially homogeneous evolving universe models. For $q=0$ ($r=3$), there are a large family of Bianchi universes, spatially homogeneous but anisotropic, characterized into nine types according to the structure constants of the Lie algebra of the three-dimensional symmetry group G_3 . These can be “orthogonal”: the fluid flow is orthogonal to the surfaces of homogeneity, or “tilted”; the latter case can have fluid rotation or acceleration, but the former cannot. They exhibit a large variety of behaviors, including power-law, oscillatory, and nonscalar singularities (Tipler *et al.* 1980). A vexed question is whether truly chaotic behavior occurs in Bianchi IX models. The behavior of large families of these models has been characterized in dynamical systems terms (Wainwright and Ellis 1996), showing the intriguing way that higher-symmetry solutions provide a “skeleton” that guides the behavior of lower-symmetry solutions in the space of spacetimes. Many Bianchi models can be shown to isotropize at late times, particularly if viscosity is present; thus, they are asymptotic to the FL universes in the far future. In some cases, Bianchi models exhibit intermediate isotropization: they are much like FL models for a large part of their life, but are very different from it both at very early and very late stages of their evolution. These could be good models of the real universe. An important theorem by Wald (1983) shows that a cosmological constant will tend to isotropize Bianchi solutions at late

times. This is an indication that inflation can succeed in making anisotropic early states resemble FL models at later times. Observational properties like element abundances and CBR anisotropy patterns can be worked out in these models (some of them develop a characteristic isolated “hot spot” in the CBR sky). For $q=1$ ($r=4$), we have spatially homogeneous LRS models, either Kantowski Sachs or Bianchi universes, and again observations can be worked out in detail and phase planes developed showing their dynamical behavior, often isotropizing at late times. There are orthogonal and tilted cases, the latter possibly involving nonscalar singularities. For $q=3$ ($r=6$), we have the isotropic FL models, discussed above. Both the LRS and isotropic cases could be good models of the real universe.

When $t=2$, we have inhomogeneous evolving models. This is a very large family, but the LRS ($q=1, r=2$) cases have been examined in detail; in the case of pressure-free matter, these are the Tolman–Bondi inhomogeneous models (Bondi 1947) that can be integrated exactly, and have been used for many interesting astrophysical and cosmological studies. Krasiński (1997) gives a very complete catalog of these and lower-symmetry inhomogeneous models and their uses in cosmology. A considerable challenge is the dynamical systems analysis for generic inhomogeneous models, needed to properly understand the early evolution of generic universe models (Uggla *et al.* 2003), and hence to determine what is generic behavior.

The Origin of the Universe

The issue underlying all this is what led to the initial conditions for the universe, for example, providing the starting conditions for inflation. There are many approaches to studying the quantum gravity phase of cosmology, including the Wheeler–de Witt equation, the path-integral approach, string cosmology, pre-big bang theory, brane cosmology, the ekpyrotic universe, the cyclic universe, and loop quantum gravity approaches. These lie beyond the purview of the present article, except to say that they are all based on unproven extrapolations of known physics. The physically possible paths will become clearer as the nature of quantum gravity is elucidated.

It is pertinent to note that there exist nonsingular realistic cosmological solutions, possible in the light of the violations of the energy condition enabled by the supposed scalar fields that underlie inflationary universe theory. These nonsingular solutions can even avoid the quantum gravity era (Ellis *et al.* 2003). However, they have very fine-tuned initial conditions, which is nowadays considered as a disadvantage; but

there is no proof that whatever processes led to the existence of the universe preferred generic rather than fine-tuned conditions; this is a philosophical rather than physical assumption. It may well be that, as regards the start of the universe, the options are that either an initial singularity occurred, or the initial conditions were very finely tuned and allowed an infinitely existing universe. Investigation of whether this conjecture is in fact valid, and if so which is the best option, are intriguing open topics.

See also: Einstein Equations: Exact Solutions; Einstein–Cartan Theory; General Relativity: Experimental Tests; General Relativity: Overview; Gravitational Lensing; Lie Groups: General Theory; Newtonian Limit of General Relativity; Quantum Cosmology; Shock Wave Refinement of the Friedman–Robertson–Walker Metric; Spacetime Topology, Causal Structure and Singularities; String Theory: Phenomenology.

Further Reading

- Bardeen JM (1980) *Physical Review D* 22: 1882.
 Bondi H (1947) *Monthly Notices of the Royal Astronomical Society* 107: 410.
 Bonnor WB and Ellis GFR (1986) *Monthly Notices of the Royal Astronomical Society* 218: 605.
 Ehlers J (1961) *Abh Mainz Akad Wiss u Lit* (translated in *Gen Rel Grav* 25: 1225, 1993).
 Ehlers J, Geren P, and Sachs RK (1968) *Journal of Mathematical Physics* 9: 1344.
 Ehlers J and Rindler W (1989) *Monthly Notices of the Royal Astronomical Society* 238: 503.
 Einstein A (1917) *Sitz Ber Preuss Akad Wiss* (translated in *The Principle of Relativity*, 1993). Dover.
 Ellis GFR (1967) *Journal of Mathematical Physics* 8: 1171.

- Ellis GFR (1971) In: Sachs RK (ed.) *General Relativity and Cosmology*, Proc. Int. School of Physics “Enrico Fermi,” Course XLVII, p. 104. Academic Press.
 Ellis GFR and Bruni M (1989) *Physical Review D* 40: 1804.
 Ellis GFR and van Elst H (1999) In: Lachieze-Ray M (ed.) *Theoretical and Observational Cosmology*, vol. 541 [gr-qc/9812046], Nato Series C: Mathematical and Physical Sciences: Kluwer.
 Ellis GFR and Madsen M (1991) *Classical and Quantum Gravity* 8: 667.
 Ellis GFR, Murugan J, and Tsagas CG (2003) gr-qc/0307112.
 Ellis GFR, Nel SD, Stoeger W, Maartens R, and Whitman AP (1985) *Physics Reports* 124(5 and 6): 315.
 Gödel K (1949) *Reviews of Modern Physics* 21: 447.
 Guth A (1981) *Physical Review D* 23: 347.
 Hawking SW and Ellis GFR (1973) *The Large Scale Structure of Space Time*. Cambridge: Cambridge University Press.
 Krasiński A (1997) *Inhomogeneous Cosmological Models*. Cambridge: Cambridge University Press.
 Kristian J and Sachs RK (1966) *The Astrophysical Journal* 143: 379.
 Lachieze-Rey M and Luminet JP (1995) *Physics Reports* 254: 135–214.
 Robertson HP (1933) *Reviews of Modern Physics* 5: 62.
 Peacock JA (1999) *Cosmological Physics*. Cambridge: Cambridge University Press.
 Rindler W (1956) *Monthly Notices of the Royal Astronomical Society* 116: 662.
 Sachs RK and Wolfe A (1967) *Astrophysical Journal* 147: 73.
 Sandage A (1961) *Astrophysical Journal* 133: 355.
 Stoeger W, Maartens R, and Ellis GFR (1995) *Astrophysical Journal* 443: 1.
 Tipler FJ, Clarke CJS, and Ellis GFR (1980) In: Held A (ed.) *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein*, vol. 2, p. 97. Plenum.
 Uggla C, van Elst H, Wainwright J, and Ellis GFR (2003) *Physical Review D* gr-qc/0304002 (to appear).
 Wainwright J and Ellis GFR (eds.) (1996) *The Dynamical Systems Approach to Cosmology*. Cambridge: Cambridge University Press.
 Wald RM (1983) *Physical Review D* 28: 2118.

Cotangent Bundle Reduction

J-P Ortega, Université de Franche-Comté, Besançon, France

T S Ratiu, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The general symplectic reduction theory (*see* Symmetry and Symplectic Reduction) becomes much richer and has many applications if the symplectic manifold is the cotangent bundle ($T^*Q, \Omega_Q = -d\Theta_Q$) of a manifold Q . The canonical 1-form Θ_Q on T^*Q is given by $\Theta_Q(\alpha_q)(V_{\alpha_q}) = \alpha_q(T_{\alpha_q}\pi_Q(V_{\alpha_q}))$, for any $q \in Q$, $\alpha_q \in T_q^*Q$, and

tangent vector $V_{\alpha_q} \in T_{\alpha_q}(T^*Q)$, where $\pi_Q : T^*Q \rightarrow Q$ is the cotangent bundle projection and $T_{\alpha_q}\pi_Q : T_{\alpha_q}(T^*Q) \rightarrow T_qQ$ is its tangent map (or derivative) at q . In natural cotangent bundle coordinates (q^i, p_i) , we have $\Theta_Q = p_i dq^i$ and $\Omega_Q = dq^i \wedge dp_i$.

Let $\Phi : G \times Q \rightarrow Q$ be a left smooth action of the Lie group G on the manifold and Q . Denote by $g \cdot q = \Phi(g, q)$ the action of $g \in G$ on the point $q \in Q$ and by $\Phi_g : Q \rightarrow Q$ the diffeomorphism of Q induced by g . The lifted left action $G \times T^*Q \rightarrow T^*Q$, given by $g \cdot \alpha_q = T_{g \cdot q}^* \Phi_{g^{-1}}(\alpha_q)$ for $g \in G$ and $\alpha_q \in T_q^*Q$, preserves Θ_Q , and admits the equivariant momentum map $J : T^*Q \rightarrow \mathfrak{g}^*$ whose expression is $\langle J(\alpha_q), \xi \rangle = \alpha_q(\langle \xi_Q(q) \rangle)$, where $\xi \in \mathfrak{g}$, the Lie algebra of G , $\langle \cdot, \cdot \rangle : \mathfrak{g}^* \times \mathfrak{g} \rightarrow \mathbb{R}$ is the duality pairing between the dual \mathfrak{g}^* and \mathfrak{g} , and $\xi_Q(q) = d\Phi(\exp t\xi, q)/dt|_{t=0}$ is the value of the

infinitesimal generator vector field ξ_Q of the G -action at $q \in Q$ (see Hamiltonian Group Actions and Symmetries and Conservation Laws). Throughout this article, it is assumed that the G -action on Q , and hence on T^*Q , is free and proper. Recall also that $((T^*Q)_\mu, (\Omega_Q)_\mu)$ denotes the reduced manifold at $\mu \in \mathfrak{g}^*$ (see Symmetry and Symplectic Reduction), where $(T^*Q)_\mu := J^{-1}(\mu)/G_\mu$ is the orbit space of the G_μ -action on the momentum level manifold $J^{-1}(\mu)$ and $G_\mu := \{g \in G \mid \text{Ad}_g^* \mu = \mu\}$ is the isotropy subgroup of the coadjoint representation of G on \mathfrak{g}^* . The left-coadjoint representation of $g \in G$ on $\mu \in \mathfrak{g}^*$ is denoted by $\text{Ad}_{g^{-1}}^* \mu$.

Cotangent bundle reduction at zero is already quite interesting and has many applications. Let $\rho: Q \rightarrow Q/G$ be the G -principal bundle projection defined by the proper free action of G on Q , usually referred to as the shape space bundle. Zero is a regular value of J and the map $\varphi_0: ((T^*Q)_0, (\Omega_Q)_0) \rightarrow (T^*(Q/G), \Omega_{Q/G})$ given by $\varphi_0([\alpha_q])(T_q \rho(v_q)) := \alpha_q(v_q)$, where $\alpha_q \in J^{-1}(0)$, $[\alpha_q] \in (T^*Q)_0$, and $v_q \in T_q Q$, is a well-defined symplectic diffeomorphism.

This theorem generalizes in two nontrivial ways when one reduces at a nonzero value of J : an embedding and a fibration theorem.

Embedding Version of Cotangent Bundle Reduction

Let $\mu \in \mathfrak{g}^*$, $Q_\mu := Q/G_\mu$, $\rho_\mu: Q \rightarrow Q_\mu$ the projection onto the G_μ -orbit space, $\mathfrak{g}_\mu := \{\xi \in \mathfrak{g} \mid \text{ad}_\xi^* \mu = 0\}$ the Lie algebra of the coadjoint isotropy subgroup G_μ , where $\text{ad}_\xi \eta := [\xi, \eta]$ for any $\xi, \eta \in \mathfrak{g}$, $\text{ad}_\xi^*: \mathfrak{g}^* \rightarrow \mathfrak{g}^*$ the dual map, $\mu' := \mu|_{\mathfrak{g}_\mu} \in \mathfrak{g}_\mu^*$ the restriction of μ to \mathfrak{g}_μ , and $((T^*Q)_\mu, (\Omega_Q)_\mu)$ the reduced space at μ . The induced G_μ -action on T^*Q admits the equivariant momentum map $J^\mu: T^*Q \rightarrow \mathfrak{g}_\mu^*$ given by $J^\mu(\alpha_q) = J(\alpha_q)|_{\mathfrak{g}_\mu}$. Assume there is a G_μ -invariant 1-form α_μ on Q with values in $(J^\mu)^{-1}(\mu')$. Then there is a unique closed 2-form β_μ on Q_μ such that $\rho_\mu^* \beta_\mu = d\alpha_\mu$. Define the magnetic term $B_\mu := \pi_{Q_\mu}^* \beta_\mu$, where $\pi_{Q_\mu}: T^*Q_\mu \rightarrow Q_\mu$ is the cotangent bundle projection, which is a closed 2-form on T^*Q_μ . Then the map $\varphi_\mu: ((T^*Q)_\mu, (\Omega_Q)_\mu) \rightarrow (T^*Q_\mu, \Omega_{Q_\mu} - B_\mu)$ given by $\varphi_\mu([\alpha_q])(T_q \rho_\mu(v_q)) := (\alpha_q - \alpha_\mu(q))(v_q)$, for $\alpha_q \in J^{-1}(\mu)$, $[\alpha_q] \in (T^*Q)_\mu$, and $v_q \in T_q Q$, is a symplectic embedding onto a submanifold of T^*Q_μ covering the base Q_μ . The embedding φ_μ is a diffeomorphism onto T^*Q_μ if and only if $\mathfrak{g} = \mathfrak{g}_\mu$. If the 1-form α_μ takes values in the smaller set $J^{-1}(\mu)$ then the image of φ_μ is the vector sub-bundle $[T \rho_\mu(VQ)]^\circ$ of T^*Q_μ , where $VQ \subset TQ$ is the vertical vector sub-bundle consisting of vectors tangent to the G -orbits, that is, its fiber at $q \in Q$ equals $V_q Q = \{\xi_Q(q) \mid \xi \in \mathfrak{g}\}$, and $^\circ$ denotes the annihilator relative to the natural duality pairing

between TQ_μ and T^*Q_μ . Note that if \mathfrak{g} is abelian or $\mu=0$, the embedding φ_μ is always onto and thus the reduced space is again, topologically, a cotangent bundle.

It should be noted that there is a choice in this theorem, namely the 1-form α_μ . Whereas the reduced symplectic space $((T^*Q)_\mu, (\Omega_Q)_\mu)$ is intrinsic, the symplectic structure on the space T^*Q_μ depends on α_μ . The theorem above states that no matter how α_μ is chosen, there is a symplectic diffeomorphism, which also depends on α_μ , of the reduced space onto a submanifold of T^*Q_μ .

Connections

The 1-form α_μ is usually obtained from a left connection on the principal bundle $\rho_\mu: Q \rightarrow Q/G_\mu$ or $\rho: Q \rightarrow Q/G$. A left connection 1-form $A \in \Omega^1(Q; \mathfrak{g})$ on the left principal G -bundle $\rho: Q \rightarrow Q/G$ is a Lie algebra-valued 1-form $A: TQ \rightarrow \mathfrak{g}$, where \mathfrak{g} denotes the Lie algebra of G , satisfying the conditions $A(\xi_Q) = \xi$ for all $\xi \in \mathfrak{g}$ and $\mathcal{A}(T_q \Phi_g(v)) = \text{Ad}_g(\mathcal{A}(v))$ for all $g \in G$ and $v \in T_q Q$, where Ad_g denotes the adjoint action of G on \mathfrak{g} . The horizontal vector sub-bundle HQ of the connection A is defined as the kernel of A , that is, its fiber at $q \in Q$ is the subspace $H_q := \ker A(q)$. The map $v_q \mapsto \text{ver}_q(v_q) := [A(q)(v_q)]_Q(q)$ is called the vertical projection, while the map $v_q \mapsto \text{hor}_q(v_q) := v_q - \text{ver}_q(v_q)$ is called the horizontal projection. Since for any vector $v_q \in T_q Q$ we have $v_q = \text{ver}_q(v_q) + \text{hor}_q(v_q)$, it follows that $TQ = HQ \oplus VQ$ and the maps $\text{hor}_q: T_q Q \rightarrow H_q Q$ and $\text{ver}_q: T_q Q \rightarrow V_q Q$ are projections onto the horizontal and vertical subspaces at every $q \in Q$.

Connections can be equivalently defined by the choice of a sub-bundle $HQ \subset TQ$ complementary to the vertical sub-bundle VQ satisfying the following G -invariance property: $H_{g \cdot q} Q = T_q \Phi_g(H_q Q)$ for every $g \in G$ and $q \in Q$. The sub-bundle HQ is called, as before, the horizontal sub-bundle and a connection 1-form A is defined by setting $A(q)(\xi_Q(q) + u_q) = \xi$, for any $\xi \in \mathfrak{g}$ and $u_q \in H_q Q$.

