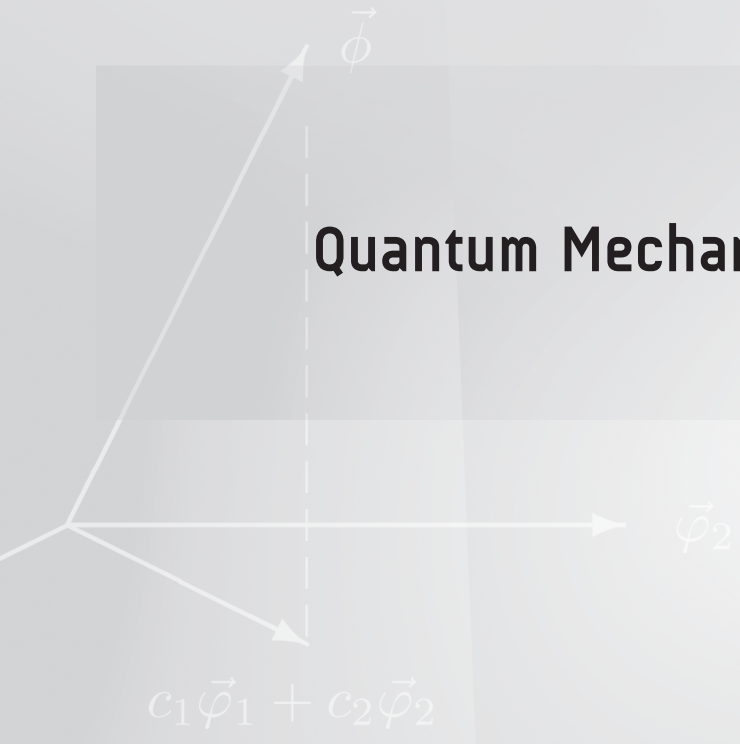
Quantum Mechanics

A Fundamental Approach

K. Kong Wan

Quantum Mechanics



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Quantum Mechanics
A Fundamental Approach

K. Kong Wan

JENNY STANFORD
PUBLISHING

Published by

Jenny Stanford Publishing Pte. Ltd.
Level 34, Centennial Tower
3 Temasek Avenue
Singapore 039190

Email: editorial@jennystanford.com
Web: www.jennystanford.com

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Quantum Mechanics: A Fundamental Approach

ISBN 978-981-4774-65-9 (Hardcover)
ISBN 978-0-203-70241-3 (eBook)

To my three beautiful grandchildren,
Freya, Lily and Leo.
Their joyful personalities are a source of
boundless love and happiness,
something which cannot be measured.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface

This book grows out of a module on the foundations of quantum mechanics given by the author to the final-year students on the degree of Master of Physics (Honours) in St Andrews and taken mainly by students in theoretical physics or in physics, in combination with or without mathematics. Already having attended previous modules on quantum mechanics, from introductory to advanced ones which contain many methods and applications, they have good knowledge of the subject. However, many of them still do not feel confident and comfortable with the subject. The mathematical formalism of quantum theory in terms of vectors and operators in infinite-dimensional complex vector spaces is very abstract. The definitions of many mathematical quantities used do not seem to have an intuitive meaning. This makes it difficult to appreciate and feel confident about the mathematical formalism, hampering the understanding of quantum mechanics.

My approach is to provide intuition for the mathematics involved by introducing the mathematics in its simplest and familiar form and then making the generalisation to obtain the abstract mathematics required for quantum theory. Feeling comfortable with the mathematics used would help the appreciation and understanding of the formalism and concepts of quantum mechanics.

This book is divided into four parts. Part I is a brief review of the general properties of classical and quantum systems. A general discussion of probability theory is also included which aims to help in understanding the probability theories relevant to quantum mechanics.

Part II is a detailed study of the mathematics for quantum mechanics. The study starts with a review of the familiar vectors and their operations in three dimensions. A geometric approach

is adopted, since three-dimensional vectors and operations on the vectors can be visualised geometrically and graphically. The aim is to provide intuition for vectors, operators, and their properties. An important example is the introduction of projection operators to describe the operation to project vectors onto any given direction. The fact that a projection operation in three dimensions can be visualised geometrically and graphically helps us to develop an understanding of projection operators in higher dimensions. Moreover, this will help the appreciation of other operators such as selfadjoint operators, which can be built up in terms of projection operators. Matrices are also discussed which can provide an explicit and numerical representation of abstract quantities.

The relevance of vectors and operators to quantum mechanics is due to the fact that certain operators together with vectors can be used to formulate probability theories—that is, vectors and operators can generate probability distributions. This can be demonstrated clearly with vectors and operators in three dimensions. In other words, vectors and operators in three dimensions can provide the framework to formulate a probability theory in a transparent and simple manner. Once we understand how probability theories can be formulated in terms of vectors and operators in three dimensions, we can extend the formalism to complex vectors in higher or lower dimensions.

Various quantities and their properties introduced in the three-dimensional vector space are re-introduced in finite-dimensional complex vectors spaces, and then again in infinite-dimensional Hilbert spaces. This apparent repetition serves as a learning process to help the students to gradually gain good understanding and hence feel confident in their applications. Readers familiar with some of mathematical preliminaries can skip this discussion. Some detailed mathematical discussion can also be omitted during a first reading.

The mathematics is presented with a certain degree of rigour to demonstrate the beauty quantum formalism. To avoid excessive technical discussions, many of the theorems and results are presented without proof. Many of the proofs are incorporated within questions in the exercises and problems section at the end of each chapter. Readers can check these questions and their solutions in

the Solutions Manual. References are provided for those who wish to pursue the matter further.

Part III presents quantum mechanics in a series of postulates. Six groups of postulates are presented to describe orthodox quantum systems. Each statement of a postulate is supplemented with a detailed discussion. To make them easier to understand, the postulates for discrete observables are presented before those for continuous observables.

Non-orthodox quantum systems, i.e., systems with superselection rules, and some conceptual issues such as quantum paradoxes and measurement are also discussed.

The last part of the book presents several illustrative applications. These include the harmonic oscillator, charged particles in external magnetic fields and the Aharonov–Bohm effect.

For easy reference, definitions, theorems, examples, comments, properties and results are labelled with section numbers. For example, Definition 6.4.2(2) is the second definition in section 6.4.2, E7.7.3(1) is the first example in section 7.7.3, and P9.2.2(3) is the third property in section 9.2.2. Exercises and problems given at the end of each chapter contain questions designed to help understand the matter presented in that chapter—for example, questions on proving some of the theorems and results. These questions are labelled by chapter number, e.g., Q12(1) is the first question in Exercises and Problems for Chapter 12.

Various symbols and notations are adopted to distinguish different quantities explicitly and to avoid misrepresentation. For example, a state is denoted by ϕ^s, a vector is denoted by $\vec{\phi}$, while a function defining the vector $\vec{\phi}$ is denoted by $\phi(x)$, and operators with the same operator expression, e.g., differential operators having the expression $-i\hbar d/dx$, but acting on different domains, are distinguished by subscripts and other symbols. There are extensive footnotes which serve to clarify things and show cross-references. There is a list of symbols under the Notation section preceding the bibliography and the index at the end of the book. Both Notation and Index are listed under a system of classification which aims to make it easier to find the references in the book. For example, *selfadjoint operators* and *spectral theorem* are listed under *operators on different vectors spaces*. If the definition of *selfadjoint operators*

in a Hilbert space is required, then one should look at *selfadjoint operators* under *operators in a Hilbert space*. The term *probability* is listed under different titles such as *probability (basics, discrete)* and *probability (quantum)*.

This book is self-contained both mathematically and physically. Hopefully, this would make it accessible and useful to a wider readership, including astronomers, mathematicians and philosophers of science who are interested in the foundations of quantum mechanics.

A Solutions Manual which presents the answers to all the questions given at the end of each chapter of this book is available and published separately by Jenny Stanford Publishing.

I am indebted to all the students, both undergraduates and postgraduates, who attended my Foundations of Quantum Mechanics module for their feedback on my lectures and lecture notes, which has led to numerous improvements, much of which are included in the present books.

K. Kong Wan
St Andrews

Section I

Classical and Quantum Systems

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 1

Structure of Physical Theories

Before going into any specific theory, it would be useful to have an idea of the general structure of physical theories. This would tell us how to proceed to formulate a theory generally. A physical theory aims to give a description of the properties and behaviour of a physical system, both qualitatively and quantitatively. Qualitatively a theory should explain the behaviour of a given physical system in terms of some fundamental properties of the system. Quantitatively a theory employs mathematical models to explain and quantify the properties and behaviour of a physical system. To achieve this, a theory requires abstraction and idealisation in order to be able to describe physical quantities in precise mathematical terms. First, we must find a mathematical framework with suitable mathematical constructs to describe various idealised physical quantities. We must be able to set up mathematical equations to describe the relationships between all the physical quantities. A good knowledge in mathematics is essential to a proper understanding and formulation of a physical theory. The study of the mathematical aspects of theoretical physics is often referred to as mathematical physics. Apart from the requirement of mathematical consistency, a theory is judged by how well its predictions agree

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

with empirical observations, especially by its ability to make new and experimentally verifiable predictions.

A physical system possesses many physical quantities which reflect the properties of the system. A physical quantity is a property of the system which can be physically observed and measured to yield numerical values. To emphasise their measurability, these quantities are called **observables**. A sufficiently large collection of these measured values at any given time characterises the system at that time. Such a characterisation defines a *state* of the system, i.e.,[1]

> *A characterisation of the system which embodies the values of a sufficiently large set of observables is called a* **state** *of the system.*

The relationship between states and observables can be complicated, depending on whether we are dealing with classical systems or quantum systems. While all observables of a classical system can be *simultaneously measured* the same is not true for a quantum system. It follows that for classical systems a state is associated with the values of all observables of the system while a state is *not* associated with the values of all observables of a quantum system.[2]

A physical theory should contain the following four basic components:

1. *Basic mathematical framework* This comprises:

(1) A set of elements endowed with some specific mathematical structure and properties. In mathematics such a set is generally known as a **space**.[3]

(2) Rules for manipulating the elements of the space.

2. *Description of states*

(1) States are described by elements of the space in the chosen mathematical framework. For this reason, the space is called the **state space** of the system.

---

[1] Bub p. 13. Isham pp. 56–59.
[2] Isham p. 58.
[3] Not to be confused with the *physical space* we live in.

(2) Every state corresponds to an element of the state space and different states would correspond to different elements of the state space. The converse is not necessarily true, i.e., the state space may have more elements needed to describe all the states of the system. For example, the states of a quantum system may not correspond one-to-one to all the elements of its state space. It is then necessary to explain clearly the relationship between the states and the elements of the state space.

3. *Description of observables*

(1) Observables are to be described by quantities defined on the state space. The description should yield all possible values of observables.