The curvature of the connection A is the Lie algebra-valued 2-form on Q defined by $B(u_q, v_q) = dA(\text{hor}_q(u_q), \text{hor}_q(v_q))$. When one replaces vectors in the exterior derivative with their horizontal projections, then the result is called the exterior covariant derivative and the preceding formula for B is often written as $B = DA$. Curvature measures the lack of integrability of the horizontal distribution, namely $B(u, v) = -A([\text{hor}(u), \text{hor}(v)])$ for any two vector fields u and v on Q . The Cartan structure equations state that $B(u, v) = d\mathcal{A}(u, v) - [\mathcal{A}(u), \mathcal{A}(v)]$, where the bracket on the right hand side is the Lie bracket in \mathfrak{g} .

Since the connection A is a Lie algebra-valued 1-form, for each $\mu \in \mathfrak{g}^*$ the formula $\alpha_\mu(q) := A(q)^*(\mu)$, where $A(q)^*: \mathfrak{g}^* \rightarrow T_q^*Q$ is the dual of the linear map $A(q): T_qQ \rightarrow \mathfrak{g}$, defines a usual 1-form on Q . This 1-form α_μ takes values in $J^{-1}(\mu)$ and is equivariant in the following sense: $\Phi_g^* \alpha_\mu = \alpha_{\text{Ad}_g^* \mu}$ for any $g \in G$.

Magnetic Terms and Curvature

There are two methods to construct the 1-form α_μ from a connection. The first is to start with a connection 1-form $A^\mu \in \Omega^1(Q; \mathfrak{g}_\mu)$ on the principal G_μ -bundle $\rho_\mu: Q \rightarrow Q/G_\mu$. Then the 1-form $\alpha_\mu := \langle \mu|_{\mathfrak{g}_\mu}, A^\mu \rangle \in \Omega^1(Q)$ is G_μ -invariant and has values in $(J^\mu)^{-1}(\mu|_{\mathfrak{g}_\mu})$. The magnetic term B_μ is the pullback to $T^*(Q/G_\mu)$ of the $\mu|_{\mathfrak{g}_\mu}$ -component $d\alpha_\mu$ of the curvature of A^μ thought of as a 2-form on the base Q/G_μ .

The second method is to start with a connection $A \in \Omega^1(Q, \mathfrak{g})$ on the principal bundle $\rho: Q \rightarrow Q/G$, to define $\alpha_\mu := \langle \mu, A \rangle \in \Omega^1(Q)$, and to observe that this 1-form is G_μ -invariant and has values in $J^{-1}(\mu)$. The magnetic term B_μ is in this case the pullback to $T^*(Q/G_\mu)$ of the μ -component $d\alpha_\mu$ of the curvature of A thought of as a 2-form on the base Q/G_μ .

The Mechanical Connection

If $(Q, \langle \langle \cdot, \cdot \rangle \rangle)$ is a Riemannian manifold and G acts by isometries, there is a natural connection on the bundle $\rho: Q \rightarrow Q/G$, namely, define the horizontal space at a point to be the metric orthogonal to the vertical space. This connection is called the mechanical connection and its horizontal bundle consists of all vectors $v_q \in TQ$ such that $J(\langle \langle v_q, \cdot \rangle \rangle) = 0$.

To determine the Lie algebra-valued 1-form A of this connection, the notion of locked inertia tensor needs to be introduced. This is the linear map $\mathbb{I}(q): \mathfrak{g} \rightarrow \mathfrak{g}^*$ depending smoothly on $q \in Q$ defined by the identity $\langle \mathbb{I}(q)\xi, \eta \rangle = \langle \langle \xi_Q(q), \eta_Q(q) \rangle \rangle$ for any $\xi, \eta \in \mathfrak{g}$. Since the G -action is free, each $\mathbb{I}(q)$ is invertible. The connection 1-form whose horizontal space was defined above is given by $A(q)(v_q) = \mathbb{I}(q)^{-1}(J(\langle \langle v_q, \cdot \rangle \rangle))$.

Denote by $K: T^*Q \rightarrow \mathbb{R}$ the kinetic energy of the metric $\langle \langle \cdot, \cdot \rangle \rangle$ on the cotangent bundle, that is, $K(\langle \langle v_q, \cdot \rangle \rangle) := (1/2)\|v_q\|^2$. The 1-form $\alpha_\mu = A(\cdot)^* \mu$ is characterized for the mechanical connection A by the condition $K(\alpha_\mu(q)) = \inf \{K(\beta_q) \mid \beta_q \in J^{-1}(\mu) \cap T_q^*Q\}$.

The Amended Potential

A simple mechanical system is a Hamiltonian system on a cotangent bundle T^*Q whose Hamiltonian function is the sum of the kinetic energy of a Riemannian metric on Q and a potential function

$V: Q \rightarrow \mathbb{R}$. If there is a Lie group G acting on Q by isometries and leaving the potential invariant, then we have a simple mechanical system with symmetry. The amended or effective potential $V_\mu: Q \rightarrow \mathbb{R}$ at $\mu \in \mathfrak{g}^*$ is defined by $V_\mu := H \circ \alpha_\mu$, where α_μ is the 1-form associated to the mechanical connection. Its expression in terms of the locked moment of inertia tensor is given by $V_\mu(q) := V(q) + (1/2)\langle \mu, \mathbb{I}(q)^{-1}\mu \rangle$. The amended potential naturally induces a smooth function $\widehat{V}_\mu \in C^\infty(Q/G_\mu)$.

The fundamental result about simple mechanical systems with symmetry is the following. The push-forward by the embedding $\varphi_\mu: ((T^*Q)_\mu, (\Omega_Q)_\mu) \rightarrow (T^*Q_\mu, \Omega_{Q_\mu} - B_\mu)$ of the reduced Hamiltonian $H_\mu \in C^\infty((T^*Q)_\mu)$ of a simple mechanical system $H = K + V \circ \pi_Q \in C^\infty(T^*Q)$ is the restriction to the vector sub-bundle $\varphi_\mu((T^*Q)_\mu) \subset T^*(Q/G_\mu)$, which is also a symplectic submanifold of $(T^*(Q/G_\mu), \Omega_{Q/G_\mu} - B_\mu)$, of the simple mechanical system on $T^*(Q/G_\mu)$ whose kinetic energy is given by the quotient Riemannian metric on Q/G_μ and whose potential is \widehat{V}_μ . However, Hamilton's equations on $T^*(Q/G_\mu)$ for this simple mechanical system are computed relative to the magnetic symplectic form $\Omega_{Q/G_\mu} - B_\mu$.

There is a wealth of applications starting from this classical theorem to mechanical systems, spanning such diverse areas as topological characterization of the level sets of the energy–momentum map to methods of proving nonlinear stability of relative equilibria (block-diagonalization of the stability form in the application of the energy–momentum method).

Fibration Version of Cotangent Bundle Reduction

There is a second theorem that realizes the reduced space of a cotangent bundle as a locally trivial bundle over shape space Q/G . This version is particularly well suited in the study of quantization problems and in control theory. The result is the following. Assume that G acts freely and properly on Q . Then the reduced symplectic manifold $(T^*Q)_\mu$ is a fiber bundle over $T^*(Q/G)$ with fiber the coadjoint orbit \mathcal{O}_μ . How this is related to the Poisson structure of the quotient $(T^*Q)/G$ will be discussed later.

The Kaluza–Klein Construction

The extra term in the symplectic form of the reduced space is called a magnetic term because it has this interpretation in electromagnetism. To understand why B_μ is called a magnetic term, consider the problem of a particle of mass m and charge e moving in \mathbb{R}^3 under the influence of a given

magnetic field $B = B_x i + B_y j + B_z k, \operatorname{div} B = 0$. The Lorentz force law (written in the International System) gives the equations of motion

$$m \frac{d\mathbf{v}}{dt} = e\mathbf{v} \times \mathbf{B} \tag{1}$$

where e is the charge and $\mathbf{v} = (\dot{x}, \dot{y}, \dot{z}) = \dot{\mathbf{q}}$ is the velocity of the particle. What is the Hamiltonian description of these equations?

There are two possible answers to this question. To formulate them, associate to the divergence free vector field \mathbf{B} the closed 2-form $B = B_x dy \wedge dz - B_y dx \wedge dz + B_z dx \wedge dy$. Also, write $\mathbf{B} = \operatorname{curl} \mathbf{A}$ for some other vector field $\mathbf{A} = (A_x, A_y, A_z)$ on \mathbb{R}^3 , called the magnetic potential.

Answer 1 Take on $T^*\mathbb{R}^3$ the symplectic form $\Omega_B = dx \wedge dp_x + dy \wedge dp_y + dz \wedge dp_z - eB$, where $(p_x, p_y, p_z) = \mathbf{p} := m\mathbf{v}$ is the momentum of the particle, and $h = m\|\mathbf{v}\|^2/2 = m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)/2$ is the Hamiltonian, the kinetic energy of the particle. A direct verification shows that $dh = \Omega_B(X_h, \cdot)$, where

$$\begin{aligned} X_h = & \dot{x} \frac{\partial}{\partial x} + \dot{y} \frac{\partial}{\partial y} + \dot{z} \frac{\partial}{\partial z} + e(B_z \dot{y} - B_y \dot{z}) \frac{\partial}{\partial p_x} \\ & + e(B_x \dot{z} - B_z \dot{x}) \frac{\partial}{\partial p_y} + e(B_y \dot{x} - B_x \dot{y}) \frac{\partial}{\partial p_z} \end{aligned} \tag{2}$$

which gives the equations of motion [1].

Answer 2 Take on $T^*\mathbb{R}^3$ the canonical symplectic form $\Omega = dx \wedge dp_x + dy \wedge dp_y + dz \wedge dp_z$ and the Hamiltonian $h_A = \|\mathbf{p} - e\mathbf{A}\|^2/2m$. A direct verification shows that $dh_A = \Omega(X_{h_A}, \cdot)$, where X_{h_A} has the same expression [2].

Next we show how the magnetic term in the symplectic form Ω_B is obtained by reduction from the Kaluza–Klein system. Let $Q = \mathbb{R}^3 \times S^1$ with the circle $G = S^1$ acting on Q , only on the second factor. Identify the Lie algebra \mathfrak{g} of S^1 with \mathbb{R} . Since the infinitesimal generator of this action defined by $\xi \in \mathfrak{g} = \mathbb{R}$ has the expression $\xi_Q(\mathbf{q}, \theta) = (\mathbf{q}, \theta; \mathbf{0}, \xi)$, if TS^1 is trivialized as $S^1 \times \mathbb{R}$, a momentum map $J : T^*Q = \mathbb{R}^3 \times S^1 \times \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathfrak{g}^* = \mathbb{R}$ is given by $J(\mathbf{q}, \theta; \mathbf{p}, p)\xi = (\mathbf{p}, p) \cdot (\mathbf{0}, \xi) = p\xi$, that is, $J(\mathbf{q}, \theta; \mathbf{p}, p) = p$. In this case, the coadjoint action is trivial, so for any $\mu \in \mathfrak{g}^* = \mathbb{R}$, we have $G_\mu = S^1, \mathfrak{g}_\mu = \mathbb{R}$, and $\mu' = \mu$. The 1-form $\alpha_\mu = \mu(A_x dx + A_y dy + A_z dz + d\theta) \in \Omega^1(Q)$, where $d\theta$ denotes the length 1-form on S^1 , is clearly $G_\mu = S^1$ -invariant, has values in $J^{-1}(\mu) = \{(\mathbf{q}, \theta; \mathbf{p}, \mu) \mid \mathbf{q}, \mathbf{p} \in \mathbb{R}^3, \theta \in S^1\}$, and its exterior differential equals $d\alpha_\mu = \mu B$. Thus, the closed 2-form β_μ on the base $Q_\mu = Q/G_\mu = Q/S^1 = \mathbb{R}^3$ equals μB and hence the magnetic term, that is, the closed 2-form $B_\mu = \pi_{Q_\mu}^* \beta_\mu$ on $T^*Q_\mu = T^*\mathbb{R}^3$, is also μB since $\pi_{Q_\mu} : Q = \mathbb{R}^3 \times S^1 \rightarrow Q/G_\mu = \mathbb{R}^3$ is the projection. Therefore, the reduced space $(T^*Q)_\mu$ is

symplectically diffeomorphic to $(T^*\mathbb{R}^3, dx \wedge dp_x + dy \wedge dp_y + dz \wedge dp_z - \mu B)$, which coincides with the phase space in Answer 1 if we put $\mu = e$. This also gives the physical interpretation of the momentum map $J : T^*Q = \mathbb{R}^3 \times S^1 \times \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathfrak{g}^* = \mathbb{R}$, $J(\mathbf{q}, \theta; \mathbf{p}, p) = p$ and hence of the variable conjugate to the circle variable θ : p represents the charge. Moreover, the magnetic term in the symplectic form is, up to a charge factor, the magnetic field.

The kinetic energy Hamiltonian

$$h(\mathbf{q}, \theta; \mathbf{p}, p) := \frac{1}{2m} \|\mathbf{p}\|^2 + \frac{1}{2} p^2$$

of the Kaluza–Klein metric, that is, the Riemannian metric obtained by keeping the standard metrics on each factor and declaring \mathbb{R}^3 and S^1 orthogonal, induces the reduced Hamiltonian

$$h_\mu(\mathbf{q}) = \frac{1}{2m} \|\mathbf{p}\|^2 + \frac{1}{2} \mu^2$$

which, up to the constant $\mu^2/2$, equals the kinetic energy Hamiltonian in Answer 1. Note that this reduced system is not the geodesic flow of the Euclidean metric because of the presence of the magnetic term in the symplectic form. However, the equations of motion of a charged particle in a magnetic field are obtained by reducing the geodesic flow of the Kaluza–Klein metric.

A similar construction is carried out in Yang–Mills theory where A is a connection on a principal bundle and B is its curvature. Magnetic terms also appear in classical mechanics. For example, in rotating systems the Coriolis force (up to a dimensional factor) plays the role of the magnetic term.

Reconstruction of Dynamics for Cotangent Bundles

A general reconstruction method of the dynamics from the reduced dynamics was given in (see Symmetry and Symplectic Reduction). For cotangent bundles, using the mechanical connection, this method simplifies considerably.

Start with the following general situation. Let G act freely on the configuration manifold Q ; let $h : T^*Q \rightarrow \mathbb{R}$ be a G -invariant Hamiltonian, $\mu \in \mathfrak{g}^*, \alpha_\mu \in J^{-1}(\mu)$, and $c_\mu(t)$ the integral curve of the reduced system with initial condition $[\alpha_\mu] \in (T^*Q)_\mu$ given by the reduced Hamiltonian function $h_\mu : (T^*Q)_\mu \rightarrow \mathbb{R}$. In terms of a connection $A \in \Omega^1(J^{-1}(\mu); \mathfrak{g}_\mu)$ on the left G_μ -principal bundle $J^{-1}(\mu) \rightarrow (T^*Q)_\mu$ the reconstruction procedure proceeds in four steps:

- *Step 1:* Horizontally lift the curve $c_\mu(t) \in (T^*Q)_\mu$ to a curve $d(t) \in J^{-1}(\mu)$ with $d(0) = \alpha_\mu$.
- *Step 2:* Set $\xi(t) = A(d(t))(X_h(d(t))) \in \mathfrak{g}_\mu$.

- *Step 3:* With $\xi(t) \in \mathfrak{g}_\mu$ determined in step 2, solve the nonautonomous differential equation $\dot{g}(t) = T_e L_{g(t)} \xi(t)$ with initial condition $g(0) = e$, where L_g denotes left translation on G ; this is the step that involves “quadratures” and is the main obstacle to finding explicit formulas.
- *Step 4:* The curve $c(t) = g(t) \cdot d(t)$, with $d(t)$ found in step 1 and $g(t)$ found in step 3 is the integral curve of X_h with initial condition $c(0) = \alpha_q$.

This method depends on the choice of the connection $A \in \Omega^1(J^{-1}(\mu); \mathfrak{g}_\mu)$. Here are several particular cases when this procedure simplifies.

(a) *One-dimensional coadjoint isotropy group.* If $G_\mu = S^1$ or $G_\mu = \mathbb{R}$, identify \mathfrak{g}_μ with \mathbb{R} via the map $a \in \mathbb{R} \leftrightarrow a\zeta \in \mathfrak{g}_\mu$, where $\zeta \in \mathfrak{g}_\mu, \zeta \neq 0$, is a generator of \mathfrak{g}_μ . Then a connection 1-form on the S^1 (or \mathbb{R}) principal bundle $J^{-1}(\mu) \rightarrow (T^*Q)_\mu$ is the 1-form A on $J^{-1}(\mu)$ given by $A = (1/\langle \mu, \zeta \rangle) \theta_\mu$, where θ_μ is the pullback of the canonical 1-form $\theta \in \Omega^1(T^*Q)$ to the submanifold $J^{-1}(\mu)$. The curvature of this connection is the 2-form on $(T^*Q)_\mu$ given by $\text{curv}(A) = -(1/\langle \mu, \zeta \rangle) \omega_\mu$, where ω_μ is the reduced symplectic form on $(T^*Q)_\mu$. In this case, the curve $\xi(t) \in \mathfrak{g}_\mu$ in step 2 is given by $\xi(t) = \Lambda[h](d(t))$, where $\Lambda \in \mathfrak{X}(T^*Q)$ is the Liouville vector field characterized by the property of being the unique vector field on T^*Q that satisfies the relation $d\theta(\Lambda, \cdot) = \theta$. In canonical coordinates (q^i, p_i) on T^*Q , $\Lambda = p_i \frac{\partial}{\partial p_i}$.