(2) The relationship between observables and states should be explicitly stated. The following two cases are of particular interest:

   (a) For a deterministic theory like classical mechanics, a state should determine the values of all observables.

   (b) For a probabilistic theory like quantum mechanics, a state should determine the probability distribution of the values of all observables.

4. *Description of time evolution (dynamics)*

We have adopted the common approach to establish the mathematical framework to describe the states of the system first. Observables are then introduced in terms of mathematical quantities defined on the set of states. It is possible to start with a mathematical framework to describe observables first with states defined through their relationship with observables. The *quantum logic* formulation of quantum mechanics is an example of such an approach.[4]

[4] Mackey pp. 56–81, Jauch pp. 67–110. Beltrametti and Gassinelli Chapter 10.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 2

Classical Systems

From our daily contact with the physical world, we gradually acquire a strong intuition as to how objects around us would behave. These are objects which we can examine directly through our senses and some simple instruments are known as *classical systems*. Typically a classical system consists of a number of particles. A huge number of these particles can group together to form a large object which is often idealised as being continuous. In other words, classical systems are idealised as either *discrete* or *continuous*. We shall discuss a number of guiding principles of the behaviour of classical systems in this chapter.

2.1 Discrete Systems

A system may consist of a single particle. More complex systems can be built up from a finite number of particles. The studies of such systems, known as *discrete systems*, lead to a large body of theories known as *classical mechanics*. Discrete systems possess a number of characteristic features which are set out in the following subsections.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

2.1.1 Intuition and Description

The properties of a discrete classical system are intuitively meaningful and their existence is taken for granted.[1] These properties can be given a mathematical description. Consider the motion of a particle in the three-dimensional *physical space* we live in. A particle of mass m is idealised as a point object. The particle's position is determined by its *position vector* $\vec{x}$ which is a three-component quantity with one component along each of the three coordinate axes in a chosen *Cartesian coordinate system*.[2] These components are denoted by x, y, z or more conveniently by x_1, x_2, x_3. Symbolically we can express $\vec{x}$ by a number of different expressions, e.g.,

$$\vec{x} = (x, y, z) = (x_1, x_2, x_3) = \{x_j,\ j = 1, 2, 3\}. \tag{2.1}$$

Observables are physical properties which can be quantified by numerical values obtainable to arbitrary accuracy by physical measurement. For example, consider the concept of instantaneous velocity $\vec{v}$. We should be able to give it a precise mathematical definition, i.e., $\vec{v} := d\vec{x}/dt$, and we should be able to measure its value accurately.[3] The same goes with the notion of energy, e.g., the kinetic energy is defined by $K := \frac{1}{2} m\vec{v}^{\,2}$. Observables are simultaneously measurable to arbitrary accuracy and there is no general restrictions of the values of observables, e.g., there is no restriction of the energy of a system to a discrete set of values. It follows that at any instant of time all the observables of a given system would have a definite value and these values serve to characterise the system. Such a characterisation of the system by the values of all its observables constitutes a *state* of the system. In other words, a state determines the values of all observables and different states mean different sets of observable values.

---

[1] *See* Isham pp. 53–56 for a brief conceptual discussion.

[2] Wan p. 14. The system is also known as a rectangular or rectangular Cartesian coordinate system. Cartesian is a term to indicate something relating to the French philosopher and mathematician Descartes (1596–1650). The position vector will be formally introduced in §6.2.2. The coordinate variables are called Cartesian coordinates.

[3] The **notation** := is generally used to signify a definition, i.e., $\vec{v}$ is defined by $d\vec{x}/dt$.

2.1.2 Determinism

There are several aspects of determinism:

(1) For many systems the values of all observables can be determined by the values of a small number of observables. This means that a state can be determined by the values of a small number of observables. For example, the state of a single particle can be determined by the values of its position and velocity. The resulting state then determines the values of all other observables, e.g., the values of its potential and kinetic energy.
(2) The dynamics must be such that the future behaviour of the system is uniquely determined by the initial behaviour of the system. Since the behaviour of a system is determined by the values of its observables this means that
    (a) The state at a later time is uniquely determined by the state at the initial time.
    (b) The future values of all observables are uniquely determined by their initial values.
(3) Classical statistical mechanics which is based on probability theory does not contradict the above determinism. In classical physics the need for probability arises due to our ignorance of all the initial conditions. Probabilistic behaviour appears if we have insufficient initial information to determine the behaviour of the system. This happens when dealing with a system composed of a large number of particles; it is practically impossible to ascertain all initial conditions.

2.1.3 Unambiguous Objective Reality

There is an underling objective reality and our task is to describe and to reveal such reality. For example, a particle moving along would have an objective kinetic energy value and this value can be found by measurement. A measurement only serves to reveal the value already possessed by the system. In other words, a measurement produces a value which the system possesses before the measurement. The disturbance of any measurement on a system

can be made arbitrarily small so that the measured value would also be the value after the measurement. We can also perform any number of simultaneous measurements of different observables without significantly disturbing the system.

2.1.4 Structure of Classical Mechanics

For simplicity we shall examine a classical particle of mass m in three-dimensional motion in Newtonian theory.[4] The basic theory consists of the following components:

1. *Basic mathematical framework* Newtonian mechanics is based on the mathematics of vectors and vector calculus in three dimensions.

2. *Description of states*

(1) A state at any given time can be described by two three-dimensional vectors, one for the particle's position $\vec{x}$ and one for the particle's velocity $\vec{v}$.

(2) Alternatively we can form a six-dimensional space with each element of the space specified by six components, first three for position and the remaining three for velocity. A state at any given time can then be described by a single element in this six-dimensional space which constitutes the **state space** of the system, i.e., a single element of the state space specifies a state of the system and vice versa.[5]

3. *Description of observables* We can divide classical observables into kinematic ones and dynamical ones.

(1) *Kinematic observables* Kinematic observables are defined by real-valued functions of the state and they are not explicitly dependent on external factors and interaction. The relations between states and observables are explicitly given by the

[4] Newton (1642–1727) was a British physicist and mathematician.

[5] Replacing velocity by momentum leads to the concept of *phase space*. A phase space has a complicated mathematical structure, known as a *Hamiltonian* or *symplectic* structure, although it is common for a phase space to be considered as a vector space (*see* Wan pp. 63–64. Isham p. 60).

functions which define the observables. The values of these observables are the values of the corresponding functions. Once a state is given, the values of all kinematic observables are fixed. The position $\vec{x}$ and velocity $\vec{v}$ are kinematic observables. The *linear momentum* $\vec{p}$ defined by

$$\vec{p} := m\vec{v} \tag{2.2}$$

is another example. It is also referred to as the *kinematic momentum*, when we want to emphasise the kinematic nature of the quantity.[6] The kinetic energy K is a kinematic observable. Another important example is the angular momentum $\vec{L}$ defined by the vector product of $\vec{x}$ and $\vec{p}$, i.e.,

$$\vec{L} := \vec{x} \times \vec{p}. \tag{2.3}$$

To emphasise the kinematic nature of this quantity we also call it the *kinematic angular momentum*. The components of $\vec{L}$ along the x, y and z directions are

$$L_x = yp_z - zp_y, \tag{2.4}$$

$$L_y = zp_x - xp_z, \tag{2.5}$$

$$L_z = xp_y - yp_x. \tag{2.6}$$

These components can be conveniently worked out in terms of the following determinant:[7]

$$\vec{L} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix}. \tag{2.7}$$

(2) *Dynamical observables* These are observables involving external interactions, e.g., they are dependent on external factors such as external forces. An obvious example is the total energy of the particle which contains a potential energy term. External factors such as potential energy are generally dependent on the position of the particle. They can also depend on the particle's velocity.[8] As an example consider a charged

[6] Feynman, Leighton and Sands p. 21–5.

[7] *See* Eq. (7.39) for determinants of matrices.

[8] An example is the magnetic potential energy of a charged particle in an external magnetic field as shown in Eq. (37.35) in §37.3.1 in Chapter 37.

particle in an external magnetic field. In addition to its linear momentum the particle has another momentum, known as its *canonical momentum*, which contains the external magnetic vector potential. Details are discussed in §27.1.1 and §27.1.2.[9] This is a dynamical observable, and hence it is also known as the *dynamical momentum*.[10] The particle also has a *canonical angular momentum*.[11] Once the external factors are known the values of these observables are determined by the state.

Classical observables are easily defined, but it is a real challenge to relate classical observables to quantum ones. As will be discussed in detail in Chapters 27 and 37, the distinction between kinematic momentum and canonical momentum is of crucial importance in the quantum theory of a charged particle in a magnetic field, since it is the canonical momentum, not the linear momentum, which is directly quantised.

4. *Dynamics*

(1) *Evolution of the state* This is given by Newton's laws.

(2) *Evolution of observables* There is no need to set up a separate equation of motion for observables since the time dependence of the state automatically leads to the time dependence of observables which are just functions of the state.[12]

2.2 Continuous Systems

Continuous Systems are idealised systems. An example is that of a continuous string lying along the x-axis and vibrating along the y-axis. The state of vibration requires a set of values for its characterisation, i.e., its vibration along the y-axis has to be described by a function $Y(x, t)$ with a set of values, one for each x and t. This function is called a *wave*. A sinusoidal wave propagating

[9] *See* Eq. (27.44) for an expression of this canonical momentum.
[10] Feynman, Leighton and Sands pp. 21–5.
[11] *See* Eq. (27.49) for an expression of this canonical angular momentum.
[12] We can establish an explicit equation of motion of an arbitrary observable if we want to. This will be discussed in detail in §27.1.

to the right along the x-axis is given by a function of the form

$$Y_+(x, t) = A \sin(kx - \omega t), \quad k, \ \omega > 0, \tag{2.8}$$

where A is the amplitude, k is known as the angular wave number and ω is the angular frequency. These quantities are related to the wavelength λ, the frequency f and the period T by

$$\lambda = 2\pi/k, \quad f = \omega/2\pi \quad \text{and} \quad T = 2\pi/\omega. \tag{2.9}$$

The above sinusoidal wave is a solution of the following equation:

$$\frac{\partial Y^2}{\partial x^2} = \frac{1}{v^2}\frac{\partial Y^2}{\partial t^2} \quad \text{with} \quad v = \frac{\omega}{k}. \tag{2.10}$$

This is known as the *classical wave equation* in one-dimension.

To be more specific consider a uniform elastic string of length L and mass per unit length ρ lying along the x-axis with both ends fixed, one end at $x = 0$ and the other at $x = L$. The string is subject to a tension of magnitude τ. When the string is set to vibrate transversely along the y-axis its configuration has to be given by a function $Y = Y(x, t)$ in order to describe the displacement of the string along the y-axis for all $x \in (0, L)$.[13] The vibration of the string can be shown to obey the above classical wave equation with the constant v related to the tension and mass density by $v = \sqrt{\tau/\rho}$.[14] Solutions of the wave equation which describe a sinusoidal vibration of the string are

$$Y_n(x, t) = \sqrt{2/L} \sin k_n x \cos \omega_n t, \tag{2.11}$$

where[15]

$$k_n = n\pi/L, \quad \omega_n = k_n\sqrt{\tau/\rho}, \quad n = 1, 2, 3, \cdots. \tag{2.12}$$

These vibrations are known as *normal modes*. The spatial part of the function

$$Y_n(x, 0) = \sqrt{2/L} \sin k_n x \tag{2.13}$$

[13] The symbol $\in$ means "belonging to" or "is in", e.g., $x \in (0, L)$ means x is in the open interval $(0, L)$. In contrast $\notin$ means "not belonging to". Closed intervals are denoted by $[0, L]$.

[14] Crawford Chapter 2.

[15] The amplitude $\sqrt{2/L}$ is chosen to satisfy Eq. (2.14) on orthonormality.

is referred to as an *eigenfunction* of the system. These eigenfunctions are *orthonormal* in the sense that

$$\int_0^L y_n(x) y_m(x)\, dx = \delta_{nm}, \tag{2.14}$$

where δ_{nm} is the *Kronecker delta* which is equal to 1 when $n = m$ and zero otherwise.[16] We also refer to $Y_n(x, t)$ as an eigenfunction. The string can execute more complicated vibrations which can be described by non-sinusoidal functions $Y(x, t)$. The mathematics of Fourier series tells us that a general solution $Y(x, t)$ of the wave equation can be obtained by an appropriate *superposition* of sinusoidal functions in Eq. (2.11), i.e., we have[17]

$$Y(x, t) = \sum_{n=1}^{\infty} A_n\, Y_n(x, t), \tag{2.15}$$

where A_n are constants.

A solution of the wave equation describes a *state* of the system. Normal modes are basic states of a vibrating string with fixed ends. We shall call the states described by the eigenfunctions as *eigenstates*. An arbitrary solution corresponds to an arbitrary state.

Many of the guiding principles in the formulation of classical mechanics still apply to classical continuous systems. However, continuous systems do have some distinctive properties not possessed by discrete systems discussed in §2.1:

P2.2(1) Observables are related to the state in a complicated manner. In an eigenstate $Y_n(x, t)$ the vibration has a definite wavelength $\lambda_n = 2\pi/k_n$, and the wavelengths corresponding to all the eigenstates form a discrete set of values. In an arbitrary state $Y(x, t)$ shown in Eq. (2.15), the vibration does not have a wavelength. In other words, some basic observables of the system do not have a definite value in an arbitrary state.

P2.2(2) While the system is in an arbitrary state $Y(x, t)$, we can perform a wavelength measurement by physically forcing the vibration into a normal mode, e.g., $Y_n(x, t)$, from which we can observe a definite wavelength, e.g., λ_n. But this is not the wavelength

[16] Kronecker (1823–1891) was a German mathematician.
[17] Crawford p. 60. Fourier (1768–1830) was a French mathematician.

of the vibration before the measurement. Indeed there is no wavelength in state $Y(x, t)$. Moreover such a measurement changes the state of the system from an arbitrary state to an eigenstate.

P2.2(3) There may exist a set of basic states from which an arbitrary state can be obtained by superposition, as shown in Eq. (2.15). These basic states or eigenstates whose superposition can give rise to an arbitrary one are said to constitute a *complete set of states*. The superposition of eigenstates gives rise to the phenomenon of *interference*.

These properties can help us to appreciate the behaviour of quantum systems later.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 3

Probability Theory for Discrete Variables

The behaviour of many classical systems cannot be predicted with certainty, despite the best efforts of the experimenter, e.g., the tossing of a coin. In the context of deterministic behaviour of classical systems, the indeterminate nature arises from insufficient initial knowledge of the particular system in question, a fact already pointed out in §2.1.2. Many of these processes can be described by a probability theory. A summary of probability theory for discrete variables is set out in this chapter.

3.1 Physical Concept of Probability

Generally an experiment can yield many outcomes. The nature of an experiment may be such that we cannot predict exactly which outcome will occur, i.e., under identical conditions a repetition of the experiment may yield a different outcome. Fortunately many of these experiments exhibit certain regularity so that the results of many independent repetitions of the experiment under identical conditions are describable in terms of a probability theory.[1] We

[1] Penrose Chapter I §4.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

call such experiments *statistical experiments*. A typical situation may involve a physical system with an observable A which possesses a discrete set of numerical values a_ℓ, $\ell = 1, 2, 3, \ldots, n$ such that

(1) each individual experiment to find the value of A yields a value a_ℓ in an unpredictable fashion, and

(2) if the experiment is carried out N times under identical conditions and if the value a_ℓ is obtained N_ℓ times, then the ratio (N_ℓ/N) appears to converge as N becomes larger and larger.

The regularity here is the convergence of the ratio (N_ℓ/N) as N becomes larger and larger. This enables us to introduce the quantity

$$\wp_N^A(a_\ell) := \frac{N_\ell}{N}, \tag{3.1}$$

for N large enough for the convergence to become apparent. This quantity is then identified with the *probability* that a measured value of A is a_ℓ. It follows that such an assignment of probabilities satisfy the following properties:

$$\wp_N^A(a_\ell) \in [0, 1] \quad \text{and} \quad \sum_{\ell=1}^{n} \wp_N^A(a_\ell) = 1. \tag{3.2}$$

The average value $\mathcal{E}(\wp_N^A)$ over N experiments is defined by

$$\mathcal{E}(\wp_N^A) := \frac{1}{N} \sum_{\ell=1}^{n} a_\ell N_\ell = \sum_{\ell=1}^{n} a_\ell \wp_N^A(a_\ell). \tag{3.3}$$

This empirical understanding of probability is known as the *frequency interpretation* of probability.[2] Ideally we would like N to tend to infinity. Of course in any practical experiment we can only achieve a large but finite N. A formulation of probability theory based on this frequency interpretation is considered by many to be unsatisfactory.[3] A mathematical formulation should be axiomatic

[2] Penrose Chapter I §4. There are many other interpretations of the notion of probability. An example would be the statement saying there is a 50% probability of raining tomorrow. Clearly we cannot have a large set of replicas of "tomorrows" to test the statement in the sense of a frequency interpretation. Such a statement is really a reflection of a subjective belief based on experience.

[3] Penrose Chapters I, II. It may be argued that a probability theory based on a frequency interpretation is not logically satisfactory since the definition of a probability cannot even be made precise on account of our failure to achieve the limiting value of $\wp_\infty^A(a_\ell)$ in any practical experiment. There are other problems also. We can have $N_1 > 0$ remaining finite as $N \to \infty$. This means that $\wp_\infty^A(a_1) = 0$ yet a measurement can still yield the value a_1 since $N_1 \neq 0$.

and independent of any particular interpretation. This we shall do in the remainder of this chapter. We shall start by reviewing some basic mathematics of set theory.

3.2 Sets, Mappings and Functions

3.2.1 Set Operations

A *set* D is just a collection of objects d, known as *elements* or *members*, which can be abstract or numerical quantities. A set is said to be *discrete* or *countable* if we can assign a positive integer ℓ to every element of the set with different elements assigned to different integers. It follows that an element can be denoted by d_ℓ and the set can be written as

$$D = \{d_\ell, \ell = 1, 2, 3, \ldots, n\}. \tag{3.4}$$

When n is infinitely large the set is said to be *countably infinite*. A set with an infinite number of elements may not be countable. For example, the set of real numbers in the interval [0, 1] has an infinite number of elements and these elements are not countable. A *subset* S of D is a set containing some elements of D, i.e., every element of S is an element of D while the converse is not necessarily true, e.g., $S = \{d_1, d_3\}$ is a subset of D. The relationship between S and D is signified by the notation

$$S \subset D \quad \text{or} \quad D \supset S. \tag{3.5}$$

A more explicit notation S_D can be used when we wish to indicate explicitly that S is a subset of D. We may have S equal to D.[4] If we wish to emphasise that S is not equal to D we call S a *proper subset* of D. We can also have subsets containing a single element, e.g., $S = \{d_2\}$. A set containing only one element is called a *singleton set*, e.g., $S = \{d_2\}$ is a singleton set. If $S \subset D$, the *complement* of S with respect to D is the subset S^c containing all the elements of D which are not in S, i.e.[5]

$$S^c := \{d : d \in D \text{ and } d \notin S\}. \tag{3.6}$$

[4] For this reason the notation $S \subseteq D$ is often used by some authors.
[5] The absence of subscript ℓ indicates that D may be continuous.

A useful concept is that of the *empty set*. This is a set containing no element at all, and it will be denoted by the symbol $\emptyset$. Two sets D_1 and D_2 are said to be *disjoint* if they have no element in common.

We have the following familiar operations:[6]

(1) Let D_1 and D_2 be two arbitrary sets. Then:

(a) Their *difference*, denoted by $D_1 - D_2$, is the set composed of all the elements of D_1 which are not in D_2, i.e.,

$$D_1 - D_2 := \{d : d \in D_1 \text{ and } d \notin D_2\}. \tag{3.7}$$

It follows that the complement S^c of a subset of a set D is equal to the difference $D - S$, i.e.,

$$S^c = D - S. \tag{3.8}$$

If D_1 and D_2 are disjoint then $D_1 - D_2$ is equal to D_1

(b) Their *intersection*, denoted by $D_1 \cap D_2$, is the set containing elements which are in both D_1 and D_2. It follows that D_1 and D_2 are *disjoint* if

$$D_1 \cap D_2 = \emptyset. \tag{3.9}$$

(c) Their *union*, denoted by $D_1 \cup D_2$, is the set containing both the elements of D_1 and D_2.

(2) Let $\{D_\ell, \ell = 1, 2, \cdots\}$ be a sequence of sets. The sequence may be finite or infinite. Then:

(a) Their union, denoted by $\cup_\ell D_\ell$ and defined by

$$\cup_\ell D_\ell := \{d \in D_\ell \quad \text{for at least one } \ell\}, \tag{3.10}$$

is called a *countable union*.

(b) Their intersection, denoted by $\cap_\ell D_\ell$ and defined by

$$\cap_\ell D_\ell := \{d \in D_\ell \quad \text{for all } \ell\}, \tag{3.11}$$

is called a *countable intersection*.

(3) Often we need to consider ordered pairs of elements from two sets. This leads to the notion of the *Cartesian product* of two sets, i.e., the Cartesian product of two sets D and G, denoted by $D \times G$, is the set of all ordered pairs (d, g):

$$D \times G := \{(d, g) : d \in D, g \in G\}. \tag{3.12}$$

[6] Lipschutz p. 104. Roman Vol. 1 pp. 6–12. Kingman and Taylor pp. 9–11.