(b) *Induced connection.* Any connection $\mathcal{A} \in \Omega^1(Q; \mathfrak{g}_\mu)$ on the left principal bundle $Q \rightarrow Q/G_\mu$ induces a connection $A \in \Omega^1(J^{-1}(\mu); \mathfrak{g}_\mu)$ by $A(\alpha_q) \times (V_{\alpha_q}) := \mathcal{A}(q)(T_{\alpha_q} \pi_Q(V_{\alpha_q}))$, where $q \in Q, \alpha_q \in T_q^*Q, V_{\alpha_q} \in T_{\alpha_q}(T^*Q)$, and $\pi_Q: T^*Q \rightarrow Q$ is the cotangent bundle projection. In this case, the curve $\xi(t) \in \mathfrak{g}_\mu$ in step 2 is given by $\xi(t) = \mathcal{A}(q(t))(\mathbb{F}h(d(t)))$, where $q(t) := \pi_Q(d(t))$ is the base integral curve and the vector bundle morphism $\mathbb{F}h: T^*Q \rightarrow TQ$ is the fiber derivative of h given by

$$\mathbb{F}h(\alpha_q)(\beta_q) := \left. \frac{d}{dt} \right|_{t=0} h(\alpha_q + t\beta_q)$$

for any $\alpha_q, \beta_q \in T_a^*Q$. Two particular instances of this situation are noteworthy.

- (b1) Assume that the Hamiltonian h is that of a simple mechanical system with symmetry. Choosing A to be the mechanical connection A_{mech} , the curve $\xi(t) \in \mathfrak{g}_\mu$ in step 2 is given by $\xi(t) = A_{\text{mech}}(q(t))(\langle d(t), \cdot \rangle)$.
- (b2) If $Q = G$ is a Lie group, $\dim G_\mu = 1$, and ζ is a generator of \mathfrak{g}_μ , then the connection $A \in \Omega^1(G)$ can be chosen to equal $\mathcal{A}(g) := (1/\langle \mu, \zeta \rangle) T_g^* R_{g^{-1}}(\mu)$, where ζ is a generator of \mathfrak{g}_μ and R_g is right translation on G .

(c) *Reconstruction of dynamics for simple mechanical systems with symmetry.* The case of simple mechanical systems with symmetry deserves special attention since several steps in the reconstruction method can be simplified. For simple mechanical systems, the knowledge of the base integral curve $q(t)$ suffices to determine the entire integral curve on T^*Q . Indeed, if $h = K + V \circ \pi_q$ is the Hamiltonian, the Legendre transformation $\mathbb{F}h: T^*Q \rightarrow TQ$ determines the Lagrangian system on TQ given by $\ell(u_q) = (1/2)\|u_q\|^2 - V(u_q)$, for $u_q \in T_qQ$. Lagrange’s equations are second-order and thus the evolution of the velocities is given by the time derivative $\dot{q}(t)$ of the base integral curve. Since $\mathbb{F}h = (\mathbb{F}\ell)^{-1}$, the solution of the Hamiltonian system is given by $\mathbb{F}\ell(\dot{q}(t))$. Using the explicit expression of the mechanical connection and the notation given in the general procedure, the method of reconstruction simplifies to the following steps. To find the integral curve $c(t)$ of the simple mechanical system with G -symmetry $h = K + V \circ \pi_Q$ on T^*Q with initial condition $c(0) = \alpha_q \in T_q^*Q$, knowing the integral curve $c_\mu(t)$ of the reduced Hamiltonian system on $(T^*Q)_\mu$ given by the reduced Hamiltonian function $h_\mu: (T^*Q)_\mu \rightarrow \mathbb{R}$ with initial condition $c_\mu(0) = [\alpha_q]$ one proceeds in the following manner. Recall the symplectic embedding $\varphi_\mu: ((T^*Q)_\mu, (\Omega_Q)_\mu) \rightarrow (T^*(Q/G_\mu), \Omega_{Q/G_\mu} - B_\mu)$. The curve $\varphi_\mu(c_\mu(t)) \in T^*(Q/G_\mu)$ is an integral curve of the Hamiltonian system on $(T^*(Q/G_\mu), \Omega_{Q/G_\mu} - B_\mu)$ given by the function that is the sum of the kinetic energy of the quotient Riemannian metric and the quotient amended potential \hat{V}_μ . Let $q_\mu(t) := \pi_{Q/G_\mu}(c_\mu(t))$ be the base integral curve of this system, where $\pi_{Q/G_\mu}: T^*(Q/G_\mu) \rightarrow Q/G_\mu$ is the cotangent bundle projection.

- *Step 1:* Relative to the mechanical connection $A_{\text{mech}} \in \Omega^1(Q; \mathfrak{g}_\mu)$, horizontally lift $q_\mu(t) \in Q/G_\mu$ to a curve $q_b(t) \in Q$ passing through $q_b(0) = q$.
- *Step 2:* Determine $\xi(t) \in \mathfrak{g}_\mu$ from the algebraic system $\langle \langle \xi(t)_Q(q_b(t)), \eta_Q(q_b(t)) \rangle \rangle = \langle \mu, \eta \rangle$ for all $\eta \in \mathfrak{g}_\mu$, where $\langle \langle \cdot, \cdot \rangle \rangle$ is the G -invariant kinetic energy Riemannian metric on Q . This implies that $\dot{q}_b(0)$ and $\xi(0)_Q(q)$ are the horizontal and vertical components of the vector $\alpha_q^\sharp \in T_qQ$ which is associated by the metric $\langle \langle \cdot, \cdot \rangle \rangle$ to the initial condition α_q .
- *Step 3:* Solve $\dot{g}(t) = T_e L_{g(t)} \xi(t)$ in G_μ with initial condition $g(0) = e$.
- *Step 4:* The curve $q(t) := g(t) \cdot q_b(t)$, with $q_b(t)$ and $g(t)$ determined in steps 2 and 4, respectively, is the base integral curve of the simple mechanical system with symmetry defined by the function h satisfying $q(0) = 0$. The curve $(\mathbb{F}h)^{-1}(\dot{q}(t)) \in T^*Q$ is the integral curve of this system with initial

condition $c(0) = \alpha_q$. In addition, $q'(t) = g(t) \cdot (\dot{q}_b(t) + \xi(t)_Q(q_b(t)))$ is the horizontal plus vertical decomposition relative to the connection induced on $J^{-1}(\mu) \rightarrow (T^*Q)_\mu$ by the mechanical connection $A_{\text{mech}} \in \Omega^1(Q; \mathfrak{g}_\mu)$.

There are several important situations when step 3, the main obstruction to an explicit solution of the reconstruction problem, can be carried out. We shall review some of them below.

- (c1) *The case $G_\mu = S^1$.* If G_μ is abelian, the equation in step 3 has the solution $g(t) = \exp \int_0^t \xi(s) ds$. If, in addition, $G_\mu = S^1$, then $\xi(s)$ can be explicitly determined by step 2. Indeed, if $\zeta \in \mathfrak{g}_\mu$ is a generator of \mathfrak{g}_μ , writing $\xi(s) = a(s)\zeta$ for some smooth real-valued function a defined on some open interval around the origin, the algebraic equation in step 2 implies that $\langle \langle a(s)\xi(t)_Q(q_b(t)), \zeta_Q(q_b(t)) \rangle \rangle = \langle \mu, \zeta \rangle$, which gives $a(s) = \langle \mu, \zeta \rangle / \|\zeta_Q(q_b(s))\|^2$. Therefore, the base integral curve of the solution of the simple mechanical system with symmetry on T^*Q passing through q is

$$q(t) = \exp \left(\langle \mu, \zeta \rangle \int_0^t \frac{ds}{\|\zeta_Q(q_b(s))\|^2} \zeta \right) \cdot q_b(t)$$

and

$$\begin{aligned} \dot{q}(t) = & \exp \left(\langle \mu, \zeta \rangle \int_0^t \frac{ds}{\|\zeta_Q(q_b(s))\|^2} \zeta \right) \\ & \times \left(\dot{q}_b(t) + \frac{\langle \mu, \zeta \rangle}{\|\zeta_Q(q_b(s))\|^2} \zeta_Q(q_b(t)) \right) \end{aligned}$$

- (c2) *The case of compact Lie groups.* An obvious situation when the differential equation in step 3 can be solved is if $\xi(t) = \xi$ for all t , where ξ is a given element of \mathfrak{g}_μ . Then the solution is $g(t) = \exp(t\xi)$. However, step 2 puts certain restrictions under this hypothesis, because it requires that $\langle \langle \xi(t)_Q(q_b(t)), \eta_Q(q_b(t)) \rangle \rangle = \langle \mu, \eta \rangle$ for any $\eta \in \mathfrak{g}_\mu$. This is satisfied if there is a bilinear nondegenerate form (\cdot, \cdot) on \mathfrak{g} satisfying $(\zeta, \eta) = \langle \langle \zeta_Q(q), \eta_Q(q) \rangle \rangle$ for all $q \in Q$ and $\zeta, \eta \in \mathfrak{g}$. This implies that (\cdot, \cdot) is positive definite and invariant under the adjoint action of G on \mathfrak{g} , so semisimple Lie algebras of noncompact type are excluded. If G is compact, which ensures the existence of a positive adjoint invariant inner product on \mathfrak{g} , and $Q = G$, this condition implies that the kinetic energy metric is invariant under the adjoint action. There are examples in which such conditions are natural, such as in Kaluza–Klein theories. Thus, if G is a compact Lie

group and (\cdot, \cdot) is a positive-definite metric invariant under the adjoint action of G on \mathfrak{g} satisfying $(\zeta, \eta) = \langle \langle \zeta_Q(q), \eta_Q(q) \rangle \rangle$ for all $q \in Q$ and $\zeta, \eta \in \mathfrak{g}$, then the element $\xi(t)$ in step 2 can be chosen to be constant and is determined by the identity $(\xi, \cdot) = \mu|_{\mathfrak{g}_\mu}$ on \mathfrak{g}_μ . The solution of the equation on step 3 is then $g(t) = \exp(t\xi)$.

- (c3) *The case when $\xi(t)$ is proportional to $\xi(t)$.* Try to find a real-valued function $f(t)$ such that $g(t) = \exp(f(t)\xi(t))$ is a solution of the equation $\dot{g}(t) = T_e L_{g(t)} \xi(t)$ with $f(0) = 0$. This gives, for small t , the equation $f(t)\xi(t) + f(t)\dot{\xi}(t) = \xi(t)$, that is, it is necessary that $\xi(t)$ and $\dot{\xi}(t)$ be proportional. So, if $\dot{\xi}(t) = \alpha(t)\xi(t)$ for some known smooth function $\alpha(t)$, then this gives $f(t) = \int_0^t \exp(\int_r^s \alpha(r) dr) ds$.
- (c4) *The case of G_μ solvable.* Write $g(t) = \exp(f_1(t)\xi_1) \exp(f_2(t)\xi_2) \cdots \exp(f_n(t)\xi_n)$, for some basis $\{\xi_1, \xi_2, \dots, \xi_n\}$ of \mathfrak{g}_μ and some smooth real-valued functions $f_i, i = 1, 2, \dots, n$, defined around zero. It is known that if G_μ is solvable, the equation in step 3 can be solved by quadratures for the f_i .

Reconstruction Phases for Simple Mechanical Systems with S^1 Symmetry

Consider a simple mechanical system with symmetry G on the Riemannian manifold $(Q, \langle \cdot, \cdot \rangle)$ with G -invariant potential $V \in C^\infty(Q)$. If $\mu \in \mathfrak{g}^*$, let V_μ be the amended potential and $\tilde{V}_\mu \in C^\infty(Q/G_\mu)$ the induced function on the base. Let $c: [0, T] \rightarrow T^*Q$ be an integral curve of the system with Hamiltonian $h = K + V \circ \pi_Q$ and suppose that its projection $c_\mu: [0, T] \rightarrow (T^*Q)_\mu$ to the reduced space is a closed integral curve of the reduced system with Hamiltonian h_μ . The reconstruction phase associated to the loop $c_\mu(t)$ is the group element $g \in G_\mu$, satisfying the identity $c(T) = g \cdot c(0)$. We shall present two explicit formulas of the reconstruction phase for the case when $G_\mu = S^1$. Let $\zeta \in \mathfrak{g}_\mu = \mathbb{R}$ be a generator of the coadjoint isotropy algebra and write $c(T) = \exp(\varphi\zeta) \cdot c(0)$; in this case, φ is identified with the reconstruction phase and, as we shall see in concrete mechanical examples, it truly represents an angle.

If $G_\mu = S^1$, the G_μ -principal bundle $\pi_\mu: J^{-1}(\mu) \rightarrow (T^*Q)_\mu := J^{-1}(\mu)/G_\mu$ admits two natural connections: $A = (1/\mu\zeta)\theta_\mu \in \Omega^1(J^{-1}(\mu))$, where θ_μ is the pullback of the canonical 1-form on the cotangent bundle to the momentum level submanifold $J^{-1}(\mu)$, and $\pi_Q^* A_{\text{mech}} \in \Omega^1(J^{-1}(\mu))$. There is no reason to choose one connection over the other and thus there are two natural formulas for the reconstruction phase in this case. Let $c_\mu(t)$ be a periodic orbit of period T of the reduced system and denote also by h_μ the value of the Hamiltonian function on it.

Assume that D is a two-dimensional surface in $(T^*Q)_\mu$ whose boundary is the loop $c_\mu(t)$. Since the manifolds $(T^*Q)_\mu$ and $T^*(Q/S^1)$ are diffeomorphic (but not symplectomorphic), it makes sense to consider the base integral curve $q_\mu(t)$ obtained by projecting $c_\mu(t)$ to the base Q/S^1 , which is a closed curve of period T . Denote by

$$\langle \widehat{V}_\mu \rangle := \frac{1}{T} \int_0^T \widehat{V}_\mu(q_\mu(t)) dt$$

the average of \widehat{V}_μ over the loop $q_\mu(t)$. Let $q_b(t) \in Q$ be the A_{mech} -horizontal lift of $q_\mu(t)$ to Q and let χ be the A_{mech} -holonomy of the loop $q_\mu(t)$ measured from $q(0)$, the base point of $c(0)$; its expression is given by $\exp \chi = \exp(-\int \int_D B)$, where B is the curvature of the mechanical connection. Denote by ω_μ the reduced symplectic form on $(T^*Q)_\mu$. With these notations the phase φ is given by

$$\begin{aligned} \varphi &= \frac{1}{\mu\zeta} \iint_D \omega_\mu + \frac{2(b_\mu - \langle \widehat{V}_\mu \rangle)T}{\mu\zeta} \\ &= \chi + \mu\zeta \int_0^T \frac{ds}{\|\zeta_Q(q_b(s))\|^2} \end{aligned} \quad [3]$$

The first terms in both formulas are the so-called geometric phases because they carry only geometric information given by the connection, whereas the second terms are called the dynamic phases since they encapsulate information directly linked to the Hamiltonian. The expression of the total phase as a sum of a geometric and a dynamic phase is not intrinsic and is connection dependent. It can even happen that one of these summands vanishes. We shall consider now two concrete examples: the free rigid body and the heavy top.

Reconstruction Phases for the Free Rigid Body

The motion of the free rigid body is a geodesic with respect to a left-invariant Riemannian metric on $SO(3)$ given by the moment of inertia of the body. The phase space of the free rigid body motion is $T^*SO(3)$ and a momentum map $J: T^*SO(3) \rightarrow \mathbb{R}^3$ of the lift of left translation to the cotangent bundle is given by right translation to the identity element. We have identified here $\mathfrak{so}(3)$ with \mathbb{R}^3 by the Lie algebra isomorphism $x \in (\mathbb{R}^3, \times) \mapsto \hat{x} \in (\mathfrak{so}(3), [\cdot, \cdot])$, where $\hat{x}(y) = x \times y$, and $\mathfrak{so}(3)^*$ with \mathbb{R}^3 by the inner product on \mathbb{R}^3 . The reduced manifold $J^{-1}(\mu)/G_\mu$ is identified with the sphere $S^2_{\|\mu\|}$ in \mathbb{R}^3 of radius $\|\mu\|$ with the symplectic form $\omega_\mu = -dS/\|\mu\|$, where dS is the standard area form on $S^2_{\|\mu\|}$ and $G_\mu \cong S^1$ is the group of rotations around the axis μ . These concentric spheres are the coadjoint orbits of the Lie-Poisson space $\mathfrak{so}(3)^*$ and represent the level sets of the

Casimir functions that are all smooth functions of $\|\Pi\|^2$, where $\Pi \in \mathbb{R}^3$ denotes the body angular momentum.

The Hamiltonian of the rigid body on the Lie-Poisson space $T^*SO(3)/SO(3) \cong \mathbb{R}^3$ is given by

$$h(\Pi) := \frac{1}{2} \left(\frac{\Pi_1^2}{I_1} + \frac{\Pi_2^2}{I_2} + \frac{\Pi_3^2}{I_3} \right)$$

where $I_1, I_2, I_3 > 0$ are the principal moments of inertia of the body. Let $\mathbb{I} := \text{diag}(I_1, I_2, I_3)$ denote the moment of inertia tensor diagonalized in a principal-axis body frame. The Lie-Poisson bracket on \mathbb{R}^3 is given by $\{f, g\}(\Pi) = -\Pi \cdot (\nabla f(\Pi) \times \nabla g(\Pi))$ and the equation of motions are $\dot{\Pi} = \Pi \times \Omega$, where $\Omega \in \mathbb{R}^3$ is the body angular velocity given in terms of Π by $\Omega_i := \Pi/I_i$, for $i = 1, 2, 3$, that is, $\Omega = \mathbb{I}^{-1}\Pi$. The trajectories of these equations are found by intersecting a family of homothetic energy ellipsoids with the angular momentum concentric spheres. If $I_1 > I_2 > I_3$, one immediately sees that all orbits are periodic with the exception of four centers (the two possible rotations about the long and the short moment of inertia axis of the body), two saddles (the two rotations about the middle moment of inertia axis of the body), and four heteroclinic orbits connecting the two saddles.

Suppose that $\Pi(t)$ is a periodic orbit on the sphere $S^2_{\|\mu\|}$ with period T . After time T , by how much has the rigid body rotated in space? The answer to this question follows directly from [3]. Taking $\zeta = \mu/\|\mu\|$ and the potential $v \equiv 0$ we get

$$\begin{aligned} \varphi &= -\Lambda + \frac{2b_\mu T}{\|\mu\|} \\ &= \iint_D \frac{2\|\mathbb{I}\Pi(s)\|^2 - (\Pi(s) \cdot \mathbb{I}\Pi(s))(\text{tr } \mathbb{I})}{(\Pi(s) \cdot \mathbb{I}\Pi(s))^2} ds \\ &\quad + \|\mu\|^3 \int_0^T \frac{ds}{(\Pi(s) \cdot \mathbb{I}\Pi(s))} \end{aligned}$$

where D is one of the two spherical caps on $S^2_{\|\mu\|}$ whose boundary is the periodic orbit $\Pi(t)$, b_μ is the value of the total energy on the solution $\Pi(t)$, and Λ is the oriented solid angle, that is,

$$\Lambda := -\frac{1}{\|\mu\|} \iint_D \omega_\mu, \quad |\Lambda| = \frac{\text{area}D}{\|\mu\|^2}$$

Reconstruction Phases for the Heavy Top

The heavy top is a simple mechanical systems with symmetry S^1 on $T^*SO(3)$ whose Hamiltonian function is given by $h(\alpha_b) := (1/2)\|\alpha_b^\sharp\|^2 + Mgl\mathbf{k} \cdot b\chi$, where $b \in SO(3)$, $\alpha_b \in T_b^*SO(3)$, \mathbf{k} is the unit vector of the spatial Oz axis (pointing in the direction opposite to

that of the gravity force), $M \in \mathbb{R}$ is the total mass of the body, $g \in \mathbb{R}$ is the value of the gravitational acceleration, the fixed point about which the body moves is the origin, and χ is the unit vector of the straight line segment of length ℓ connecting the origin to the center of mass of the body. This Hamiltonian is left invariant under rotations about the spatial Oz axis. A momentum map induced by this S^1 -action is given by $J: T^*\text{SO}(3) \rightarrow \mathbb{R}, J(\alpha_b) = T_e^*L_b(\alpha_b) \cdot \mathbf{k}$; recall that $T_e^*L_b(\alpha_b) =: \Pi \in \mathbb{R}^3$ is the body angular momentum. The reduced space $J^{-1}(\mu)/S^1$ is generically the cotangent bundle of the unit sphere endowed with the symplectic structure given by the sum of the canonical form plus a magnetic term; equivalently, this is the coadjoint orbit in the dual of the Euclidean Lie algebra $\mathfrak{se}(3)^* = \mathbb{R}^3 \times \mathbb{R}^3$ given by $\mathcal{O}_\mu = \{(\Pi, \Gamma) \mid \Pi \cdot \Gamma = \mu, \|\Gamma\|^2 = 1\}$. The projection map $J^{-1}(\mu) \rightarrow \mathcal{O}_\mu$ implementing the symplectic diffeomorphism between the reduced space and the coadjoint orbit in $\mathfrak{se}(3)^*$ is given by $\alpha_b \mapsto (\Pi, \Gamma) := (T_e^*L_b(\alpha_b), b^{-1}\mathbf{k})$. The orbit symplectic form ω_μ on \mathcal{O}_μ has the expression $\omega_\mu(\Pi, \Gamma)((\Pi \times \mathbf{x} + \Gamma \times \mathbf{y}, \Gamma \times \mathbf{x}), (\Pi \times \mathbf{x}' + \Gamma \times \mathbf{y}', \Gamma \times \mathbf{x}')) = -\Pi \cdot (\mathbf{x} \times \mathbf{x}') - \Gamma \cdot (\mathbf{x} \times \mathbf{y}' - \mathbf{x}' \times \mathbf{y})$ for any $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}' \in \mathbb{R}^3$. The heavy-top equations $\dot{\Pi} = \Pi \times \Omega + \text{Mgl}\Gamma \times \chi, \dot{\Gamma} = \Gamma \times \Omega$ are Lie-Poisson equations on $\mathfrak{se}(3)^*$ for the Hamiltonian $h(\Pi, \Gamma) = (1/2)\Pi \cdot \Omega + \text{Mgl}\Gamma \cdot \chi$ and the Lie-Poisson bracket $\{f, g\}(\Pi, \Gamma) = -\Pi \cdot (\nabla_\Pi f \times \nabla_\Pi g) - \Gamma \cdot (\nabla_\Pi f \times \nabla_\Gamma g - \nabla_\Pi g \times \nabla_\Gamma f)$, where ∇_Π and ∇_Γ denote the partial gradients.

Let $(\Pi(t), \Gamma(t))$ be a periodic orbit of period T of the heavy-top equations. After time T , by how much has the heavy top rotated in space? The answer is provided by [3]:

$$\begin{aligned} \varphi &= \frac{1}{\mu} \iint_D \omega_\mu + \frac{1}{\mu} \left(2b_\mu T - 2\text{Mgl} \int_0^T \Gamma(s) \cdot \chi ds \right) \\ &= \iint_D \frac{2\|\Pi\Gamma(s)\|^2 - (\Gamma(s) \cdot \Pi\Gamma(s))(\text{tr } \mathbb{I})}{(\Gamma(s) \cdot \Pi\Gamma(s))^2} ds \\ &\quad + \int_0^T \frac{ds}{\Gamma(s) \cdot \Pi\Gamma(s)} \end{aligned}$$

where D is the spherical cap on the unit sphere whose boundary is the closed curve $\Gamma(t)$ and \mathcal{D} is a two-dimensional submanifold of the orbit \mathcal{O}_μ bounded by the closed integral curve $(\Pi(t), \Gamma(t))$. The first terms in each summand represent the geometric phase and the second terms the dynamic phase.

Gauged Poisson Structures

If the Lie group G acts freely and properly on a smooth manifold Q , then $(T^*Q)/G$ is a quotient Poisson manifold (see Poisson Reduction), where the quotient is taken relative to the (left) lifted cotangent

action. The leaves of this Poisson manifold are the orbit reduced spaces $J^{-1}(\mathcal{O}_\mu)/G$, where $\mathcal{O}_\mu \subset \mathfrak{g}^*$ is the coadjoint G -orbit through $\mu \in \mathfrak{g}^*$ (see Symmetry and Symplectic Reduction). Is there an explicit formula for this reduced Poisson bracket on a manifold diffeomorphic to $(T^*Q)/G$? It turns out that this question has two possible answers, once a connection on the principal bundle $\pi: Q \rightarrow Q/G$ is introduced. The discussion below will also link to the fibration version of cotangent bundle reduction.

In order to present these answers, we review two bundle constructions. Let G act freely and properly on the manifold P and consider the a (left) principal G -bundle $\rho: P \rightarrow P/G := M$. Let $\tau: N \rightarrow M$ be a surjective submersion. Then the pullback bundle $\tilde{\rho}: (n, p) \in \tilde{P} := \{(n, p) \in N \times P \mid \rho(p) = \tau(n)\} \mapsto n \in N$ over N is also a principal (left) G -bundle relative to the action $g \cdot (n, p) := (n, g \cdot p)$.

If there is a (left) G -action a manifold V , then the diagonal G -action $g \cdot (p, v) = (g \cdot p, g \cdot v)$ on $P \times V$ is also free and proper and one can form the associated bundle $P \times_G V := (P \times V)/G$ which is a locally trivial fiber bundle $\rho_E: [p, v] \in E := P \times_G V \mapsto \rho(p) \in M$ over M with fibers diffeomorphic to V . Analogously, one can form the associated fiber bundle $\rho_{\tilde{E}}: \tilde{E} := \tilde{P} \times_G V \rightarrow N$. Summarizing, the associated bundle $\tilde{E} = \tilde{P} \times_G V \rightarrow N$ is obtained from the principal bundle $\rho: P \rightarrow M$, the surjective submersion $\tau: N \rightarrow M$, and the G -manifold V by pullback and association, in this order.

These operations can be reversed. First, form the associated bundle $\rho_E: E = P \times_G V \rightarrow M$ and then pull it back by the surjective submersion $\tau: N \rightarrow M$ to N to get the pullback bundle $\tilde{\rho}_E: \tilde{E} \rightarrow N$. The map $\Phi: \tilde{P} \times_G V \rightarrow \tilde{E}$ defined by $\Phi([(n, p), v]) := (n, [p, v])$ is an isomorphism of locally trivial fiber bundles.

These general considerations will be used now to realize the quotient Poisson manifold $(T^*Q)/G$ in two different ways. Let Q be a manifold and G a Lie group (with Lie algebra \mathfrak{g}) acting freely and properly on it. Let $A \in \Omega^1(Q; \mathfrak{g})$ be a connection 1-form on the left G -principal bundle $\pi: Q \rightarrow Q/G$. Pull back the G -bundle $\pi: Q \rightarrow Q/G$ by the cotangent bundle projection $\pi_{Q/G}: T^*(Q/G) \rightarrow Q/G$ to $T^*(Q/G)$ to obtain the G -principal bundle $\tilde{\pi}_{Q/G}: (\alpha_{[q]}, q) \in \tilde{Q} := \{(\alpha_{[q]}, q) \mid [q] = \pi(q), q \in Q\} \mapsto \alpha_{[q]} \in T^*(Q/G)$. This bundle is isomorphic to the annihilator $(VQ)^\circ \subset T^*Q$ of the vertical bundle $VQ := \ker T\pi \subset TQ$. Next, form the coadjoint bundle $\rho_S: S := \tilde{Q} \times_G \mathfrak{g}^* \rightarrow T^*(Q/G)$ of $\tilde{Q}, \rho_S((\alpha_{[q]}, q), \mu) = \alpha_{[q]}$, that is, the associated vector bundle to the G -principal bundle $\tilde{Q} \rightarrow T^*(Q/G)$ given by the coadjoint representation of G on \mathfrak{g}^* . The connection-dependent map $\Phi_A: S \rightarrow (T^*Q)/G$ defined by $\Phi_A((\alpha_{[q]}, q), \mu) := [T_q^*\pi(\alpha_{[q]}) + A(q)^*\mu]$, where $q \in Q, \alpha_q \in T_q^*Q$, and

$\mu \in \mathfrak{g}^*$, is a vector bundle isomorphism over Q/G . The Sternberg space is the Poisson manifold $(S, \{ \cdot, \cdot \}_S)$, where $\{ \cdot, \cdot \}_S$ is the pullback to S by Φ_A of the quotient Poisson bracket on $(T^*Q)/G$.

Next, we proceed in the opposite order. Construct first the coadjoint bundle $\rho_{\mathfrak{g}^*} : [q, \mu] \in \tilde{\mathfrak{g}}^* := Q \times_G \mathfrak{g}^* \mapsto [q] \in Q/G$ associated to the principal bundle $\pi : Q \rightarrow Q/G$ and then pull it back by the cotangent bundle projection $\pi_{Q/G} : T^*(Q/G) \rightarrow Q/G$ to $T^*(Q/G)$ to obtain the vector bundle $\rho_W : W := \{(\alpha_{[q]}, [q, \mu]) \mid \pi_{Q/G}(\alpha_{[q]}) = \rho_{\mathfrak{g}^*}([q, \mu]) = [q]\}$, $\rho_W(\alpha_{[q]}, [q, \mu]) = \alpha_{[q]}$ over $T^*(Q/G)$. Note that $W = T^*(Q/G) \oplus \tilde{\mathfrak{g}}^*$ and hence W is also a vector bundle over Q/G . Let HQ be the horizontal sub-bundle defined by the connection A ; thus, $TQ = HQ \oplus VQ$, where $H_qQ := \ker A(q)$. For each $q \in Q$, the linear map $T_q\pi|_{H_qQ} : H_qQ \rightarrow T_{[q]}(Q/G)$ is an isomorphism. Let $\text{hor}_q := (T_q\pi|_{H_qQ})^{-1} : T_{[q]}(Q/G) \rightarrow H_qQ \subset T_qQ$ be the horizontal lift operator induced by the connection A . Thus, $\text{hor}_q^* : T_q^*Q \rightarrow T_{[q]}^*(Q/G)$ is a linear surjective map whose kernel is the annihilator $(H_qQ)^\circ$ of the horizontal space. The connection-dependent map $\Psi_A : (T^*Q)/G \rightarrow W$ defined by $\Psi_A([\alpha_q]) := (\text{hor}_q^*(\alpha_q), [q, J(\alpha_q)])$, where $q \in Q, \alpha_q \in T_q^*Q$, and $J : T^*Q \rightarrow \mathfrak{g}^*$ is the momentum map of the lifted action, $\langle J(\alpha_q), \xi \rangle = \alpha_q(\xi_Q(q))$ for $\xi \in \mathfrak{g}$, is a vector bundle isomorphism over Q/G and $\Psi_A \circ \Phi_A = \Phi$. The Weinstein space is the Poisson manifold $(W, \{ \cdot, \cdot \}_W)$, where $\{ \cdot, \cdot \}_W$ is the push-forward by Ψ_A of the Poisson bracket of $(T^*Q)/G$. In particular, $\Phi : S \rightarrow W$ is a connection independent Poisson diffeomorphism. The Poisson brackets on S and on W are called gauged Poisson brackets. They are expressed explicitly in terms of various covariant derivatives induced on S and on W by the connection $A \in \Omega^1(Q; \mathfrak{g})$.

Recall that the connection A on the principal bundle $\pi : Q \rightarrow Q/G$ naturally induces connections on pullback bundles and affine connections on associated vector bundles. Thus, both S and W carry covariant derivatives induced by A . They are given, according to general definitions, in the cases under consideration, by:

- If $f \in C^\infty(S), s = [(\alpha_{[q]}, q), \mu] \in S$, and $v_{\alpha_{[q]}} \in T_{\alpha_{[q]}} T^*(Q/G)$, then $d_{\tilde{A}}^S f(s) \in T_{\alpha_{[q]}} T^*(Q/G)$ is defined by $d_{\tilde{A}}^S f(s)(v_{\alpha_{[q]}}) := df(s)(T_{((\alpha_{[q]}, q), \mu)} \pi_{\tilde{Q} \times \mathfrak{g}^*}^{-1}((v_{\alpha_{[q]}}), \text{hor}_q(T_{\alpha_{[q]}} \tau(v_{\alpha_{[q]}}))), 0))$ where $\pi_{\tilde{Q} \times \mathfrak{g}^*} : \tilde{Q} \times \mathfrak{g}^* \rightarrow \tilde{Q} \times_G \mathfrak{g}^* = S$ is the orbit map. The symbol $d_{\tilde{A}}^S$ signifies that this is a covariant derivative on the associated bundle S induced by the connection \tilde{A} on the principal G -pullback bundle $\tilde{Q} \rightarrow T^*(Q/G)$. This connection \tilde{A} is the pullback connection defined by A .
- If $f \in C^\infty(W), w = (\alpha_{[q]}, [q, \mu]) \in W$, and $v_{\alpha_{[q]}} \in T_{\alpha_{[q]}} T^*(Q/G)$, then $\tilde{\nabla}_A^W f(w) \in T_{\alpha_{[q]}} T^*(Q/G)$ is defined

by $\tilde{\nabla}_A^W f(w)(v_{\alpha_{[q]}}) = df(w)(v_{\alpha_{[q]}}), T_{(q, \mu)} \pi_{Q \times \mathfrak{g}^*}^{-1}(\text{hor}_q(T_{\alpha_{[q]}} \tau_{Q/G}(v_{\alpha_{[q]}})), 0))$ where $\pi_{Q \times \mathfrak{g}^*} : Q \times \mathfrak{g}^* \rightarrow \tilde{Q} \times_G \mathfrak{g}^* = \tilde{\mathfrak{g}}^*$ is the orbit map. The symbol $\tilde{\nabla}_A$ signifies that this is a covariant derivative on the pullback bundle W induced by the covariant derivative ∇_A on the coadjoint bundle $\tilde{\mathfrak{g}}^*$. This covariant derivative ∇_A is induced on $\tilde{\mathfrak{g}}^*$ by the connection A .

- For $f \in C^\infty(W)$, we have $d_{\tilde{A}}^S(f \circ \Phi) = (\tilde{\nabla}_A^W f) \circ \Phi$.