The operations of complement, union and intersection are related by the following De Morgan's theorem.[7]

Theorem 3.2.1(1) *Let S_ℓ, $\ell = 1, 2, \cdots$ be a sequence of subsets of a set D. Then the complement of the union $\cup_\ell S_\ell$ of S_ℓ is equal to the intersection $\cap_\ell S_\ell^c$ of their complements S_ℓ^c, and the complement of the intersection $\cap_\ell S_\ell$ of S_ℓ is equal to the union $\cup_\ell S_\ell^c$ of their complements i.e.,*

$$\left(\cup_\ell S_\ell\right)^c = \cap_\ell S_\ell^c, \quad \left(\cap_\ell S_\ell\right)^c = \cup_\ell S_\ell^c. \tag{3.13}$$

3.2.2 Mappings and Functions

A *mapping* f is a relation which associates each element d of a set D a unique element g of another set G. This is denoted by

$$f : D \to G \quad \text{with} \quad f(d) = g. \tag{3.14}$$

We call f a mapping from D into G with $g = f(d)$ referred to as the *image* of d under f. A mapping is also called a *function*, i.e., the above mapping is also referred to as a *function from D into G*, or a *function on D into G*. In particular, we have mappings into a set of real numbers. The set of all real numbers τ ranges from $-\infty$ to ∞ is denoted by the symbol $\mathbb{R}$, i.e., $\mathbb{R} := \{\tau : \tau \in (-\infty, \infty)\}$. We can represent the set $\mathbb{R}$ of real numbers geometrically in terms of a horizontal line, called the *real line*, in that each real number corresponds to a point in the real line. The half real line $\mathbb{R}^+$ consists of the set of numbers $[0, \infty)$, i.e., $\mathbb{R}^+ := [0, \infty)$, and the other half real line $\mathbb{R}^-$ consists of the set of numbers $(-\infty, 0]$.

It is a common practice to denote a function from D into G shown in Eq. (3.14) by the notation $f(d)$, although strictly speaking a function from D into G should be denoted by a single symbol, e.g., f, and the notation $f(d)$ should denote the image of d under the function. When the notation $f(d)$ is used to denote a function the symbol d should be regarded as a variable, not a specific member of D.[8] There are different kinds of functions:

(1) A mapping from D into the set $\mathbb{R}$ of real numbers is called a *real-valued function* on D.

[7] Roman Vol. 1 pp. 9–10. De Morgan (1806–1871) was a British mathematician.
[8] For example, a function f on $\mathbb{R}$ is often written as $f(x)$ and called a function of x.

(2) A mapping from D into the set $\mathbb{C}$ of complex numbers is called a *complex-valued function* on D.

(3) Let $\{S\}$ denote a family of subsets S of D. Then a mapping from $\{S\}$ into G is called a *set function*. An example would be a mapping f from $\{S\}$ into $\mathbb{R}$, i.e.,

$$S \to f(S) \in \mathbb{R}. \tag{3.15}$$

This is known as a *real-valued set function*.

(4) If G consists of operators a mapping from D into G is called an *operator-valued function*. In §15.1 we shall present a study of *projector-valued functions*. These are mappings into G which is composed of a special class of operators known as projectors.

There are a few terms frequently used in relation to mappings:

1. *Domain, codomain and range* For a mapping shown in Eq. (3.14), the set D is called the *domain* and the set G is called the *codomain* of the mapping. The collection R of all the images, i.e.,

$$R := \{\, g : g = f(d),\ d \in D \,\}, \tag{3.16}$$

is called the *range* of the mapping. Generally f maps D into a subset of G, i.e., $R \subset G$.

For a set function which maps a family $\{S\}$ of subsets S into G, the domain of the set function is the family $\{S\}$ of subsets.

2. *Into and onto mappings* Generally the range R is a subset of the codomain G and the mapping is said to be a mapping from D *into* G. If the range R coincides with G then the mapping is said to be a mapping from D *onto* G. For an onto mapping, we can rewrite G as R, i.e., we have a mapping from D onto R.

3. *One-to-one mappings* If different elements of D are associated by f with different elements of G, i.e.,

$$d_1 \neq d_2 \quad \Rightarrow \quad g_1 = f(d_1) \neq g_2 = f(d_2), \tag{3.17}$$

then f is said to be a *one-to-one* mapping. For a one-to-one mapping f from D onto R every element r of R has a unique element d of D associated with it by the one-to-one mapping.

4. *One-to-one correspondence* A one-to-one mapping f of D onto R is said to define a *one-to-one correspondence* between the sets D and

R. We shall denote a one-to-one correspondence by a "two-sided" arrow, i.e.,

$$f : D \leftrightarrow R. \tag{3.18}$$

5. *Identity mapping* The mapping $I\!I$ from D onto D, i.e., onto itself, defined by $I\!I(d) = d \ \forall d \in D$ is called the *identity mapping on* D.[9]

6. *Image and inverse image* The *image* of a subset S_D of the domain D under a function f, denoted by $f(S_D)$, is the set in the range given by

$$f(S_D) := \{r \in R : r = f(d), \ d \in S_D\}. \tag{3.19}$$

The *inverse image* of a subset S_R of the range R under f, denoted by $f^{-1}(S_R)$, is the set in the domain D which is mapped onto S_R by f, i.e.,

$$f^{-1}(S_R) := \{d \in D : f(d) = r \in S_R\}. \tag{3.20}$$

For a mapping from D into G, the range is generally a proper subset of G, i.e., there are subsets of G which are outside the range of the mapping. Then no elements in D are mapped to any element of such subsets so that the inverse image of such a subset of G does not exist. We shall signify this by saying that the inverse image of such a subset is the empty set $\emptyset$. With this understanding we can say that the inverse image of any subset of G exists.

7. *Inverse* Let f be a one-to-one mapping from D onto R. Since every element r of R has a unique element d of D associated with it by the one-to-one correspondence $r = f(d)$ we can define a one-to-one mapping f^{-1} from R onto D by this association, i.e., we have

$$f^{-1} : R \to D \quad \text{defined by} \quad f^{-1}(r) = d. \tag{3.21}$$

We call this the *inverse* of f.

A function may not have an inverse. The existence of inverse images does not imply the existence of an inverse function.

[9] The symbol $\forall$ means "for every" or "for all".

3.3 Discrete Sample Spaces and Events

Consider a statistical experiment which has a discrete set

$$S_{\text{am}} = \{s_1, s_2, \cdots\} \tag{3.22}$$

of possible outcomes. Let us introduce the following terms:

1. *Sample space* The set S_{am} of outcomes is called the *sample space* of the experiment.

2. *Events* A subset E of S_{am}, i.e., a set of outcomes, is called an *event*. For the statistical experiment on tossing of a die, the sample space consists of six outcomes, i.e., six numbers:

$$S_{\text{am}} = \{1, 2, 3, 4, 5, 6\}. \tag{3.23}$$

The event E_{e} that an even number occurs consists of three outcomes, i.e., $E_{\text{e}} = \{2, 4, 6\}$. The empty set and the sample space are formally regarded as events. The empty set which contains no outcomes is known as the *impossible event* and the sample space S_{am} is known as the *sure event*.[10]

3. *Elementary events* A subset of S_{am} containing a single outcome, e.g., $\{s_2\}$, is called an *elementary event*. If a run of the experiment yields the outcome s_2 we say either the outcome s_2 occurs or the elementary event $E_2 = \{s_2\}$ occurs, and any event containing the outcome s_2 is also said to occur.

4. *Disjoint events* Two events are said to be *disjoint* or *mutually exclusive* if they have no outcome in common so that they cannot both occur in a single run of the experiment.

5. *Intersection of events* The set $E_1 \cap E_2$ is the event that occurs if both E_1 *and* E_2 occurs. Two events E_1 and E_2 are disjoint or mutually exclusive if $E_1 \cap E_2 = \emptyset$. As an example we can see that an event E is disjoint from the event $E^c = S_{\text{am}} - E$.[11]

6. *Union of events* The set $E_1 \cup E_2$ is the event that occurs if either E_1 *or* E_2 occurs.

[10] We assume that every run of the experiment would yield an outcome $s_\ell \in S_{\text{am}}$. It is then impossible to have $\emptyset$ as an outcome.

[11] The set E^c is the complement of E.

It is useful to visualise these operations in terms of *Venn diagrams* used in set theory, e.g., we can picture the sample space as a big rectangle and various events as smaller rectangles within the big rectangle.[12]

The probabilistic nature of a statistical experiment with a discrete sample space can be described by assigning a probability to each outcome. Since every outcome lies in some event an assignment of probability to all events will also characterise the probabilistic nature of the experiment.

3.4 Probability Mass Functions and Measures

Consider a statistical experiment with a discrete sample space $S_{\text{am}} = \{s_1, s_2, \cdots, s_n\}$, where n may be finite or infinite. Let $\{E\}$ denote the set of all events E, including the sure event S_{am} and the impossible event $\emptyset$. There are two conventional ways to introduce probabilities on S_{am}. One way is to assign a probability to each outcome s_ℓ in terms of a function from S_{am} into the interval $[0, 1]$ of real numbers.

Definition 3.4(1) *A* **probability mass function** $\wp$ *is a function from the sample space* $S_{am} = \{s_1, s_2, \cdots, s_n\}$ *of a statistical experiment into the interval* $[0, 1]$ *satisfying the conditions on probabilities in* Eq. (3.2), i.e.,

$$\wp(s_\ell) \in [0, 1], \quad \sum_{\ell=1}^{n} \wp(s_\ell) = 1. \tag{3.24}$$

The value $\wp(s_\ell)$ is interpreted as the probability of occurrence of outcome s_ℓ. The sum in Eq. (3.24) is the total probability which is equal to 1.[13]

Alternatively we can assign a probability to each event E in terms of a set function from the collection $\{E\}$ of events into the interval $[0, 1]$. Such a set function is known as a *probability measure*.

Definition 3.4(2) *A set function* $\mathcal{M}_p$ *on the collection* $\{E\}$ *of all the events* E *of a statistical experiment is called a* **probability measure**

[12] Lipschutz p. 5. Venn (1834–1923) was a British logician and philosopher.

[13] We will take the *frequency interpretation* of probability with the assumption that $\wp(s_\ell) = 0$ means that s_ℓ never occur and $\wp(s_\ell) = 1$ mean that s_ℓ is certain to occur.

on the set $\{E\}$ *of events if the following properties hold:*[14]

PM3.4(1) *Non-negativity* $\mathcal{M}_{\mathrm{p}}(E) \in [0, 1]$.
PM3.4(2) *Normalisation* $\mathcal{M}_{\mathrm{p}}(S_{\mathrm{am}}) = 1$.
PM3.4(3) *Additivity* $\mathcal{M}_{\mathrm{p}}(\cup_\ell E_\ell) = \sum_\ell \mathcal{M}_{\mathrm{p}}(E_\ell)$, where $E_\ell \cap E_{\ell'} = \emptyset, \quad \ell \neq \ell'$.