To write the two gauged Poisson brackets on S and on W explicitly, we denote by $\tilde{\mathfrak{g}} = Q \times_G \mathfrak{g}$ the adjoint bundle of $\pi : Q \rightarrow Q/G$, by $\Omega_{Q/G}$ the canonical symplectic structure on $T^*(Q/G)$, by $B \in \Omega^2(Q; \mathfrak{g})$ the curvature of A , and by \mathcal{B} the $\tilde{\mathfrak{g}}$ -valued 2-form $\mathcal{B} \in \Omega^2(Q/G; \tilde{\mathfrak{g}})$ on the base Q/G defined by $\mathcal{B}([q])(u_{[q]}, v_{[q]}) = [q, B(q)(u_q, v_q)]$, for any $u_q, v_q \in T_qQ$ that satisfy $T_q\pi(u_q) = u_{[q]}$ and $T_q\pi(v_q) = v_{[q]}$. Note that both S^* and W^* are Lie algebra bundles, that is, their fibers are Lie algebras and the fiberwise Lie bracket operation depends smoothly on the base point. If $f \in C^\infty(S)$, denote by $df/ds \in S^* = \tilde{Q} \times_G \mathfrak{g}$ the usual fiber derivative of f . Similarly, if $f \in C^\infty(W)$ denote by $df/dw \in W^*$ the usual fiber derivative of f . Finally, $\sharp : T^*(T^*(Q/G)) \rightarrow T(T^*(Q/G))$ is the vector bundle isomorphism induced by $\Omega_{Q/G}$. The Poisson bracket of $f, g \in C^\infty(S)$ is given by

$$\begin{aligned} \{f, g\}_S(s) &= \Omega_{Q/G}(\alpha_{[q]}) \left(d_{\tilde{A}}^S f(s)^\sharp, d_{\tilde{A}}^S g(s)^\sharp \right) \\ &\quad - \left\langle s, \left[\frac{\delta f}{\delta s}, \frac{\delta g}{\delta s} \right] \right\rangle \\ &\quad + \left\langle v, (\pi_{Q/G}^* \mathcal{B})(\alpha_{[q]}) \left(d_{\tilde{A}}^S f(s)^\sharp, d_{\tilde{A}}^S g(s)^\sharp \right) \right\rangle \end{aligned}$$

where $v = [q, \mu] \in \tilde{\mathfrak{g}}^*$. The Poisson bracket $f, g \in C^\infty(W)$ is given by

$$\begin{aligned} \{f, g\}_W(w) &= \Omega_{Q/G}(\alpha_{[q]}) \left(\tilde{\nabla}_A^W f(w)^\sharp, \tilde{\nabla}_A^W g(w)^\sharp \right) \\ &\quad - \left\langle w, \left[\frac{\delta f}{\delta w}, \frac{\delta g}{\delta w} \right] \right\rangle \\ &\quad + \left\langle v, (\pi_{Q/G}^* \mathcal{B})(\alpha_{[q]}) \left(\tilde{\nabla}_A^W f(w)^\sharp, \tilde{\nabla}_A^W g(w)^\sharp \right) \right\rangle \end{aligned}$$

Note that their structure is of the form: “canonical” bracket plus a (left) “Lie–Poisson” bracket plus a curvature coupling term.

The Symplectic Leaves of the Sternberg and Weinstein Spaces

The map $\varphi_A : \tilde{Q} \times \mathfrak{g}^* \rightarrow T^*Q$ given by $\varphi_A((\alpha_{[q]}, q), \mu) := T_q^* \pi(\alpha_{[q]}) + A(q)^* \mu$, where $((\alpha_{[q]}, q), \mu) \in \tilde{Q} \times \mathfrak{g}^*$, is a G -equivariant diffeomorphism; the G -action on T^*Q is by cotangent lift and on $\tilde{Q} \times \mathfrak{g}^*$ is $g \cdot ((\alpha_{[q]}, q), \mu) = ((\alpha_{[q]}, g \cdot q), \text{Ad}_{g^{-1}}^* \mu)$. The pullback J_A

of the momentum map to $\tilde{Q} \times \mathfrak{g}^*$ has the expression $J_A((\alpha_{[q]}, q), \mu) = \mu$, so if $\mathcal{O} \subset \mathfrak{g}^*$ is a coadjoint orbit we have $J_A^{-1}(\mathcal{O}) = \tilde{Q} \times \mathcal{O}$, and hence the orbit reduced manifold $J_A^{-1}(\mathcal{O})/G$, whose connected components are the symplectic leaves of S , equals $\tilde{Q} \times_G \mathcal{O}$. Its symplectic form is the Sternberg minimal coupling form $\tilde{\omega}_{\tilde{Q}} + \rho_S^* \Omega_{\tilde{Q}/G}$.

In this formula, the 2-form $\tilde{\omega}_{\tilde{Q}}$ has not been defined yet. It is uniquely defined by the identity $\pi_{\tilde{Q} \times \mathfrak{g}^*}^* \tilde{\omega}_{\tilde{Q}} = d\hat{A} + \Pi_{\mathcal{O}} \omega_{\tilde{Q}}$, where $\omega_{\tilde{Q}}$ is the minus orbit symplectic form on \mathcal{O} (see Symmetry and Symplectic Reduction), $\Pi_{\mathcal{O}}: \tilde{Q} \times \mathcal{O} \rightarrow \mathcal{O}$ is the projection on the second factor, and $\hat{A} \in \Omega^2(\tilde{Q} \times \mathcal{O})$ is the 2-form given by $\hat{A}((\alpha_{[q]}, q), \mu) \langle (u_{\alpha_{[q]}}, v_q), \nu \rangle = -\langle \mu, A(q)(v_q) \rangle$ for $((\alpha_{[q]}, q), \mu) \in \tilde{Q} \times \mathcal{O}$, $(u_{\alpha_{[q]}}, v_q) \in T_{(\alpha_{[q]}, q)} \tilde{Q}$, and $\nu \in \mathfrak{g}^*$.

The symplectic leaves of the Weinstein space W are obtained by pushing forward by Φ the symplectic leaves of the Sternberg space. They are the connected components of the symplectic manifolds $(T^*(Q/G) \oplus (Q \times_G \mathcal{O}), \Pi_{T^*(Q/G)}^* \Omega_{Q/G} + \Pi_{Q \times_G \mathcal{O}}^* \omega_{\tilde{Q} \times_G \mathcal{O}})$, where \mathcal{O} is a coadjoint orbit in \mathfrak{g}^* , $\Omega_{Q/G}$ is the canonical symplectic form on $T^*(Q/G)$, $\omega_{\tilde{Q} \times_G \mathcal{O}}$ is a closed 2-form on $Q \times_G \mathcal{O}$ to be defined below, and $\Pi_{T^*(Q/G)}: T^*(Q/G) \oplus (Q \times_G \mathcal{O}) \rightarrow T^*(Q/G)$, $\Pi_{Q \times_G \mathcal{O}}: T^*(Q/G) \oplus (Q \times_G \mathcal{O}) \rightarrow Q \times_G \mathcal{O}$ are the projections. The closed 2-form $\omega_{\tilde{Q} \times_G \mathcal{O}} \in \Omega^2(Q \times_G \mathcal{O})$ is uniquely determined by the identity $\pi_{Q \times \mathcal{O}}^* \omega_{\tilde{Q} \times_G \mathcal{O}} = \omega_{\tilde{Q} \times \mathcal{O}}$, where $\pi_{Q \times \mathcal{O}}: Q \times \mathcal{O} \rightarrow Q \times_G \mathcal{O}$ is the orbit space projection, $\omega_{\tilde{Q} \times \mathcal{O}} \in \Omega^2(Q \times \mathcal{O})$ is closed and given by $\omega_{\tilde{Q} \times \mathcal{O}}(q, \mu) \langle (u_q, -\text{ad}_\xi^* \mu), (v_q, -\text{ad}_\eta^* \mu) \rangle := -d(A \times \text{id}_{\mathcal{O}})(q, \mu) \langle (u_q, -\text{ad}_\xi^* \mu), (v_q, -\text{ad}_\eta^* \mu) \rangle + \omega_{\tilde{Q}}(\mu) \langle \text{ad}_\xi^* \mu, \text{ad}_\eta^* \mu \rangle$, and $A \times \text{id}_{\mathcal{O}} \in \Omega^1(Q \times \mathfrak{g}^*)$ is given by $(A \times \text{id}_{\mathcal{O}})(q, \mu) \langle u_q, -\text{ad}_\xi^* \mu \rangle = \langle \mu, A(q)(u_q) \rangle$, for $q \in Q$, $\mu \in \mathfrak{g}^*$, $u_q, v_q \in T_q Q$, $\xi, \eta \in \mathfrak{g}$.

Thus, on the Sternberg and Weinstein spaces, both the Poisson bracket as well as the symplectic form on the leaves have explicit connection dependent formulas (see Gauge Theory: Mathematical Applications for a general treatment of gauge theories).

See also: Gauge Theory: Mathematical Applications; Hamiltonian Group Actions; Poisson Reduction; Symmetries and Conservation Laws; Symmetry and Symplectic Reduction.

Further Reading

- Abraham R and Marsden JE (1978) *Foundations of Mechanics*, 2nd edn. Reading, MA: Addison-Wesley.
- Guichardet A (1984) On rotation and vibration motions of molecules. *Annales de l'Institut Henri Poincaré. Physique Théorique* 40: 329–342.
- Iwai T (1987) A geometric setting for classical molecular dynamics. *Annales de l'Institut Henri Poincaré. Physique Théorique*. 47: 199–219.
- Kummer M (1981) On the construction of the reduced phase space of a Hamiltonian system with symmetry. *Indiana University Mathematics Journal* 30: 281–291.
- Lewis D, Marsden JE, Montgomery R, and Ratiu TS (1986) The Hamiltonian structure for dynamic free boundary problems. *Physica D* 18: 391–404.
- Marsden JE, Montgomery R, and Ratiu TS (1990) Reduction, symmetry, and phases in mechanics. *Memoirs of the American Mathematical Society* 88(436).
- Marsden JE, Misiolek G, Ortega J-P, Perlmutter M, and Ratiu TS (2005) *Hamiltonian Reduction by Stages*, Lecture Notes in Mathematics. Springer.
- Marsden JE and Perlmutter M (2000) The orbit bundle picture of cotangent bundle reduction. *Comptes Rendus Mathématiques de l'Académie des Sciences. La Société Royale du Canada* 22: 33–54.
- Marsden JE and Ratiu TS (2003) *Introduction to Mechanics and Symmetry*, 2nd edn. second printing; 1st edn. (1994), Texts in Applied Mathematics, vol. 17. New York: Springer.
- Montgomery R (1984) Canonical formulations of a particle in a Yang–Mills field. *Letters in Mathematical Physics* 8: 59–67.
- Montgomery R (1991) How much does a rigid body rotate? A Berry's phase from the eighteenth century. *American Journal of Physics* 59: 394–398.
- Montgomery R, Marsden JE, and Ratiu TS (1984) Gauged Lie Poisson structures. In: Marsden J (ed.) *Fluids and Plasmas: Geometry and Dynamics*, Contemporary Mathematics, vol. 28, pp. 101–114. Providence, RI: American Mathematical Society.
- Satzer WJ (1977) Canonical reduction of mechanical systems invariant under abelian group actions with an application to celestial mechanics. *Indiana, University Mathematics Journal* 26: 951–976.
- Simo JC, Lewis D, and Marsden JE (1991) The stability of relative equilibria. Part I: The reduced energy–momentum method. *Archive for Rational Mechanics and Analysis* 115: 15–59.
- Smale S (1970) Topology and mechanics. *Inventiones Mathematicae* 10: 305–331, 11: 45–64.
- Sternberg S (1977) Minimal coupling and the symplectic mechanics of a classical particle in the presence of a Yang–Mills field. *Proceedings of the National Academy of Sciences* 74: 5253–5254.
- Weinstein A (1978) A universal phase space for particles in Yang–Mills fields. *Letters in Mathematical Physics* 2: 417–420.
- Zaalani N (1999) Phase space reduction and Poisson structure. *Journal of Mathematical Physics* 40(7): 3431–3438.

Critical Phenomena in Gravitational Collapse

C Gundlach, University of Southampton,
Southampton, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Sufficiently dense concentrations of mass–energy in general relativity collapse irreversibly and form black holes. More precisely, the singularity theorems state that once a closed trapped surface has developed, some world lines will only extend to a finite length in the future – they end in a spacetime singularity. Furthermore, the cosmic censorship hypothesis states that this singularity is hidden away inside a black hole. One can, therefore, classify initial data in general relativity which describe an isolated system with no black hole present into those which remain regular, and those which form a black hole during their evolution.

Theorems on the stability of Minkowski spacetime, and similar results for some types of matter coupled to gravity, imply that sufficiently weak (in some technical sense) initial data will remain regular. On the other hand, no necessary or sufficient criterion for black hole formation is known. For very strong data the existence of a closed trapped surface implies black hole formation, but although the data themselves may be regular, the trapped surface must already be inside the black hole. Between the very weak and very strong regime, there is a middle regime of initial data for which one cannot decide if they will or will not form a black hole, other than evolving them in time.

The threshold between collapse and dispersion was first explored systematically by Choptuik (1992). He concentrated on the simple model of a spherically symmetric massless scalar (matter) field $\phi(r, t)$. In this model, the scalar-field matter must either form a black hole, or disperse to infinity – it cannot form stable stars. Choptuik explored the space of initial data by means of one-parameter families of initial data which interpolate between strong data (say with large parameter p) that form a black hole and weak data (with small p) that disperse. The critical value p_* of the parameter p can be found for each family by evolving many data sets from that family. Near the black hole threshold, Choptuik found the following phenomena:

1. *Mass scaling.* By fine-tuning the initial data to the threshold along any one-parameter family, one can make arbitrarily small black holes. Near the threshold, the black hole mass scales as

$$M \simeq C(p - p_*)^\gamma \quad \text{for } p \geq p_* \quad [1]$$

for the black hole mass M in the limit $p \rightarrow p_*$ from above.

2. *Universality.* While p_* and C depend on the particular one-parameter family of data, the critical exponent γ has a universal value, $\gamma \simeq 0.374$, for all one-parameter families of scalar-field data. Furthermore, for a finite time in a finite region of space, the solutions generated by all near-critical data approach one and the same solution ϕ_* , called the critical solution:

$$\phi(r, t) \simeq \phi_*\left(\frac{r}{L}, \frac{t - t_*}{L}\right) \quad [2]$$

The constants t_* and L depend again on the family of initial data, but $\phi_*(r, t)$ is universal. This universal phase ends when the evolution decides between black hole formation and dispersion. The universal critical solution is approached by any initial data that are sufficiently close to the black hole threshold, on either side, and from any one-parameter family.

3. *Scale-echoing.* The critical solution $\phi_*(r, t)$ is unchanged when one rescales space and time by a factor e^Δ :

$$\phi_*(r, t) = \phi_*(e^\Delta r, e^\Delta t) \quad [3]$$

where $\Delta \simeq 3.44$ for the scalar field.

The same phenomena were quickly discovered in many other types of matter coupled to gravity, and even in vacuum gravity (where gravitational waves can form black holes). The echoing period Δ and critical exponent γ depend on the type of matter, but the existence of the phenomena appears to be generic. For some types of matter (e.g., perfect fluid matter), the critical solution is continuously scale invariant (or continuously self-similar, CSS) in the sense that

$$\phi_*(r, t) = \phi_*(r/t) \quad [4]$$

rather than scale-periodic (or discretely self-similar, DSS) as in [3]. (We use the notation $\phi_*(x)$ for the function of one variable r/t .) We have described scale invariance and scale-echoing here in terms of coordinates, but these do admit geometric, coordinate-invariant definitions, which are not restricted to spherical symmetry.

There is also another kind of critical behavior at the black hole threshold. Here, too, the evolution goes through a universal critical solution, but it is static, rather than scale invariant. As a consequence, the mass of black holes near the threshold takes a universal finite value (some fixed fraction of the mass of the critical solution), instead of showing power-law

scaling. In an analogy with first- and second-order phase transitions in statistical mechanics, the critical phenomena with a finite mass at the black hole threshold are called type I, and the critical phenomena with power-law scaling of the mass are called type II.

At this point, we characterize the degree of rigor of the various parts of the theory that is summarized in this article. Critical phenomena were discovered in the numerical time evolution of generic asymptotically flat initial data. Numerical evolution of many elements of a specific one-parameter family, and fine-tuning to the black hole threshold along that family showed self-similarity and mass scaling near the threshold. Doing this for a number of randomly chosen one-parameter families suggests that these phenomena, and in particular the echoing scale Δ and mass-scaling exponent γ , are universal between initial data within one model (e.g., the spherical scalar field). Numerical experiments, however, can only explore a finite-dimensional subspace of the infinite-dimensional space of initial data (phase space) of the field theory, and so cannot prove universality.

We go further by applying the theory of dynamical systems to general relativity. The arguments summarized in the next section would be difficult to make rigorous, as the dynamical system under consideration is infinite dimensional, but they suggest a focus on fixed points of the dynamical system and their linear perturbations. Even though the dynamical systems motivation is not mathematically rigorous, the linearized analysis itself is a well-defined problem that can be solved numerically to essentially arbitrary precision. This proves universality on a perturbative level, and provides numerical values of Δ and γ . A combination of the global dynamical systems analysis and perturbative analysis even predicts further critical exponents for black hole charge and angular momentum. Finally, critical phenomena have been discovered in a number of systems (different types of matter and symmetry restrictions), and this suggests that they may be generic for some large class of field theories (although details such as the numerical values of γ and Δ do depend on the system), but there is no conclusive evidence for this at present.

The Dynamical Systems Picture

When we consider general relativity as an infinite-dimensional dynamical system, a solution curve is a spacetime. Points along the curve are Cauchy surfaces in the spacetime, which can be thought of as moments of time. An important difference between general relativity and other field theories

is that the same spacetime can be sliced in many different ways, none of which is preferred. Therefore, to turn general relativity into a dynamical system, one has to fix a slicing (and in practice also coordinates on each slice). In the example of the spherically symmetric massless scalar field, using polar slicing and an area radial coordinate r , a point in phase space can be characterized by the two functions

$$Z = \left\{ \phi(r), r \frac{\partial \phi}{\partial t}(r) \right\} \quad [5]$$

In spherical symmetry, there are no degrees of freedom in the scalar field, and Cauchy data for the metric can be reconstructed from Z using the Einstein constraints.

The phase space consists of two halves: initial data whose time evolution always remains regular, and data which contain a black hole or form one during time evolution. The numerical evidence collected from individual one-parameter families of data suggests that the black hole threshold that separates the two is a smooth hypersurface. The mass-scaling law [1] can, therefore, be restated without explicit reference to one-parameter families. Let P be any function on phase space such that data sets with $P > 0$ form black holes, and data with $P < 0$ do not, and which is analytic in a neighborhood of the black hole threshold $P = 0$. The black hole mass as a function on phase space is then given by

$$M \simeq F(P) P^\gamma \quad [6]$$

for $P > 0$, where $F(P) > 0$ is an analytic function.

Consider now the time evolution in this dynamical system, near the threshold (“critical surface”) between black hole formation and dispersion. A phase-space trajectory that starts out in a critical surface by definition never leaves it. A critical surface is, therefore, a dynamical system in its own right, with one dimension fewer. If it has an attracting fixed point, such a point is called a critical point. It is an attractor of codimension 1, and the critical surface is its basin of attraction. The fact that the critical solution is an attractor of codimension 1 is visible in its linear perturbations: it has an infinite number of decaying perturbation modes tangential to (and spanning) the critical surface, and a single growing mode not tangential to the critical surface.

Any trajectory beginning near the critical surface, but not necessarily near the critical point, moves almost parallel to the critical surface toward the critical point. As the phase point approaches the critical point, its movement parallel to the surface

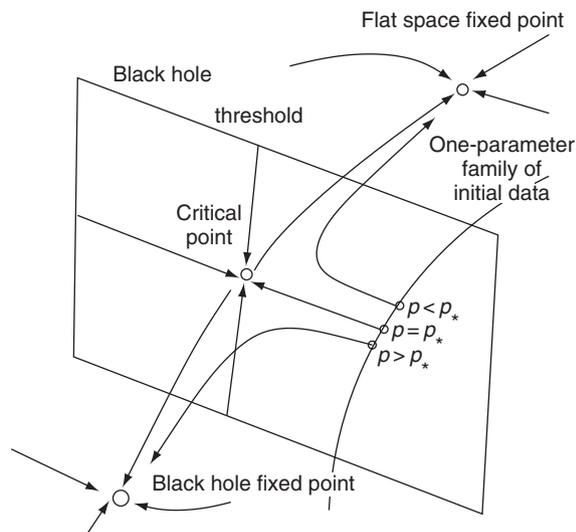


Figure 1 The phase-space picture for the black hole threshold in the presence of a critical point. The arrow lines are time evolutions, corresponding to spacetimes. The line without an arrow is not a time evolution, but a one-parameter family of initial data that crosses the black hole threshold at $p = p_*$. (Reproduced with permission from Gundlach C (2003) Critical phenomena in gravitational collapse. *Physics Reports* 376: 339–405.)

slows down, while its distance and velocity out of the critical surface are still small. The phase point spends sometime moving slowly near the critical point. Eventually, it moves away from the critical point in the direction of the growing mode, and ends up on an attracting fixed point.

This is the origin of universality: any initial data set that is close to the black hole threshold (on either side) evolves to a spacetime that approximates the critical spacetime for sometime. When it finally approaches either the dispersion fixed point or the black hole fixed point, it does so on a trajectory that appears to be coming from the critical point itself. All near-critical solutions are passing through one of these two funnels. All details of the initial data have been forgotten, except for the distance from the black hole threshold: the closer the initial phase point is to the critical surface, the more the solution curve approaches the critical point, and the longer it will remain close to it.

In all systems that have been examined, the black hole threshold contains at least one critical point. A fixed point of the dynamical system represents a spacetime with an additional continuous symmetry that generic solutions do not have. If the critical spacetime is time independent in the usual sense, we have type I critical phenomena; if the symmetry is scale invariance, we have type II critical phenomena. The attractor within the critical surface may also be a limit cycle, rather than a fixed point. In spacetime

terms this corresponds to a discrete symmetry (DSS rather than CSS in type II, or a pulsating critical solution, rather than a stationary one, in type I).

Self-Similarity and Mass Scaling

Type II critical phenomena occur where the critical solution is scale invariant (self-similar, CSS or DSS). Using suitable spacetime coordinates, a CSS solution can be characterized as independent of a time coordinate τ which is also a logarithmic scale. Similarly, a DSS solution can be characterized as periodic in τ . For example, starting from the scale periodicity [3] in polar-radial coordinates, we replace r and t by new coordinates

$$x \equiv -\frac{r}{t-t_*}, \quad \tau \equiv -\ln\left(-\frac{t-t_*}{L}\right) \quad [7]$$

where the accumulation time t_* and scale L must be matched to the one-parameter family under consideration. τ has been defined so that it increases as t increases and approaches t_* from below. It is useful to think of r , t , and L as having dimension length in units $c=G=1$, and of x and τ as dimensionless. Choptuik’s observation, expressed in these coordinates, is that in any near-critical solution there is a spacetime region where the fields Z are well approximated by the critical solution, or

$$Z(x, \tau) \simeq Z_*(x, \tau) \quad [8]$$

with

$$Z_*(x, \tau + \Delta) = Z_*(x, \tau) \quad [9]$$

Note that the time parameter of the dynamical system must be chosen as τ if a CSS solution is to be a fixed point, or a DSS solution a cycle. More generally (going beyond spherical symmetry), on any self-similar spacetime one can introduce coordinates $x^\mu = (\tau, x^1, x^2, x^3)$ in which the metric is of the form

$$g_{\mu\nu} = e^{-2\tau} \bar{g}_{\mu\nu} \quad [10]$$

and where $\bar{g}_{\mu\nu}$ is independent of τ for a CSS spacetime, and periodic in τ for a DSS spacetime. These coordinates are not unique.

The critical exponent γ can be calculated from the linear perturbations of the critical solution. In order to keep the notation simple, the discussion will be restricted to a critical solution that is spherically symmetric and CSS, which is correct, for example, for perfect-fluid matter.

Let us assume that we have fine-tuned initial data close to the black hole threshold so that in a region the resulting spacetime is well approximated by the CSS critical solution. This part of the spacetime

corresponds to the section of the phase-space trajectory that lingers near the critical point. In this region, we can linearize around Z_* . As Z_* does not depend on τ , its linear perturbations can depend on τ only exponentially. Labeling the perturbation modes by i , a single mode perturbation is of the form

$$\delta Z = C_i e^{\lambda_i \tau} Z_i(x) \quad [11]$$

In the near-critical regime, we can therefore approximate the solution as

$$Z(x, \tau) \simeq Z_*(x) + \sum_{i=0}^{\infty} C_i(p) e^{\lambda_i \tau} Z_i(x) \quad [12]$$

The notation $C_i(p)$ is used because the perturbation amplitudes C_i depend on the initial data, and hence on the parameter p that controls the initial data.

If Z_* is a critical solution, by definition there is exactly one λ_i with positive real part (in fact, it is purely real), say λ_0 . As $t \rightarrow t_*$ from below, which corresponds to $\tau \rightarrow \infty$, all other perturbations decay and can be neglected. By definition, the critical solution corresponds to $p = p_*$, and so we must have $C_0(p_*) = 0$. Linearizing around p_* , we obtain

$$Z(x, \tau) \simeq Z_*(x) + \left. \frac{dC_0}{dp} \right|_{p_*} (p - p_*) e^{\lambda_0 \tau} Z_0(x) \quad [13]$$

in a region of the spacetime.

Now we extract Cauchy data at one particular value of τ within that region, namely at τ_p defined by

$$\left. \frac{dC_0}{dp} \right|_{p_*} |p - p_*| e^{-\lambda_0 \tau_p} \equiv \epsilon \quad [14]$$

where ϵ is an arbitrary small constant, so that

$$Z(x, \tau_p) \simeq Z_*(x) \pm \epsilon Z_0(x) \quad [15]$$

where \pm is the sign of $p - p_*$, left behind because by definition ϵ is positive. As τ increases from τ_p , the growing perturbation becomes nonlinear and the approximation [13] breaks down. Then either a black hole forms (say for the positive sign), or the solution disperses (for the negative sign). We need not follow this nonlinear evolution in detail to find the black hole mass scaling in the former case: dimensional analysis is sufficient. Going back to coordinates t and r , we have

$$Z(r, t_p) \simeq Z_* \left(\frac{r}{L_p} \right) \pm \epsilon Z_0 \left(\frac{r}{L_p} \right) \quad [16]$$

where

$$L_p \equiv L e^{-\tau_p} \quad [17]$$

These Cauchy data at $t = t_p$ depend on the initial data at $t = 0$ only through the overall scale L_p , and through the sign in front of ϵ . If the field equations themselves are scale invariant, or asymptotically scale invariant at scales L_p and smaller, the black hole mass, which has dimensions of length in gravitational units, must be proportional to the initial data scale L_p , the only length scale that is present. Therefore,

$$M \propto L_p \propto (p - p_*)^{1/\lambda_0} \quad [18]$$

and we have found the critical exponent to be $\gamma = 1/\lambda_0$.

The Analogy with Statistical Mechanics

The existence of a threshold where a qualitative change takes place, universality, scale invariance, and critical exponents suggest that there is a mathematical analogy between type II critical phenomena and critical phase transitions in statistical mechanics.

In equilibrium statistical mechanics, observable macroscopic quantities, such as the magnetization of a ferromagnetic material, are derived as statistical averages over microstates of the system. The expected value of an observable is

$$\langle A \rangle = \sum_{\text{microstates}} A(\text{microstate}) e^{-H(\text{microstate}, \mu)} \quad [19]$$

The Hamiltonian H depends on the parameters μ , which comprise the temperature, parameters characterizing the system such as interaction energies of the constituent molecules, and macroscopic forces such as the external magnetic field. The objective of statistical mechanics is to derive relations between the macroscopic quantities A and parameters μ .

Phase transitions in thermodynamics are thresholds in the space of external forces μ at which the macroscopic observables A , or one of their derivatives, change discontinuously. In a ferromagnetic material at high temperatures, the magnetization \mathbf{m} of the material (alignment of atomic spins) is determined by the external magnetic field \mathbf{B} . At low temperatures, the material shows a spontaneous magnetization even at zero external field, which breaks rotational symmetry. With increasing temperature, the spontaneous magnetization \mathbf{m} decreases and vanishes at the Curie temperature T_* as

$$|\mathbf{m}| \sim (T_* - T)^\gamma \quad [20]$$

In the presence of a very weak external field, the spontaneous magnetization aligns itself with the external field \mathbf{B} , while its strength is, to leading order, independent of \mathbf{B} . The function $\mathbf{m}(\mathbf{B}, T)$,

therefore, changes discontinuously at $B=0$. The line $B=0$ for $T < T_*$ is, therefore, a line of first-order phase transitions between the possible directions of the spontaneous magnetization (in a one-dimensional system, between m up and m down). This line ends at the critical point ($B=0, T=T_*$) where the order parameter $|m|$ vanishes. The role of $B=0$ as the critical value of B is obscured by the fact that $B=0$ is singled out by symmetry.

A critical phase transition involves scale-invariant physics. One sign of this is that fluctuations appear on a large range of length scales between the underlying atomic scale and the scale of the sample. In particular, the atomic scale, and any dimensionful parameters associated with that scale, must become irrelevant at the critical point. This can be taken as the starting point for obtaining properties of the system at the critical point.

One first defines a semigroup acting on microstates: the renormalization group. Its action is to group together a small number of particles as a single particle of a fictitious new system, using some averaging procedure. Alternatively, this can also be done in Fourier space. One then defines a dual action of the renormalization group on the space of Hamiltonians by demanding that the partition function is invariant under the renormalization group action:

$$\sum_{\text{microstates}} e^{-H} = \sum_{\text{microstates}'} e^{-H'} \quad [21]$$

The renormalized Hamiltonian H' is in general more complicated than the original one, but it can be approximated by a fixed expression where only a finite number of parameters μ are adjusted. Fixed points of the renormalization group correspond to Hamiltonians with the parameters μ at their critical values. The critical value of any dimensional parameter μ must be zero (or infinity). Only dimensionless combinations can have nontrivial critical values.

The behavior of thermodynamical quantities at the critical point is in general not trivial to calculate. But the action of the renormalization group on length scales is given by its definition. The blowup of the correlation length ξ at the critical point is, therefore, the easiest critical exponent to calculate. We make contact with critical phenomena in gravitational collapse by considering the time evolution in coordinates (τ, x) as a renormalization group action. The calculation of the critical exponent for the black hole mass M is the precise analog of the calculation of the critical exponent for the correlation length ξ , substituting $T_* - T$ for $p - p_*$, and

taking into account that the τ -evolution in critical collapse is toward smaller scales, while the renormalization group flow goes toward larger scales: therefore, ξ diverges at the critical point, while M vanishes.

We have shown above that the black hole mass is controlled by one global function P on phase space. Clearly, P is the gravity equivalent of $T - T_*$ in the ferromagnet. But it is tempting to speculate (Gundlach 2002) that there is also a gravity equivalent of the external magnetic field B , which gives rise to a second independent critical exponent. At least in some situations, the angular momentum of the initial data can play this role. Note that, like B , angular momentum is a vector, with a critical value that is zero because all other values break rotational symmetry. Furthermore, the final black hole can have nonvanishing angular momentum, which must depend on the angular momentum of the initial data. The former is analogous to the magnetization m , the latter to the external field B . It can be shown that this analogy holds perturbatively for small angular momentum. Future numerical simulations will show if it goes further.

Universality and Cosmic Censorship

Critical phenomena in gravitational collapse first generated interest because a complicated self-similar structure and dimensionless numbers γ and Δ arise from generic initial data evolved by quite simple field equations. Another point of interest is the rather detailed analogy of phenomena in a deterministic field theory with critical phase transitions in statistical mechanics. But critical phenomena are important for general relativity mostly for a different reason.

Black holes are among the most important solutions of general relativity because of their universality: the black hole uniqueness theorems state that stable black holes are completely determined by their mass, angular momentum, and electric charge – the Kerr–Newman family of black holes. Perturbation theory shows that any perturbations of black holes from the Kerr–Newman solutions must be radiated away.

Critical solutions have a similar importance because they are generic intermediate states of the evolution that are also independent of the initial data. An important distinction is that critical solutions depend on the matter model, and are therefore less universal than black holes. However, critical phenomena in gravitational collapse seem to arise in axisymmetric vacuum spacetimes, and so are apparently not linked to the

presence of matter. Furthermore, they also arise in perfect-fluid matter with the equation of state $p = \rho/3$, which is that of an ultrarelativistic gas. This is a good approximation for matter at very high density, such as in the big bang. This is important because critical phenomena probe arbitrarily large matter densities or spacetime curvatures as the initial data are fine-tuned to the black hole threshold. At even higher densities, presumably on the Planck scale, scale invariance is again broken by quantum-gravity effects, and so critical phenomena will end there.

The cosmic censorship conjecture states that naked singularities do not arise from suitably generic initial data for suitably well-behaved matter. Critical phenomena in gravitational collapse have forced a tightening of this conjecture. Type II (self-similar) critical solutions contain a naked singularity, that is, a point of infinite spacetime curvature from which information can reach a distant observer. (By contrast, the singularity inside a black hole is hidden from distant observers.) On a kinematical level, this could be seen already from the form [10] of the metric. Because the critical solution is the end state for all initial data that are exactly on the black hole threshold, all initial data on the black hole threshold form a naked singularity. As type II critical phenomena appear to be generic at least in spherical symmetry, this means that in generic self-gravitating systems, the space of regular initial data that form naked singularities is larger than expected, namely of codimension 1. Excluding naked singularities from generic initial data may be the sharpest version of cosmic censorship one can now hope to prove.

Another point of interest in critical collapse is that it allows one to make a small region of arbitrarily high curvature from finite-curvature initial data. This may be a route for probing quantum-gravity effects. Similarly, one can make black holes that are much smaller than any length scale present in the initial data or the matter equation of state. An application has been suggested for this in cosmology, where primordial black holes could have masses much smaller than the Hubble scale at which they are created, rather than of the order of this scale.

Outlook

Critical phenomena in gravitational collapse are now well understood in spherical symmetry, both theoretically and in numerical simulations. In some matter models, the phenomenology is quite complicated, but it still fits into the basic picture outlined here.

The crucial question as to what happens beyond spherical symmetry remains largely unanswered at the time of writing. Perturbation theory around spherical symmetry suggests that critical phenomena are not restricted to exactly spherical situations. This is also supported by simulations in axisymmetric (highly nonspherical) vacuum gravity. Other simulations of nonspherical gravitational collapse which cover the necessary range of spacetime scales required to see critical phenomena are only just becoming available, and the results are not yet clear-cut. For collapse with angular momentum, no high-resolution calculations have yet been carried out. As the necessary techniques become available, one should be prepared for numerical simulations to make dramatic extensions or corrections to the picture of critical collapse drawn up here.