The above properties can be interpreted and understood as follows:

PM3.4(1) The value $\mathcal{M}_{\mathrm{p}}(E)$ is interpreted as the probability of occurrence of event E. The probability of E *not* occurring is given by $\mathcal{M}_{\mathrm{p}}(E^c)$.

PM3.4(2) Each run of the experiment must produce an outcome so that the sure event S_{am} which contains all outcomes occurs with probability 1.

PM3.4(3) The probabilities of exclusive events add up to produce the probability for the union of all these exclusive events. This property, known as *countable additivity*, requires the probability of the union of a countable collection of mutually exclusive events to be equal to the sum of the probabilities of all the individual events. This property applies both to finite and infinite unions of mutually exclusive events.[15]

Some useful properties of probability measures are summarised in the following theorem.[16]

Theorem 3.4(1) *A probability measure* $\mathcal{M}_{\mathrm{p}}$ *on a discrete sample space* S_{am} *with a corresponding set of events* E *satisfies:*

$$\mathcal{M}_{\mathrm{p}}(\emptyset) = 0. \tag{3.25}$$

$$\mathcal{M}_{\mathrm{p}}(E^c) = 1 - \mathcal{M}_{\mathrm{p}}(E). \tag{3.26}$$

$$E_1 \subset E_2 \Rightarrow \mathcal{M}_{\mathrm{p}}(E_1) \leq \mathcal{M}_{\mathrm{p}}(E_2). \tag{3.27}$$

$$\mathcal{M}_{\mathrm{p}}(E_1 \cap E_2) = \mathcal{M}_{\mathrm{p}}(E_1) - \mathcal{M}_{\mathrm{p}}(E_1 - E_2). \tag{3.28}$$

$$\mathcal{M}_{\mathrm{p}}(E_1 \cup E_2) = \mathcal{M}_{\mathrm{p}}(E_1) + \mathcal{M}_{\mathrm{p}}(E_2) - \mathcal{M}_{\mathrm{p}}(E_1 \cap E_2). \tag{3.29}$$

[14] Lipschutz pp. 40–41.
[15] Countably infinite additivity is required only for infinite sample spaces.
[16] Lipschutz pp. 40–41. Using Venn diagrams for the various operations of events one can visualise the theorem.

Probability mass functions and probability measures are related in a one-to-one manner:

(1) A probability measure $\mathcal{M}_{\mathrm{p}}$ generates a probability mass function $\wp$ by

$$\wp(s_\ell) = \mathcal{M}_{\mathrm{p}}(\{s_\ell\}) \quad \forall s_\ell \in S_{\mathrm{am}}. \tag{3.30}$$

In other words, the probability of occurrence of an outcome s_ℓ is the same as the probability of occurrence of the corresponding elementary event $\{s_\ell\}$.

(2) A probability mass function $\wp$ gives rise to a probability measure $\mathcal{M}_{\mathrm{p}}$ by

$$\mathcal{M}_{\mathrm{p}}(E) = \sum_{\ell} \wp(s_\ell) \quad \forall\, s_\ell \in E. \tag{3.31}$$

The set of probabilities for all the outcomes and events in a sample space is called a **probability distribution** of the sample space. A probability distribution can be characterised by either a probability mass function or a probability measure. Probability measures become indispensable for the description of probability distributions of continuous sample spaces.

3.5 Expectation Values, Variances and Uncertainties

Consider a statistical experiment to measure an observable A which has a set of values $a_1, a_2, a_3, \cdots, a_n$. These values define the sample space. Let the probability distribution of the sample space be given by a probability mass function $\wp^A$. To quantify the average (mean) measured value and the deviation from the mean of individual values we can introduce the following definitions.

Definition 3.5(1)

(1) *The expectation value $\mathcal{E}(\wp^A)$ of observable A is defined to be*

$$\mathcal{E}(\wp^A) = \sum_{\ell=1}^{n} a_\ell \wp^A(a_\ell). \tag{3.32}$$

(2) *The variance* $\mathrm{Var}\,(\wp^A)$ *of observable* A *is defined to be the mean-square-deviation from the mean, i.e.,*

$$\mathrm{Var}\,(\wp^A) = \sum_{\ell=1}^{n} \left(a_\ell - \mathcal{E}(\wp^A)\right)^2 \wp^A(a_\ell). \tag{3.33}$$

(3) *The uncertainty* $\Delta(\wp^A)$ *of observable* A *is defined to be the root-mean-square-deviation from the mean, i.e.,*

$$\Delta(\wp^A) = \sqrt{\mathrm{Var}\,(\wp^A)}, \tag{3.34}$$

or explicitly,[17]

$$\Delta(\wp^A) = \left(\sum_{\ell=1}^{n} \left(a_\ell - \mathcal{E}(\wp^A)\right)^2 \wp^A(a_\ell)\right)^{1/2}. \tag{3.35}$$

Equation (3.32) corresponds to Eq. (3.3). In practical applications we can interpret $\mathcal{E}(\wp^A)$ as the *average value* or the *mean value* of A in the experiment. The variance serves as an indication of the spread of the values from the expectation value. The mean of individual deviations $a_\ell - \mathcal{E}(\wp^A)$, i.e., $\sum_\ell \left(a_\ell - \mathcal{E}(\wp^A)\right)\wp^A(a_\ell)$, may produce a misleading result of zero if individual values are evenly spread out on both sides of the expectation value, with some deviation being negative and some being positive. It is necessary to use the squares of the deviations in the definition. We can evaluate the variance and standard deviation in terms of expectation values as seen in the following theorem which can be easily verified.

Theorem 3.5(1) *The variance is equal to the difference between the mean of the square of* A *and the square of the mean of* A, i.e.,

$$\mathrm{Var}\,(\wp^A) = \mathcal{E}\left(\wp^{A^2}\right) - \mathcal{E}\left(\wp^A\right)^2, \tag{3.36}$$

where

$$\mathcal{E}\left(\wp^{A^2}\right) = \sum_{\ell=1}^{n} a_\ell^2 \wp^A(a_\ell), \quad \mathcal{E}\left(\wp^A\right)^2 = \left(\sum_{\ell=1}^{n} a_\ell \wp^A(a_\ell)\right)^2. \tag{3.37}$$

In a general statistical experiment, the outcomes may not be numerical, e.g., in a coin tossing experiment. The notion of expectation value does not apply. However, we can quantify non-numerical outcomes by introducing a function to assign a value to each non-numerical

[17] This is also known as the *standard deviation*.

outcome. Such functions are known in probability theory as *random variables*. Since we shall only consider experiments with numerical outcomes we shall avoid introducing this notion of random variables altogether.

3.6 Probability Distribution Functions

We can introduce a new function on the real line $\mathbb{R}$, known as a *probability distribution function*, to correspond to a given probability mass function $\wp$ defined on the (numerical) sample space $S_{\text{am}} = \{a_1, a_2, \ldots, a_n\}$ of a statistical experiment.

Definition 3.6(1) *The probability distribution function corresponding to a probability mass function $\wp$ is a function $\mathcal{F}$ on the real line $\mathbb{R}$ with its value $\mathcal{F}(\tau)$ defined to be the probability of an outcome being less than or equal to τ for every τ in $\mathbb{R}$.*

Explicitly $\mathcal{F}(\tau)$ is related to $\wp(a_\ell)$ by

$$\mathcal{F}(\tau) = \begin{cases} 0, & \tau < a_1 \\ \wp(a_1), & a_1 \le \tau < a_2 \\ \wp(a_1) + \wp(a_2), & a_2 \le \tau < a_3 \\ \cdots & \\ \sum_{\ell=1}^{j} \wp(a_\ell), & a_j \le \tau < a_{j+1} \\ \cdots & \\ \sum_{\ell=1}^{n} \wp(a_\ell) = 1, & a_n \le \tau \end{cases} . \tag{3.38}$$

The function $\mathcal{F}(\tau)$ is piecewise-constant with discontinuous steps at $\tau = a_\ell$. It is useful to introduce the following limiting processes:

$$\text{from the right} \quad \mathcal{F}(\tau + 0) := \lim_{\epsilon \to 0} \mathcal{F}(\tau + \epsilon), \quad \epsilon > 0. \tag{3.39}$$

$$\mathcal{F}(a_\ell + 0) := \lim_{\epsilon \to 0} \mathcal{F}(a_\ell + \epsilon), \quad \epsilon > 0. \tag{3.40}$$

$$\text{from the left} \quad \mathcal{F}(\tau - 0) := \lim_{\epsilon \to 0} \mathcal{F}(\tau - \epsilon), \quad \epsilon > 0. \tag{3.41}$$

$$\mathcal{F}(a_\ell - 0) := \lim_{\epsilon \to 0} \mathcal{F}(a_\ell - \epsilon), \quad \epsilon > 0. \tag{3.42}$$

The function is continuous from the right for all τ, even at a discontinuity, but discontinuous from the left at a discontinuity, i.e.,

$$\mathcal{F}(\tau + 0) = \mathcal{F}(\tau) \ \forall \tau, \quad \mathcal{F}(a_\ell - 0) \neq \mathcal{F}(a_\ell). \tag{3.43}$$

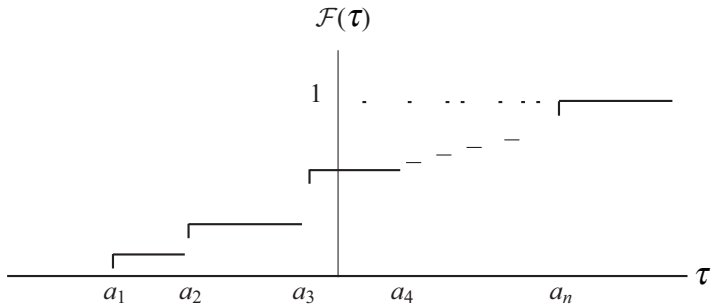


Figure 3.1 Probability distribution function.

This is shown in the above plot of $\mathcal{F}(\tau)$ against τ.

Probability distribution functions possess some characteristic physical and mathematical properties.[18]

Physical properties:

PP3.6(1) $\mathcal{F}^A(\tau) = 0$ if τ is less than a_1 and $\mathcal{F}(\tau) = 1$ if τ is bigger than or equal a_n. Generally each value of function is related to the probability of many outcomes. Probabilities for many outcomes are known as **cumulative probabilities**.