See also: Computational Methods in General Relativity: The Theory; Spacetime Topology, Causal Structure and Singularities; Stability of Minkowski Space; Stationary Black Holes.

Further Reading

- Abrahams AM and Evans CR (1993) Critical behavior and scaling in vacuum axisymmetric gravitational collapse. *Physical Review Letters* 70: 2980–2983.
- Choptuik MW (1993) Universality and scaling in gravitational collapse of a massless scalar field. *Physical Review Letters* 70: 9–12.
- Choptuik MW (1999) Critical behavior in gravitational collapse. *Progress of Theoretical Physics* 136 (suppl.): 353–365.
- Evans CR and Coleman JS (1994) Critical phenomena and self-similarity in the gravitational collapse of radiation fluid. *Physical Review Letters* 72: 1782–1785.
- Gundlach C (1999) *Living Reviews in Relativity* 2: 4 (published electronically at <http://www.livingreviews.org>).
- Gundlach C (2002) Critical gravitational collapse with angular momentum: from critical exponents to universal scaling functions. *Physical Review D* 65: 064019.

Current Algebra

G A Goldin, Rutgers University, Piscataway, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Certain commutation relations among the current density operators in quantum field theories define an infinite-dimensional Lie algebra. The original current algebra of Gell-Mann described weak and electromagnetic currents of the strongly interacting particles (hadrons), leading to the Adler–Weisberger formula and other important physical results. This helped inspire mathematical and quantum-theoretic developments such as the Sugawara model, light cone currents, Virasoro algebra, the mathematical theory of affine Kac–Moody algebras, and non-relativistic current algebra in quantum and statistical physics. Lie algebras of local currents may be the infinitesimal representations of loop groups, local current groups or gauge groups, diffeomorphism groups, and their semidirect products or other extensions. Broadly construed, current algebra thus leads directly into the representation theory of infinite-dimensional groups and algebras. Applications have ranged across conformally invariant field theory, vertex operator algebras, exactly solvable lattice and continuum models in statistical physics, exotic particle statistics and q -commutation relations, hydrodynamics and quantized vortex motion. This brief survey describes but a few highlights.

Relativistic Local Current Algebra for Hadrons

To model superfluidity, Landau had proposed in 1941 a quantum hydrodynamics fundamentally based on local fluid densities and currents as (operator) dynamical variables. However, current algebra came into its own in theoretical physics with the ideas of Gell-Mann in the early 1960s. The basic concept, in the era just preceding quantum chromodynamics (QCD), was that even without knowing the Lagrangian governing hadron dynamics in detail, exact kinematical information – the local symmetry – could still be encoded in an algebra of currents. The local (vector and axial vector) current density operators, expressed where possible in terms of underlying quantized field operators in Hilbert space, were to form two octets of Lorentz 4-vectors, with each octet corresponding to the eight generators of the compact Lie group SU(3).

More specifically (Adler and Dashen 1968), let $\mathcal{F}_a^\mu(x)$, $a = 1, 2, \dots, 8$, $\mu = 0, 1, 2, 3$, be an octet of hadronic vector currents, where as usual $x = (x^\nu) = (x^0, \mathbf{x})$ denotes a point in four-dimensional spacetime. Likewise, introduce an axial vector octet $\mathcal{F}_a^{5\mu}(x)$. Unless otherwise specified, we use natural units, where $\hbar = 1$ and $c = 1$. Define the corresponding charges F_a and F_a^5 to be the space integrals of the time components of these currents, that is,

$$\begin{aligned} F_a(x^0) &= \int d^3x \mathcal{F}_a^0(x^0, \mathbf{x}) \\ F_a^5(x^0) &= \int d^3x \mathcal{F}_a^{50}(x^0, \mathbf{x}) \end{aligned} \quad [1]$$

where $d^3x = dx^1 dx^2 dx^3$. Then F_1, F_2, F_3 are the three components I_1, I_2, I_3 of the isotopic spin, and $Y = (2\sqrt{3}/3)F_8$ is the hypercharge. The usual electromagnetic current $J_{\text{em}}^\mu(x^0, \mathbf{x})$ is given by

$$J_{\text{em}}^\mu = q \left(\mathcal{F}_3^\mu + \frac{\sqrt{3}}{3} \mathcal{F}_8^\mu \right) \quad [2]$$

where q is the unit elementary charge, and the total charge is given by $Q = \int d^3x J_{\text{em}}^0(x^0, \mathbf{x}) = q(I_3 + Y/2)$. The hadronic part of the weak current entering an effective Lagrangian can be written as

$$\begin{aligned} J_w^\mu &= \left[(\mathcal{F}_1^\mu - \mathcal{F}_1^{5\mu}) + i(\mathcal{F}_2^\mu - \mathcal{F}_2^{5\mu}) \right] \cos \theta_C \\ &+ \left[(\mathcal{F}_4^\mu - \mathcal{F}_4^{5\mu}) + i(\mathcal{F}_5^\mu - \mathcal{F}_5^{5\mu}) \right] \sin \theta_C \end{aligned} \quad [3]$$

where θ_C is the Cabibbo angle (determined experimentally to be ~ 0.27 rad). The terms with $\mathcal{F}_1 - \mathcal{F}_1^5$ and $\mathcal{F}_2 - \mathcal{F}_2^5$ are strangeness conserving, those with $\mathcal{F}_4 - \mathcal{F}_4^5$ and $\mathcal{F}_5 - \mathcal{F}_5^5$ are not.

The main current algebra hypothesis is that the time components \mathcal{F}^0 and \mathcal{F}^{50} of these octets satisfy the equal-time commutation relations:

$$\begin{aligned} &[\mathcal{F}_a^0(x^0, \mathbf{x}), \mathcal{F}_b^0(y^0, \mathbf{y})]_{x^0=y^0} \\ &= i\delta^{(3)}(\mathbf{x} - \mathbf{y}) \sum_d c_{abd} \mathcal{F}_d^0(x^0, \mathbf{x}) \\ &[\mathcal{F}_a^0(x^0, \mathbf{x}), \mathcal{F}_b^{50}(y^0, \mathbf{y})]_{x^0=y^0} \\ &= i\delta^{(3)}(\mathbf{x} - \mathbf{y}) \sum_d c_{abd} \mathcal{F}_d^{50}(x^0, \mathbf{x}) \\ &[\mathcal{F}_a^{50}(x^0, \mathbf{x}), \mathcal{F}_b^{50}(y^0, \mathbf{y})]_{x^0=y^0} \\ &= i\delta^{(3)}(\mathbf{x} - \mathbf{y}) \sum_d c_{abd} \mathcal{F}_d^{50}(x^0, \mathbf{x}) \end{aligned} \quad [4]$$

where the c_{abd} are structure constants of the Lie algebra of SU(3), antisymmetric in the indices. Since current commutators relate bilinear expressions to

linear ones, they fix the normalizations of the currents. The chiral currents $\mathcal{F}_a^{L\mu} = (1/2)(\mathcal{F}_a^\mu - \mathcal{F}_a^{5\mu})$ and $\mathcal{F}_a^{R\mu} = (1/2)(\mathcal{F}_a^\mu + \mathcal{F}_a^{5\mu})$ commute with each other, so that the local current algebra decomposes into two independent pieces.

The Dirac δ -functions in eqns [4] require that \mathcal{F}_a^0 and \mathcal{F}_a^{50} be interpreted as (unbounded) operator-valued distributions; while the fixed-time condition suggests these should make mathematical sense as three-dimensional distributions, with x^0 held constant. Such distributions may be modeled on the test-function space \mathcal{D} of real-valued, compactly supported, C^∞ functions on the spacelike hyperplane \mathbf{R}^3 . For functions $f_a, f_a^5 \in \mathcal{D}$, one has formally the “smeared currents” that are expected to be bona fide (unbounded) operators in Hilbert space; suppressing x^0 ,

$$\begin{aligned} \mathcal{F}_a^0(f_a) &= \int_{\mathbf{R}^3} d^3x f_a(\mathbf{x}) \mathcal{F}_a^0(x^0, \mathbf{x}) \\ \mathcal{F}_a^{50}(f_a^5) &= \int_{\mathbf{R}^3} d^3x f_a^5(\mathbf{x}) \mathcal{F}_a^{50}(x^0, \mathbf{x}) \end{aligned} \tag{5}$$

Equations [4] then become

$$\begin{aligned} [\mathcal{F}_a^0(f_a), \mathcal{F}_b^0(f_b)] &= [\mathcal{F}_a^{50}(f_a), \mathcal{F}_b^{50}(f_b)] \\ &= i \sum_d \mathcal{F}_d^0(c_{abdf_a f_b}) \\ [\mathcal{F}_a^0(f_a), \mathcal{F}_b^{50}(f_b)] &= i \sum_d \mathcal{F}_d^{50}(c_{abdf_a f_b}) \end{aligned} \tag{6}$$

Let $g(\mathbf{x})$ be a C^∞ map from \mathbf{R}^3 to the Lie algebra \mathcal{G} of chiral $SU(3) \times SU(3)$, equal to zero outside a compact set. The set of all such \mathcal{G} -valued functions forms an infinite-dimensional Lie algebra under the pointwise bracket, $[g, g'](\mathbf{x}) = [g(\mathbf{x}), g'(\mathbf{x})]$. Let us call this Lie algebra $\text{map}_0(\mathbf{R}^3, \mathcal{G})$, where the subscript 0 indicates the condition of compact support when that is applicable (on compact manifolds, we omit the subscript). Expanding $g(\mathbf{x})$ with respect to a fixed basis of \mathcal{G} , we straightforwardly identify the map g with the two octets of test functions f_a and f_a^5 . Then, defining $\mathcal{F}(g) = \sum_a \mathcal{F}_a^0(f_a) + \sum_a \mathcal{F}_a^{50}(f_a^5)$, eqns [6] are interpreted (for fixed x^0) as a representation \mathcal{F} of $\text{map}_0(\mathbf{R}^3, \mathcal{G})$.

Integrating out the spatial variables entirely using eqns [1] leads to a representation at x^0 of \mathcal{G} by the charges F_a and F_a^5 . The Adler–Weisberger sum rule was first derived (in 1965) from the commutation relations of these charges, together with the assumption of a partially conserved axial-vector current (PCAC). It connected nucleon β -decay coupling with pion–nucleon scattering cross sections, agreeing well with experiment. Various low-energy theorems followed, also in accord with experiment. Shortly thereafter, Adler was able to eliminate the PCAC assumption, and derived a further sum rule going

beyond an experimental test of the algebra of charges to test the actual local current algebra. Here, the prediction pertained to structure functions in the deep inelastic scattering of neutrinos. This was elaborated by Bjorken to inelastic electron scattering. On the theoretical side, the study of the chiral current in perturbation theory led into the theory of anomalies. All these ideas were highly influential in subsequent theoretical work (Treiman *et al.* 1985, Mickelsson 1989).

It is a natural idea to try to extend eqns [4] or [6], which elegantly express the combined ideas of locality and symmetry, to an equal-time commutator algebra that would also include the space components of the local currents $\mathcal{F}_a^k, k = 1, 2, 3$. One may write without difficulty the commutators of the charges in [1] with these space components:

$$\begin{aligned} [F_a(x^0), \mathcal{F}_b^k(x^0, \mathbf{x})] &= [F_a^5(x^0), \mathcal{F}_b^{5k}(x^0, \mathbf{x})] \\ &= i \sum_d c_{abd} \mathcal{F}_d^k(x^0, \mathbf{x}) \\ [F_a(x^0), \mathcal{F}_b^{5k}(x^0, \mathbf{x})] &= [F_a^5(x^0), \mathcal{F}_b^k(x^0, \mathbf{x})] \\ &= i \sum_d c_{abd} \mathcal{F}_d^{5k}(x^0, \mathbf{x}) \end{aligned} \tag{7}$$

But the commutator of the local time component with the local space component of the current cannot be merely the obvious extrapolation from eqns [4] and [7], that is, it cannot be

$$\begin{aligned} [\mathcal{F}_a(x^0, \mathbf{x}), \mathcal{F}_b^k(y^0, \mathbf{y})]_{x^0=y^0} \\ = i\delta^{(3)}(\mathbf{x} - \mathbf{y}) \sum_d c_{abd} \mathcal{F}_d^k(x^0, \mathbf{x}) \end{aligned}$$

and so forth. Under very general conditions, for a relativistic theory based on local quantum fields or local observables, additional “Schwinger terms” are required on the right-hand sides of such commutators (Renner 1968).

Well-known difficulties in specifying the Schwinger terms are associated with the fact that operator-valued distributions are singular when regarded as if they were functions of spacetime points. Thus, the product of two distributions at a point is often singular or undefined. When the currents forming a local current algebra are written as normal-ordered products of field operator distributions and their derivatives, the Schwinger terms in their commutation relations may be calculated, for example, by “splitting points” in the arguments of the underlying fields, and subsequently letting the separation tend toward zero. The general form of a Schwinger term typically involves the derivative of a δ -function times an operator. This may be a multiple of the identity (i.e., a c -number) or not, depending on the underlying

field-theoretic model. Furthermore, when the number of spacetime dimensions is greater than $1 + 1$, the c -number Schwinger terms turn out to be infinite. Hence, we do not obtain this way a bona-fide infinite-dimensional, equal-time commutator algebra comprising all the components of the local currents.

Sugawara, Kac–Moody, and Virasoro Algebras

Since equations such as [4] and [6] are not explicitly dependent on how the currents are constructed from underlying canonical fields, one has the possibility of writing a theory entirely in terms of self-adjoint currents as the dynamical variables, bypassing the field operators entirely, and expressing a Hamiltonian operator directly in terms of such local currents. This is in the spirit of approaches to quantum field theory based on local algebras of observables. It suggests consideration of relativistic current algebras with finite c -number or operator Schwinger terms in $s + 1$ dimensions, $s \geq 1$.

The Sugawara model, which is of this type, turned out to be one of the most influential of those proposed in the late 1960s and early 1970s. Henceforth, let G be a compact Lie group, and \mathcal{G} its Lie algebra; let $F_a, a = 1, \dots, \dim \mathcal{G}$, be a basis for \mathcal{G} , with $[F_a, F_b] = i \sum_d c_{abd} F_d$. The Sugawara current algebra, at the fixed time $x^0 = y^0$ (which, from here on, we suppress in the notation), is given by

$$\begin{aligned} [J_a^0(\mathbf{x}), J_b^0(\mathbf{y})] &= i \delta^{(3)}(\mathbf{x} - \mathbf{y}) \sum_d c_{abd} J_d^0(\mathbf{x}) \\ [J_a^0(\mathbf{x}), J_b^k(\mathbf{y})] &= i \delta^{(3)}(\mathbf{x} - \mathbf{y}) \sum_d c_{abd} J_d^k(\mathbf{x}) \\ &\quad + i c \delta_{ab} \frac{\partial}{\partial x^k} \delta^{(3)}(\mathbf{x} - \mathbf{y}) I \\ [J_a^k(\mathbf{x}), J_b^\ell(\mathbf{y})] &= 0 \quad (k, \ell = 1, 2, 3) \end{aligned} \tag{8}$$

where $J_a^\mu = (J_a^0, J_a^k), k = 1, 2, 3$, is again a 4-vector, c is a finite constant, and I is the identity operator. The time components in eqns [8] behave like the local currents in eqns [4]. The Schwinger term is a c -number, while setting the commutators of the space components to zero is the simplest choice consistent with the Jacobi identity. The Sugawara Hamiltonian is given in terms of the local currents by the formal expression:

$$H = \frac{1}{2c} \sum_a \int_{\mathbb{R}^3} d^3x \left[J_a^0(\mathbf{x})^2 + \sum_{k=1}^3 J_a^k(\mathbf{x})^2 \right] \tag{9}$$

where the pointwise products of the currents require interpretation in the particular representation. This Hamiltonian leads to current conservation equations for the J_a^μ .

Related to the Sugawara current algebra, with $s = 1$ and the spatial dimension compactified, are affine Kac–Moody and Virasoro algebras (Goddard and Olive 1986, Kac 1990). Consider the infinite-dimensional Lie algebra $\text{map}(S^1, \mathcal{G})$ of smooth functions from the circle to \mathcal{G} under the pointwise bracket. This is also called a loop algebra. Referring to the basis F_a , define $T_a^{(m)}$ for integer m to be the Fourier function $\theta \rightarrow F_a \exp[-im\theta]$. The pointwise bracket in $\text{map}(S^1, \mathcal{G})$ gives $[T_a^{(m)}, T_b^{(n)}] = i \sum_d c_{abd} T_d^{(m+n)}$ for these generators. The corresponding (untwisted) affine Kac–Moody algebra is a (uniquely defined, nontrivial) one-dimensional central extension of this loop algebra – that is, the new generator commutes with all elements of the Lie algebra and, in an irreducible representation, must be a multiple of the identity. In such a representation, the new bracket can be written as

$$[T_a^{(m)}, T_b^{(n)}] = i \sum_d c_{abd} T_d^{(m+n)} + km \delta_{ab} \delta_{m,-n} I \tag{10}$$

where k is a constant. Here, $T_a^{(m=0)}$ is again a representation of \mathcal{G} . Self-adjointness of the local currents in the representation imposes the condition $T_a^{(m)*} = T_a^{(-m)}$.