PP3.6(2) The probability $\wp(\tau)$ of a single outcome τ is given by

$$\wp(\tau) = \mathcal{F}^A(\tau + 0) - \mathcal{F}^A(\tau - 0). \tag{3.44}$$

If $\tau \neq a_\ell$ then $\wp(\tau) = 0$ as expected and if $\tau = a_\ell$ then the probability for the single outcome $\tau = a_\ell$ in Eq. (3.44) is equal to $\wp(a_\ell)$, i.e.,

$$\wp(a_\ell) = \mathcal{F}^A(a_\ell + 0) - \mathcal{F}^A(a_\ell - 0). \tag{3.45}$$

Hence the probability of every outcome is obtainable from $\mathcal{F}(\tau)$.

PP3.6(3) The probability that an outcome lies in an interval $(\tau_1, \tau_2]$ is equal to $\mathcal{F}(\tau_2) - \mathcal{F}(\tau_1)$.

Mathematical Properties:

MP3.6(1) *Non-decreasing* $\mathcal{F}(\tau_1) \leq \mathcal{F}(\tau_2)$ if $\tau_1 \leq \tau_2$.

[18] Kingman and Taylor pp. 95–99.

MP3.6(2) *Values at infinities* $\mathcal{F}(-\infty) = 0,\ \mathcal{F}(+\infty) = 1.$

MP3.6(3) *Continuity from the right* $\mathcal{F}(\tau + 0) = \mathcal{F}(\tau).$

The function $\mathcal{F}(\tau)$ rises from 0 to 1. It is non-decreasing and continuous from the right. Functions with these mathematical properties play a crucial role in probability theory. We shall therefore give a general definition of these kind of functions.

Definition 3.6(2)

(1) *A real-valued function defined on the real line IR is said to be non-decreasing if it possesses property* MP3.6(1).[19]

(2) *A real-valued function defined on the real line IR is called a* **distribution function** *if it possesses properties* MP3.6(1), MP3.6(2) and MP3.6(3).[20]

A distribution function is non-negative. It can be continuous throughout or it can be piecewise-constant with discontinuous jumps. The function in Eq. (3.38) and depicted in Fig. 3.1 is a distribution function. This kind of functions are also known as *jump functions* and the increment of such a function at a discontinuity is known as the *jump* of the function at the discontinuity. Distribution functions have many other applications, e.g., they can be used to describe continuous or discrete distribution of mass along the real line.[21]

We now have three different ways to specify a probability distribution. The approaches in terms of probability measures and probability distribution functions are superfluous for discrete sample spaces. As will be seen in Chapter 4 probability measures and probability distribution functions are essential when one deals with continuous sample spaces for which the assignment of a non-zero probability to each individual outcome is generally not possible. The present discussion also serves to provide an intuition to the

[19] *See* Theorem 4.2.2(1) for further properties of non-decreasing functions.
[20] Kingman and Taylor p. 96.
[21] Smirnov pp. 18–23. *See* Lebesgue–Stieltjes measure introduced in terms of Eqs. (4.14) to (4.17).

important mathematical concept of *measures* to be discussed later in §4.1.

Exercises and Problems

Q3(1) Prove Theorem 3.4(1).[22]

Q3(2) Prove Theorem 3.5(1).

Q3(3) What is the value $\mathcal{F}(a_4) - \mathcal{F}(a_3)$ of the probability distribution function in Eq. (3.38)?

Q3(4) In an experiment of tossing a fair die a number from 1 to 6 will be obtained with equal probabilities.

(a) Write down the probability mass function $\wp$ and evaluate the expectation value and the uncertainty.

(b) Write down the corresponding probability distribution function $\mathcal{F}(\tau)$ and sketch a plot of $\mathcal{F}(\tau)$ versus τ. What are the values $\mathcal{F}(\tau)$ at $\tau = 0.9,\ 1,\ 2.5,\ 6$ and 6.1?

[22] Lipschutz pp. 40–41. Roman Vol. 1 Problems 7.2a-1 on p. 313.

Chapter 4

Probability Theory for Continuous Variables

When dealing with statistical experiments whose outcomes form a continuous set of real numbers τ, e.g., $I\!R$, we do not have to consider every subset of $I\!R$ as an event, since every run of the experiment will give a value in a certain interval. It is sufficient to identify a structured family of subsets of $I\!R$ large enough to include individual numbers, intervals and sets obtained by operations of unions, intersections and complements of intervals. The generally adopted subsets are *Borel sets* of the reals, named after the French mathematician Émile Borel (1871–1956). Borel sets, which include all open and closed intervals, individual numbers as well the empty set and $I\!R$, are large enough to be able to quantify all experimental outcomes.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

4.1 Borel Sets, Borel Functions and Measures

4.1.1 Borel Sets of the Reals and Borel Functions

Definition 4.1.1(1)[1] *Borel sets of the reals, denoted by* $\mathbb{B}(\mathbb{R})$ *or* $\mathbb{B}$*, is the smallest family of subsets* Λ *of* $\mathbb{R}$ *which is closed under complements and countable unions, i.e.,*

$$\Lambda \in \mathbb{B} \Rightarrow \Lambda^c \in \mathbb{B} \quad \text{and} \quad \Lambda_\ell \in \mathbb{B} \Rightarrow \cup_\ell \Lambda_\ell \in \mathbb{B}, \tag{4.1}$$

and contains all open intervals.

Borel sets are also closed under countable intersection.[2] These properties are needed for many applications, e.g., the closure of countable union is required in the definition of *measures* in Definition 3.4(2). All the subsets of $\mathbb{R}$ we shall encounter in this book, e.g., *open intervals* (a, b), *half-open intervals* $(a, b]$ and $[a, b)$, closed intervals $[a, b]$ and sets containing a single number, are Borel sets. The reals $\mathbb{R}$ and the empty set $\emptyset$ are also Borel sets.

For an experiment with only a countable set of outcomes, we can assign a probability to each outcome. We can then obtain the probabilities for all events. This simple approach is not possible when the outcomes form a continuum, e.g., $\mathbb{R}$. If we were to assign a non-zero probability to each outcomes $\tau \in \mathbb{R}$ we would have difficulty satisfying the normalisation requirement PM3.4(2) on probability measures that the total probability must be 1. The general approach is to assign probabilities directly to events, not to individual outcomes, i.e., for a continuous sample space $\mathbb{R}$ the approach would be as follows:

(1) We start by identifying each event with a Borel set of $\mathbb{R}$ and vice versa.
(2) A probability distribution is then described by a set function from $\mathbb{B}$ into the interval $[0, 1]$.[3]

[1] Reed and Simon Vol. 1 pp. 14–16.
[2] This follows Theorem 3.2.1(1). Some authors would include closure under countable intersection explicitly in their definition of Borel sets.
[3] As for discrete sample spaces, a *probability distribution* is the set of probabilities for all the events in a sample space.

The kind of set functions suitable for describing probability distributions are special cases of a general class of set functions known as *measures* which have many applications in mathematics.

Before looking into set functions let us start by considering real-valued functions f defined on the real line $\mathbb{R}$. Generally the image of a Borel set under an arbitrary function f is not necessarily a Borel set and similarly the inverse image of a Borel set under f is not always a Borel set. For our applications we are interested in functions under which the inverse image of a Borel set is always a Borel set. The reason for the definition in terms of a requirement on the inverse images, rather than on the images, will become clear when we consider Lebesgue integration in §4.2.2.

Let f be a real-valued function with $\mathbb{R}$ as its domain and codomain. Let Λ be a Borel set of the codomain. The inverse image $f^{-1}(\Lambda)$ of Λ may or may not be a Borel set of the domain.

Definition 4.1.1(2) *A real-valued function f on $\mathbb{R}$ is said to be a* **Borel function** *if the inverse image $f^{-1}(\Lambda)$ of every Borel set Λ defined by*

$$f^{-1}(\Lambda) := \{\tau \in \mathbb{R} : f(\tau) \in \Lambda\} \tag{4.2}$$

is a Borel set. A complex-valued function is a Borel function if both its real and imaginary parts are Borel functions.

Note that the inverse image of an interval may be a set of isolated points, a union of intervals and so on. The inverse image of a Borel set lying entirely outside the range of the function would be the empty set $\emptyset$. All the usual algebraic operations on Borel functions, e.g., sum and product, would result in Borel functions. The following are two simple but important examples of Borel functions on $\mathbb{R}$:

E4.1.1(1) *Characteristic functions* on $\mathbb{R}$ The characteristic function of a Borel set Λ on $\mathbb{R}$, denoted by χ_Λ, is defined by

$$\chi_\Lambda(\tau) := \begin{cases} 1, & \tau \in \Lambda \\ 0, & \tau \notin \Lambda \end{cases}. \tag{4.3}$$

If Λ is an interval, e.g., $\Lambda = (a, b]$, then $\chi_{(a,b]}(\tau)$ is equal to 1 inside the interval, i.e., for all $\tau \in (a, b]$, and vanishes outside the interval, i.e., for all $\tau \notin (a, b]$. This is a Borel function since the inverse image of every Borel set is again a Borel set. For example, the inverse images of the Borel sets $\{1\}$, $\{0\}$ and $\{0, 1\}$ are

$$\chi_\Lambda^{-1}(\{1\}) = \Lambda, \quad \chi_\Lambda^{-1}(\{0\}) = \Lambda^c, \quad \chi_\Lambda^{-1}(\{0, 1\}) = I\!R. \tag{4.4}$$

E4.1.1(2) *Simple functions*[4] Let Λ be a Borel set and let $\{\Lambda_\ell, \ell = 1, 2, \cdots, n\}$ be a finite collection of disjoint Borel sets forming a *partition* of Λ.[5] In other words, Λ is equal to the union of Λ_ℓ, i.e.,

$$\Lambda = \cup_\ell \Lambda_\ell, \quad \Lambda_\ell \cap \Lambda_{\ell'} = \emptyset. \tag{4.5}$$

Let $\{a_1, a_2, \cdots, a_n\}$ be a set of arbitrary real numbers, then a function $f_s(\tau)$ of the form[6]

$$f_s(\tau) := \sum_{\ell=1}^{n} a_\ell \, \chi_{\Lambda_\ell}(\tau) \tag{4.6}$$

is called a *simple function* defined on Λ which can be extended to $I\!R$. A simple function is just a weighted sum of characteristic functions, and is hence a Borel function. It has a finite number of different values if n is finite. A simple function is said to be *piecewise-constant* if the Borel sets Λ_ℓ are intervals, the probability distribution function in Eq. (3.38) being an example. The following examples, known as **step functions**, serve to illustrate the simplicity of simple functions defined on $I\!R$:[7]

E4.1.1(2)(a) The *unit step function* g_{us} is defined by[8]

$$g_{\text{us}}(\tau) := \begin{cases} 0, & \text{if } \tau < 0 \\ 1, & \text{if } \tau \geq 0 \end{cases}. \tag{4.7}$$

[4] Simple functions can be extended to include cases where n is infinite.
[5] Generally the term *partition* of a set S means a collection of subsets of S, not necessarily finite or pairwise disjoint, whose union is equal to S.
[6] Fano p. 219.
[7] Despite their simplicity, these functions have many applications.
[8] This is also known as the *Heaviside step function*.