Now the compactly supported C^∞ (tangent) vector fields on a C^∞ manifold M form a natural Lie algebra under the Lie bracket, denoted by $\text{vect}_0(M)$. In local Euclidean coordinates, for $\mathbf{g}_1, \mathbf{g}_2 \in \text{vect}_0(M)$, one can write this bracket as

$$[\mathbf{g}_1, \mathbf{g}_2] = \mathbf{g}_1 \cdot \nabla \mathbf{g}_2 - \mathbf{g}_2 \cdot \nabla \mathbf{g}_1 \tag{11}$$

As the affine Kac–Moody algebras are central extensions of the algebra of \mathcal{G} -valued functions on S^1 , so are Virasoro algebras central extensions of the algebra of vector fields on S^1 . Let $L^{(m)}$ denote the (complexified) vector field described by $\exp[-im\theta](1/i)\partial/\partial\theta$, for integer m . These generators then satisfy $[L^{(m)}, L^{(n)}] = (m - n)L^{(m+n)}$. Adjoining to the Lie algebra of vector fields a new central element (commuting with all the $L^{(m)}$), the Virasoro bracket in an irreducible representation is given by the formula

$$\begin{aligned} [L^{(m)}, L^{(n)}] &= (m - n)L^{(m+n)} \\ &\quad + c \frac{(m + 1)m(m - 1)}{12} \delta_{m,-n} I \end{aligned} \tag{12}$$

where the numerical coefficient c is called the Virasoro central charge; self-adjointness of the currents imposes $L^{(m)*} = L^{(-m)}$. It is straightforward to verify that eqn [12] satisfies the Jacobi identity. The special form of the central term in the Virasoro current algebra results from the Gelfand–Fuks cohomology on the algebra of vector fields.

The Kac–Moody and Virasoro algebras, both modeled on S^1 , may be combined to form a natural semidirect sum of Lie algebras, with the additional bracket

$$[T_a^{(m)}, L^{(n)}] = mT_a^{(m+n)} \tag{13}$$

Roughly speaking, the Kac–Moody generators correspond to Fourier transforms of charge densities on S^1 , whereas the Virasoro generators correspond to Fourier transforms of infinitesimal motions in S^1 . The central extensions provide the finite, c -number Schwinger terms. These structures have important application to light cone current algebra, conformally invariant quantum field theories in $(1 + 1)$ -dimensional spacetime, the quantum theory of strings, exactly solvable models in statistical mechanics, and many other domains.

Of greatest physical importance, both in quantum field theory and statistical mechanics, are those irreducible, self-adjoint representations of the Virasoro algebra known as highest weight representations, where the spectrum of the operator $L^{(m=0)}$ is bounded below. In these applications, one represents a pair of Virasoro algebras by mutually commuting sets of operators $L^{(m)}$ and $\bar{L}^{(m)}$. In the quantum theory, for example, one takes the total energy $H \propto \bar{L}^{(0)} + L^{(0)}$, and the total momentum $P \propto \bar{L}^{(0)} - L^{(0)}$. In a highest weight representation, there is a unique eigenstate of $L^{(0)}$ having the lowest eigenvalue \hbar ; for this “vacuum” $|\hbar\rangle$, $L^{(m)}|\hbar\rangle = 0$, $m > 0$.

Friedan, Qiu, and Shenker showed in 1984 that highest weight representations are characterized by a class of specific, non-negative values of the central charge c and, correspondingly, of \hbar : either $c \geq 1$ (and $\hbar \geq 0$) or $c = 1 - 6(\ell + 2)^{-1}(\ell + 3)^{-1}$, $\ell = 1, 2, 3, \dots$ (and \hbar assumes a corresponding, specified set of values for each value of ℓ). In a beautiful application to the study of the critical behavior of well-known statistical systems, in which the generator of dilations is proportional to $\bar{L}^{(0)} + L^{(0)}$, they discovered a direct correspondence with permitted values of the central charge; thus, $c = 1/2$ for the Ising model, $c = 7/10$ for the tricritical Ising model, $c = 4/5$ for the three-state Potts model, and $c = 6/7$ for the tricritical three-state Potts model.

Current Algebras and Groups

Local current algebras may be exponentiated to obtain corresponding infinite-dimensional topological groups (Pressley and Segal 1986, Mickelsson 1989, Kac 1990). Let G be a Lie group whose Lie algebra is \mathcal{G} . The algebra $\text{map}_0(M, \mathcal{G})$, consisting of smooth, compactly supported \mathcal{G} -valued functions on

M under the pointwise bracket, exponentiates to the local current group $\text{Map}_0(M, G)$, consisting of smooth maps from M to G that are the identity outside a compact set in M , under the pointwise group operation. When M is taken to be the four-dimensional spacetime manifold (rather than a spacelike hyperplane), the local current group modeled on M is mathematically a gauge group for nonabelian gauge field theory.

Likewise, the algebra $\text{vect}_0(M)$ exponentiates to the group $\text{Diff}_0(M)$ of compactly supported C^∞ diffeomorphisms of M (under composition). The Kac–Moody and Virasoro algebras exponentiate to central extensions of the loop group $\text{Map}(S^1, G)$ and the diffeomorphism group $\text{Diff}(S^1)$, respectively. The semidirect sums of the Lie algebras are the infinitesimal generators of semidirect products of the groups.

Under appropriate technical conditions, self-adjoint representations of current algebras generate (and may be obtained from) continuous unitary representations of the corresponding groups. The needed technical conditions have to do with the existence of a dense set of analytic vectors belonging to a common, dense invariant domain of essential self-adjointness for the currents.

Nonrelativistic Current Algebra

In nonrelativistic local current algebra, Schwinger terms do not appear. In 1968, Dashen and Sharp defined (at fixed time t , suppressed in the present notation) a mass density $\rho(\mathbf{x}) = m\psi^*(\mathbf{x})\psi(\mathbf{x})$ and a momentum density $\mathbf{J}(\mathbf{x}) = (\hbar/2i)\{\psi^*(\mathbf{x})\nabla\psi(\mathbf{x}) - [\nabla\psi^*(\mathbf{x})]\psi(\mathbf{x})\}$, where ψ is a second-quantized canonical field; here we keep \hbar in the notation. The resulting equal-time algebra is the semidirect sum:

$$\begin{aligned} [\rho(\mathbf{x}), \rho(\mathbf{y})] &= 0 \\ [\rho(\mathbf{x}), J^k(\mathbf{y})] &= -i\hbar \frac{\partial}{\partial x^k} [\delta^{(3)}(\mathbf{x} - \mathbf{y})\rho(\mathbf{x})] \\ [J^k(\mathbf{x}), J^\ell(\mathbf{y})] &= i\hbar \left\{ \frac{\partial}{\partial y^k} [\delta^{(3)}(\mathbf{x} - \mathbf{y})J^\ell(\mathbf{y})] \right. \\ &\quad \left. - \frac{\partial}{\partial x^\ell} [\delta^{(3)}(\mathbf{x} - \mathbf{y})J^k(\mathbf{x})] \right\} \end{aligned} \tag{14}$$

Since this current algebra is independent of whether ψ obeys commutation or anticommutation relations, the information as to particle statistics (Bose or Fermi) is not encoded in the Lie algebra itself but in the choice of its representation (up to unitary equivalence). Again interpreting ρ and J^k as operator-valued distributions, define $\rho(f) = \int_{\mathbb{R}^3} d^3x f(\mathbf{x})\rho(\mathbf{x})$ and $J(g) = \int_{\mathbb{R}^3} d^3x \sum_{k=1}^3 g^k(\mathbf{x})J^k(\mathbf{x})$, where f and the

components g^k of the vector field g belong to the function-space \mathcal{D} . Then the Lie algebra becomes

$$\begin{aligned} [\rho(f_1), \rho(f_2)] &= 0 \\ [\rho(f), J(g)] &= i\hbar\rho(g \cdot \nabla f) \\ [J(g_1), J(g_2)] &= -i\hbar J([g_1, g_2]) \end{aligned} \tag{15}$$

Equations [15] are a representation by self-adjoint operators of the semidirect sum of the abelian Lie algebra \mathcal{D} with $\text{vect}_0(\mathbf{R}^3)$. The corresponding group is the natural semidirect product of the space \mathcal{D} (regarded as an abelian topological group under addition) with $\text{Diff}_0(\mathbf{R}^3)$.

The construction generalizes to a general manifold M or manifold with boundary (in place of \mathbf{R}^3), and to a general set of charge densities that generate the local Lie algebra $\text{map}_0(M, \mathcal{G})$. When $M = S^1$, we have the Kac–Moody and Virasoro algebras with central charge zero. However, $L^{(0)}$ in the nonrelativistic $(1+1)$ -dimensional quantum theories is proportional to the total momentum P , and thus is unbounded above and below.

The continuous unitary representations of $\text{Diff}_0(M)$, or its semidirect product with a local current group at fixed time, thus describe nonrelativistic quantum systems (Albeverio *et al.* 1999, Goldin 2004). The unitary representation $V(\phi)$, $\phi \in \text{Diff}_0(M)$, satisfies $V(\phi_r^g) = \exp[i(r/\hbar)J(g)]$, where $r \in \mathbf{R}$ and ϕ_r^g is the one-parameter flow in $\text{Diff}_0(M)$ generated by the vector field g . Such a representation may be described very generally by means of a measure μ on a configuration space Δ , quasi-invariant under a group action of $\text{Diff}_0(M)$ on Δ , together with a unitary 1-cocycle χ on $\text{Diff}_0(M) \times \Delta$. The Hilbert space for the representation is $\mathcal{H} = L^2_{d\mu}(\Delta, \mathcal{W})$, which is the space of measurable functions $\Psi(\gamma)$, $\gamma \in \Delta$, taking values in an inner product space \mathcal{W} , and square integrable with respect to μ . The unitary representation V is given by

$$[V(\phi)\Psi](\gamma) = \chi_\phi(\gamma)\Psi(\phi\gamma)\sqrt{\frac{d\mu_\phi}{d\mu}}(\gamma) \tag{16}$$

where $\phi\gamma$ denotes the group action $\text{Diff}_0(M) \times \Delta \rightarrow \Delta$; μ_ϕ is the measure on Δ transformed by ϕ (which, by the quasi-invariance of μ , is absolutely continuous with respect to μ); $d\mu_\phi/d\mu$ is the Radon–Nikodym derivative; and $\chi_\phi(\gamma): \mathcal{W} \rightarrow \mathcal{W}$ is a system of unitary operators in \mathcal{W} obeying the cocycle equation

$$\chi_{\phi_1\phi_2}(\gamma) = \chi_{\phi_1}(\gamma)\chi_{\phi_2}(\phi_1\gamma) \tag{17}$$

Equations [16] and [17] hold outside sets of μ -measure zero in Δ . Given the quasi-invariant measure μ on Δ , one may always choose $\mathcal{W} = \mathcal{C}$

and $\chi_\phi(\gamma) \equiv 1$ to obtain a unitary group representation on complex-valued wave functions; but inequivalent cocycles describe unitarily inequivalent representations.

The configuration space $\Delta^{(N)}$, $N = 1, 2, 3, \dots$, consists of N -point subsets of \mathbf{R}^s , and $\mu^{(N)}$ is the (local) Lebesgue measure on $\Delta^{(N)}$. The corresponding diffeomorphism group and local current algebra representations describe N identical quantum particles in s -dimensional space. When $\chi \equiv 1$, we have bosonic exchange symmetry. Inequivalent cocycles on $\Delta^{(N)}$ are obtained (for $s \geq 2$) by inducing (generalizing Mackey’s method) from inequivalent unitary representations of the fundamental group $\pi_1[\Delta^{(N)}]$. For $s \geq 3$, this fundamental group is the symmetric group S_N of particle permutations; the odd representation of S_N , $N \geq 2$, gives fermionic exchange symmetry, while the higher-dimensional representations are associated with particles satisfying the parastatistics of Greenberg and Messiah.

When $s = 2$, however, $\pi_1[\Delta^{(N)}]$ is the braid group B_N . Goldin, Menikoff, and Sharp obtained induced representations of the current algebra describing the intermediate statistics proposed by Leinaas and Myrheim for identical particles in 2-space. Such excitations, subsequently termed “anyons” by Wilczek and characterized as charge-flux tube composites, are important constructs in the theory of surface phenomena such as the quantum Hall effect, and anyonic statistics has also been applied to the study of high- T_c superconductivity. Current algebra representations induced by higher-dimensional representations of B_N describe the statistics of “plektons.” Similarly, current algebra in nonsimply connected space describes the Aharonov–Bohm effect.

Let $\psi^*(h) = \int_{\mathbf{R}^s} d^s x h(x)\psi^*(x)$ denote the smeared creation field. Let the indexed set of representations ρ_N, J_N , $N = 0, 1, 2, \dots$, satisfying the current algebra [15], act in Hilbert spaces \mathcal{H}_N , where $\psi^*(h): \mathcal{H}_N \rightarrow \mathcal{H}_{N+1}$, $\psi(h): \mathcal{H}_{N+1} \rightarrow \mathcal{H}_N$, $\psi(h)|\mathcal{H}_0 = 0$, so that ψ^* and ψ intertwine the N -particle diffeomorphism group representations. Let $\rho(f)$ and $J(g)$ act on $\bigoplus_{N=0}^\infty \mathcal{H}_N$, so that $\rho(f)\Psi_N = \rho_N(f)\Psi_N$, $J(g)\Psi_N = J_N(g)\Psi_N$. Then conditions for a Fock space hierarchy are specified by commutator brackets of the fields with the currents:

$$\begin{aligned} [\rho(f), \psi^*(h)] &= \psi^*(\rho_{N=1}(f)h) \\ [J(g), \psi^*(h)] &= \psi^*(J_{N=1}(g)h) \end{aligned} \tag{18}$$

The local creation and annihilation fields for anyons in \mathbf{R}^2 , obeying [18], satisfy q -commutation relations, where q is the relative phase change associated with a single counterclockwise exchange of two anyons,

and the q -commutator $[A, B]_q = AB - qBA$. These relations generalize the canonical commutation ($q=1$) and anticommutation ($q=-1$) relations of quantum field theory.

When Δ is the configuration space of infinite but locally finite subsets of \mathbf{R}^s , nonrelativistic current algebra describes the physics of infinite gases in continuum classical or quantum statistical mechanics. Here, the most important kinds of measures μ are Poisson measures (associated with gases of noninteracting particles at fixed average density) or Gibbsian measures (associated with translation-invariant two-body interactions). These measures describe equilibrium states and correlation functions in the classical case, and specify the current algebra representations in the quantum theory.

The group of volume-preserving diffeomorphisms was taken by Arnold as the symmetry group of an ideal, classical, incompressible fluid, and Marsden and Weinstein described the hydrodynamics of such a fluid using the Lie–Poisson bracket associated with the nonrelativistic current algebra of divergenceless vector fields. The idea of using this algebra to study quantized fluid motion, included in the program proposed by Rasetti and Regge, formed the basis of the subsequent study of quantized vortex structures in superfluids from the point of view of geometric quantization on coadjoint orbits of the diffeomorphism group. This leads to quantum configuration spaces whose elements are no longer sets of points – for example, spaces of vortex filaments in \mathbf{R}^2 , or ribbons and tubes in \mathbf{R}^3 .

See also: Algebraic Approach to Quantum Field Theory; Electroweak Theory; Quantum Chromodynamics; Solitons and Kac–Moody Lie Algebras; Symmetries in Quantum Field Theory: Algebraic Aspects; Toda Lattices; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras.

Further Reading

- Adler SL and Dashen RF (1968) *Current Algebras and Applications to Particle Physics*. New York: Benjamin.
- Albeverio S, Kondratiev YuG, and Röckner M (1999) Diffeomorphism groups and current algebras: configuration space analysis in quantum theory. *Reviews in Mathematical Physics* 11: 1–23.
- Arnol'd VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. Applied Mathematical Sciences, vol. 125. Berlin: Springer.
- Gell-Mann M and Ne'eman Y (2000) *The Eightfold Way (1964) (reissued)*. Cambridge, MA: Perseus Publishing.
- Goddard P and Olive D (1986) Kac–Moody and Virasoro algebras in relation to quantum physics. *International Journal of Modern Physics A* 1: 303–414.
- Goldin GA (1996) Quantum vortex configurations. *Acta Physica Polonica B* 27: 2341–2355.
- Goldin GA (2004) Lectures on diffeomorphism groups in quantum physics. In: Govaerts J, Hounkonnou N, and Msezane AZ (eds.) *Contemporary Problems in Mathematical Physics: Proceedings of the Third International Workshop*, pp. 3–93. Singapore: World Scientific.
- Goldin GA and Sharp DH (1991) The diffeomorphism group approach to anyons. *International Journal of Modern Physics B* 5: 2625–2640.
- Ismagilov RS (1996) *Representations of Infinite-Dimensional Groups*. Translations of Mathematical Monographs, vol. 152. Providence, RI: American Mathematical Society.
- Kac V (1990) *Infinite Dimensional Lie Algebras*. Cambridge: Cambridge University Press.
- Marsden J and Weinstein A (1983) Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids. *Physica D* 7: 305–323.
- Mickelsson J (1989) *Current Algebras and Groups*. New York: Plenum.
- Ottesen JT (1995) *Infinite Dimensional Groups and Algebras in Quantum Physics*. Berlin: Springer.
- Pressley A and Segal G (1986) *Loop Groups*. Oxford: Oxford University Press.
- Renner B (1968) *Current Algebras and Their Applications*. Oxford: Pergamon.
- Sharp DH and Wightman AS (eds.) (1974) *Local Currents and Their Applications*. New York: Elsevier.
- Treiman SB, Jackiw R, Zumino B, and Witten E (1985) *Current Algebra and Anomalies*. Singapore: World Scientific.