E4.1.1(2)(b) An antisymmetric step function g_{as} satisfies $g_{\text{as}}(\tau) = -g_{\text{as}}(-\tau)$, e.g.,

$$g_{\text{as}}(\tau) := \begin{cases} -1, & \text{if } \tau < 0 \\ 0, & \text{if } \tau = 0 \\ +1, & \text{if } \tau > 0 \end{cases} . \tag{4.8}$$

E4.1.1(2)(c) A step function with step at $\tau = a_1$ and increment $g_1 - g_0$, where $g_1 > g_0$:

$$g_{\text{s}}(\tau) := \begin{cases} g_0, & \text{if } \tau < a_1 \\ g_1, & \text{if } \tau \geq a_1 \end{cases} . \tag{4.9}$$

E4.1.1(2)(d) Non-decreasing step functions g_{nd} with discontinuities at $\tau = a_1, a_2, a_3, \cdots$ where $a_1 < a_2 < a_3 < \cdots$:

$$g_{\text{nd}}(\tau) := \begin{cases} g_0, & \text{if } \quad \tau < a_1 \\ g_1, & \text{if } \quad a_1 \leq \tau < a_2 \\ g_2, & \text{if } \quad a_2 \leq \tau < a_3 \\ \vdots \end{cases} . \tag{4.10}$$

where $g_0 < g_1 < g_2 < \cdots$. The function undergoes a discontinuous jump in value by the amount $\Delta g_\ell = g_\ell - g_{\ell-1}$ at the discontinuity at $\tau = a_\ell$. This is similar to the probability distribution function depicted in Fig. 3.1.[9]

An important result is that any Borel function can be expressed as the *pointwise limit* of an appropriate sequence of simple functions.[10] This property is generally used in the definition of Lebesgue integrals to be discussed in §4.2.2.

In practice all the functions we shall encounter, including discontinuous and continuous functions, are Borel functions. Borel functions are so pervasive that it is difficult to write down a function which is not Borel. Borel functions play an important role in the theory of Lebesgue integration since they are related to the integrability of functions.

---

[9] Smirnov pp. 6–7.

[10] Kingman and Taylor p. 105. Roman Vol. 1 pp. 341–344. *Pointwise convergence* means $f_\ell(x) \to f(x)$ for all x as $\ell \to \infty$.

4.1.2 Lebesgue and Lebesgue–Stieltjes Measures

Definition 4.1.2(1)[11]

(1) *A measure* $\mathcal{M}$ *on the Borel sets* $\mathbb{B}$ *is a set function from the Borel sets into the extended reals* $\mathbb{R}_{ex}$*, i.e.,*

$$\mathcal{M} : \mathbb{B} \mapsto \mathbb{R}_{ex} \quad by \quad \Lambda \in \mathbb{B} \mapsto \mathcal{M}(\Lambda) \in \mathbb{R}_{ex}, \tag{4.11}$$

with the following properties:[12]

(a) *Non-negativity* $\mathcal{M}(\Lambda) \geq 0.$
(b) *Empty set* $\emptyset$ $\mathcal{M}(\emptyset) = 0.$
(c) *Additivity* $\mathcal{M}(\cup_\ell \Lambda_\ell) = \sum_\ell \mathcal{M}(\Lambda_\ell), \quad \Lambda_\ell \cap \Lambda_{\ell'} = \emptyset.$

(2) *A set* $\Lambda \in \mathbb{B}$ *is called a set of measure zero if* $\mathcal{M}(\Lambda) = 0.$

A measure is basically a real-valued set function which is non-negative and countably additive. Let us illustrate the concept of measures with examples:

E4.1.2(1) *Lebesgue measure on* $\mathbb{B}$[13] Intuitively the positive and the additive nature of a measure suggests that it is an assignment of a numerical value to some properties of the Borel sets. An example is a set function which assigns a length to every Borel set. To define this set function we start with its values for half-open intervals and individual numbers:

(1) We define the value of $\mathcal{M}\big((\tau_1, \tau_2]\big)$ by $\mathcal{M}\big((\tau_1, \tau_2]\big) := \tau_2 - \tau_1.$
(2) We define the value of $\mathcal{M}\big(\{\tau\}\big)$ by $\mathcal{M}\big(\{\tau\}\big) := 0.$

Using the additive property we can extend the function consistently to all intervals with the results

$$\mathcal{M}\big([\tau_1, \tau_2]\big) = \mathcal{M}\big([\tau_1, \tau_2)\big) = \mathcal{M}\big((\tau_1, \tau_2)\big) = \tau_2 - \tau_1. \tag{4.12}$$

[11] Kingman and Taylor pp. 54–55. The extended reals $\mathbb{R}_{ex}$ consists of the reals $\mathbb{R} = (-\infty, \infty)$ together with $-\infty$ and ∞, i.e., we have

$$\mathbb{R}_{ex} := -\infty \cup \mathbb{R} \cup \infty = [-\infty, \infty].$$

We can have $\mathcal{M}(\Lambda) = \infty$ for some Λ. We can also have a more general concept of measures with mappings into a set of non-numerical quantities such as operators. We shall discuss such non-numerical measures in §15.1.

[12] Here Λ_ℓ are mutually disjoint Borel sets. Note that $\ell \neq \ell'$ in (c).

[13] Lebesgue (1875–1941) was a French mathematician famous for his theory of integration.

In other words, we define the length of an interval to be $\tau_2 - \tau_1$ and we define the length of a single point $\tau \in I\!R$ to be zero. It can be shown that there is a unique extension of this set function to all Borel sets. The resulting measure is known as the *Lebesgue measure on* $I\!B$ and is denoted by $\mathcal{M}_l$.[14]

A single point in $I\!R$ does not have any length. This is reflected by the fact that a singleton set $\{\tau\}$ has a Lebesgue measure zero, and it is hence called a *set of Lebesgue measure zero* or simply *a set of measure zero*. It follows from the additive property that a countable union of isolated points, e.g., $\{\tau_1, \tau_2, \cdots\} = \{\tau_1\} \cup \{\tau_2\} \cup \cdots$ is also a set of Lebesgue measure zero. Real numbers are divided into *rational* and *irrational* numbers. A number is rational if it can be expressed as a fraction n/m of two integers n and $m \neq 0$, and it is irrational otherwise. Let $[a, b]$ be a closed interval. The set of rational numbers in $[a, b]$ is known to be countable.[15] Hence it is a set of measure zero.[16] The Lebesgue measure of the set of irrational numbers in $[a, b]$ is equal to $b - a$.

A property or an equation which holds except on a set of measure zero is said to hold **almost everywhere**. The following examples serve to illustrate the concept:

E4.1.2(1)(a) A real-valued function on $I\!R$ which is zero everywhere, i.e., $f(\tau) = 0\ \forall \tau \in I\!R$, is called the *zero function*. A function which vanishes everywhere except at a number of isolated point is equal to the zero function almost everywhere, e.g., the following function

$$f(\tau) := \begin{cases} 1 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases} \tag{4.13}$$

is equal to the zero function almost everywhere. We simply say that *this function is zero almost everywhere*.

E4.1.2(1)(b) Two functions $f_1(\tau)$ and $f_2(\tau)$ on $I\!R$ which are equal except at a number of isolated points is said to be equal almost everywhere.

---

[14] Kingman and Taylor p. 69, p. 79. Roman Vol. 1 p. 314. For the empty set, we have $\mathcal{M}_l(\emptyset) = 0$,

[15] Kingman and Taylor p. 12.

[16] Byron-Fuller Vol. 1 p. 215.

E4.1.2(2) *Lebesgue–Stieltjes measures on $\mathbb{B}$*[17] The *Lebesgue measure* is a measure of the length of an interval along the real line. There is a need for a measure of other quantities. Consider a distribution of a finite amount of mass along the real line $\mathbb{R}$ specified by a continuous density function $\rho(\tau) \geq 0, \quad \tau \in \mathbb{R}$. The total mass in the interval $(-\infty, \tau]$ is given by

$$m(\tau) = \int_{-\infty}^{\tau} \rho(\tau')d\tau', \tag{4.14}$$

which is a function of τ. This function has some of the properties of a distribution function introduced in Definition 3.6 (2), i.e., $m(\tau)$ is non-decreasing and continuous from the right. The mass in any interval $(\tau_1, \tau_2]$ is equal to $m(\tau_2) - m(\tau_1) \geq 0$. We can introduce a measure, known as a Lebesgue–Stieltjes measure on $\mathbb{B}$ and denoted by $\mathcal{M}_{\mathrm{ls,m}}$, by first defining its values on an interval and on a isolated point τ in terms of the function $m(\tau)$, i.e.,

$$\mathcal{M}_{\mathrm{ls,m}}\big((\tau_1, \tau_2]\big) := m(\tau_2) - m(\tau_1), \quad \mathcal{M}_{\mathrm{ls,m}}(\{\tau\}) := 0. \tag{4.15}$$

As in the case of the Lebesgue measure, we can extend these values to generate a measure on $\mathbb{B}$, which will give us the value of the mass in any Borel set.

The above example suggests that we can generate a measure using a function $g(\tau)$ which is non-decreasing and continuous from the right such as a distribution function in the following manner:

(1) Define a set function $\mathcal{M}_{\mathrm{ls,g}}$ on all bounded half-open intervals $(\tau_1, \tau_2]$ by

$$\mathcal{M}_{\mathrm{ls,g}}\big((\tau_1, \tau_2]\big) := g(\tau_2) - g(\tau_1). \tag{4.16}$$

(2) The value of the function for a singleton set, i.e., $\mathcal{M}_{\mathrm{ls,g}}(\{\tau\})$, is defined by[18]

$$\mathcal{M}_{\mathrm{ls,g}}(\{\tau\}) := g(\tau + 0) - g(\tau - 0). \tag{4.17}$$

This expression follows Eq. (3.44) and it is a generalisation of Eq. (4.15) to allow for discontinuities of the function $g(\tau)$.

[17] Kingman and Taylor p. 95. Stieltjes (1856–1894) was a Dutch mathematician.
[18] Since $g(\tau)$ is continuous from the right we have $g(\tau + 0) = g(\tau)$.

(3) Extend this set function to all Borel sets to produce a unique measure on the Borel sets of $\mathbb{R}$.[19]

The resulting measure, denoted by $\mathcal{M}_{\mathrm{ls,g}}$, is referred to as the *Lebesgue–Stieltjes measure generated by the function* $g(\tau)$. The function $g(\tau)$ does not need to be continuous.[20] An example would be a probability distribution function $\mathcal{F}(\tau)$ shown in Fig. 3.1.

The measure $\mathcal{M}_{\mathrm{ls,g}}$ has the following values:[21]

$$\mathcal{M}_{\mathrm{ls,g}}(\emptyset) = 0, \tag{4.18}$$

$$\mathcal{M}_{\mathrm{ls,g}}(\{\tau\}) = g(\tau) - g(\tau - 0), \tag{4.19}$$

$$\mathcal{M}_{\mathrm{ls,g}}((\tau_1, \tau_2]) = g(\tau_2) - g(\tau_1), \tag{4.20}$$

$$\mathcal{M}_{\mathrm{ls,g}}((\tau_1, \tau_2)) = g(\tau_2 - 0) - g(\tau_1), \tag{4.21}$$

$$\mathcal{M}_{\mathrm{ls,g}}([\tau_1, \tau_2]) = g(\tau_2) - g(\tau_1 - 0), \tag{4.22}$$

$$\mathcal{M}_{\mathrm{ls,g}}([\tau_1, \tau_2)) = g(\tau_2 - 0) - g(\tau_1 - 0), \tag{4.23}$$

Lebesgue–Stieltjes measures are an extension of Lebesgue measure since by letting $g(\tau) = \tau$ we recover the Lebesgue measure.

E4.1.2(3) *Probability measures on* $\mathbb{B}$ The concept of measures is not limited to assigning values to some properties of the Borel sets themselves. A measure can be used to give values to quantities associated with the Borel sets. For example, we can use a measure to describe a probability distribution. A measure $\mathcal{M}_{\mathrm{p}}$ on $\mathbb{B}$ satisfying the additional requirement

$$\mathcal{M}_{\mathrm{p}}(\mathbb{R}) = 1 \tag{4.24}$$

is called a **probability measure on** $\mathbb{B}$, because $\mathcal{M}_{\mathrm{p}}$ satisfies the properties of a probability measure in Definition 3.4(2). As an example let us consider a statistical experiment which has $\mathbb{R}$ as its sample space with Borel sets Λ of $\mathbb{R}$ corresponding to events. Since a Borel set Λ and its complement Λ^c are disjoint and since $\Lambda \cup \Lambda^c = \mathbb{R}$ we have

$$\begin{aligned} \mathcal{M}_{\mathrm{p}}(\mathbb{R}) &= \mathcal{M}_{\mathrm{p}}(\Lambda \cup \Lambda^c) \\ &= \mathcal{M}_{\mathrm{p}}(\Lambda) + \mathcal{M}_{\mathrm{p}}(\Lambda^c) \geq \mathcal{M}_{\mathrm{p}}(\Lambda). \end{aligned} \tag{4.25}$$

[19] Roman Vol. 1 pp. 320–324. Kingman and Taylor pp. 95–96.

[20] Discontinuities are from the left since the function is assumed to be continuous from the right.

[21] Roman Vol. 1 pp. 320–321. The value of $\mathcal{M}_{\mathrm{ls,g}}(\{\tau\})$ is not necessarily zero, e.g., the left limit $g(\tau - 0)$ is not necessarily equal to $g(\tau + 0)$ at a discontinuity.

The condition $\mathcal{M}_{\mathrm{p}}(I\!R) = 1$ means that

$$\mathcal{M}_{\mathrm{p}}(\Lambda) \in [0, 1], \quad \mathcal{M}_{\mathrm{p}}(\Lambda^c) = 1 - \mathcal{M}_{\mathrm{p}}(\Lambda) \in [0, 1]. \tag{4.26}$$

We can then take $\mathcal{M}_{\mathrm{p}}(\Lambda)$ as the probability for the event Λ.

We shall return to study probability measures on a continuous sample space in detail in §4.3.

E4.1.2(4) *Lebesgue–Stieltjes and probability measures* The Lebesgue–Stieltjes measure generated by a probability distribution function possesses the properties of a probability measure. The relationship between probability distribution functions and probability measures will be discussed in detail in §4.3.

4.2 Riemann and Lebesgue Integrals

4.2.1 Riemann Integrals

The mathematics of integration is complicated because there are different types of integrals. The integrals presented in elementary calculus are *Riemann integrals*.[22] To clarify the concept let us consider a real-valued function f defined on a closed and bounded interval $[a, b]$ of $I\!R$. The function is assumed to be bounded, continuous and non-negative. Such a function can be plotted as a curve of $y = f(x)$ in the x-y plane. The Riemann integral of f over the interval $[a, b]$ is conceptually interpretable as the area under the curve.

Let $\{x_0, x_1, x_2, \cdots, x_n\}$ be a set of points in $[a, b]$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b. \tag{4.27}$$

We can approximate of the area under the curve by a series of rectangles. There are two conventional ways to proceed:

(1) Divide the interval $[a, b]$ into the following n closed intervals:[23]

$$\Lambda_1 = [x_0, x_1], \ \Lambda_2 = [x_1, x_2], \ \cdots, \ \Lambda_n = [x_{n-1}, x_n]. \tag{4.28}$$

[22] Riemann (1826–1866) was a German mathematician, a pioneer of the theory of Riemannian geometry.

[23] All these smaller intervals are closed in order to form the base of a rectangle.

Choose an arbitrary point x'_j in the j^{th} interval Λ_j. Associated with the j^{th} interval we have a rectangle of height $f(x'_j)$, width $\Delta x_j = x_j - x_{j-1}$ and area $f(x'_j)\Delta x_j$. We define the Riemann integral of the given function f over $[a, b]$ as the limit of the sum of the areas of all these rectangles as every rectangle becomes vanishingly small, i.e.[24]

$$\int_{[a,b]} f(x)dx := \lim_{\Delta x_j \to 0} \sum_{j=1}^{n} f(x'_j)\,\Delta x_j. \tag{4.29}$$

A function f is said to be *Riemann integrable* over $[a, b]$ if the above limit exists. This concept of integration can be extended to functions which are positive for some regions and negative for some other regions of the interval. The integral then represents the algebraic sum of the areas above and below the x-axis, with the areas above the x-axis being positive and the areas below the x-axis being negative.

(2) Another way to look at the problem is to construct two rectangles with each interval, i.e., we construct a rectangle with height $f_{j\min}$ which is the minimum value of $f(x)$ in Λ_j and another rectangle with height $f_{j\max}$ which is the maximum value of $f(x)$ in Λ_j. We then consider the resulting *upper* and the *lower* Riemann integrals defined by

$$\int_{[a,b]}^{+} f(x)dx := \lim_{\Delta x_j \to 0} \sum_{j=1}^{n} f_{j\max}\,\Delta x_j, \tag{4.30}$$

$$\int_{[a,b]}^{-} f(x)dx := \lim_{\Delta x_j \to 0} \sum_{j=1}^{n} f_{j\min}\,\Delta x_j. \tag{4.31}$$

When these two integrals exist and agree then we have a unique value which can be defined as the Riemann integral of f over $[a, b]$. This agrees with the definition in Eq. (4.29) for continuous functions.

[24] The limit can be proved to exist for continuous functions. The limit may not exist for functions which are highly discontinuous and unbounded.

(3) It is easy to see that a function is Riemann integrable over $[a, b]$ if the function is continuous except at a finite number of discontinuities. We can go further to show that a bounded function is Riemann integrable over a finite interval $[a, b]$ *if and only if* it has only a countable number of discontinuities.[25]

Riemann integral can be extended to include integrals of functions over open intervals and many unbounded functions. For example, consider an open interval (a, b). Let $[a_\ell, b_\ell]$ be a sequence of closed intervals inside (a, b) such that $a_\ell \to a$ and $b_\ell \to b$ as $\ell \to \infty$. We can define an integral of f over the open interval (a, b) as the limiting value of the sequence of integrals over the closed intervals $[a_\ell, b_\ell]$. An example is $f = 1/\sqrt{x}$. This function is bounded and continuous in the interval $[a > 0, b]$ and hence integrable over $[a > 0, b]$. The integral over $[0, b]$ is obtained as the limit of the integral over $[a, b]$ as $a \to 0$. A similar extension applies to integration over infinite intervals, e.g., (a, ∞), $(-\infty, \infty)$, and integration of certain unbounded functions. The resulting integrals are known as *improper Riemann integrals*. Riemann integrals are related to differentiability of functions. Let $f(x)$ be a real-valued function defined on a closed interval $[a, b]$. Then *the fundamental theorem of calculus* tells us that if $f(x)$ is continuous on $[a, b]$ and if[26]

$$F(x) = F(a) + \int_a^x f(x')\, dx', \tag{4.32}$$

then $F(x)$ is continuous and differentiable in the open interval (a, b) with its derivative

$$F'(x) := \frac{df(x)}{dx} = f(x) \quad \forall\, x \in (a, b). \tag{4.33}$$

It follows that for all $x \in (a, b)$ we have

$$F(x) = F(a) + \int_a^x F'(x')\, dx', \tag{4.34}$$

and

$$F(b) - F(a) = \int_a^b F'(x')\, dx'. \tag{4.35}$$

[25] The phrase *if and only if* is widely used in mathematics. For example, the sentence "Statement A is true if and only if statement B is true" means that if A is true then B is true and if B is true then A is true.

[26] Kingman and Taylor p. 230.

Riemann integrals turn out to be inadequate in many applications:[27]

(1) Many functions are not Riemann integrable.

(2) The integral is defined over an interval, not over a more general set of real numbers such as a Borel set.

(3) Riemann integrals have difficulty with some convergence problems, e.g., if a sequence of Riemann integrable functions $f_\ell(x)$ converges to a limit function $f(x)$ pointwise the limit function may not be Riemann integrable.

A new theory of integration, the Lebesgue's theory, is needed.[28]

4.2.2 Lebesgue Integrals

Consider a real-valued function $f(x)$ on a closed and bounded interval $[a, b]$ which is bounded, continuous and non-negative. Let $f_{\min}$ and $f_{\max}$ be the greatest lower bound and lowest upper bound of f on the interval $[a, b]$. Again we want to establish the area under the curve $y = f(x)$ in the x-y plane. The range of the function is $[f_{\min}, f_{\max}]$.[29] Let $\{y_0, y_1, y_2, \cdots, y_n\}$ be a set of points such that

$$y_0 = f_{\min} < y_1 < y_2 < \cdots < y_n = f_{\max}. \tag{4.36}$$

We can divide the range $[f_{\min}, f_{\max}]$ into n smaller intervals[30]

$$\Lambda'_1 = [y_0, y_1], \ \Lambda'_2 = [y_1, y_2], \cdots, \ \Lambda'_n = [y_{n-1}, y_n]. \tag{4.37}$$

These are intervals along the y-axis. The inverse image of Λ'_j under the function f, i.e., $\Lambda_j = f^{-1}(\Lambda'_j)$, is a subset of the x-axis. Let y'_j be an arbitrary value in Λ'_j. We may be tempted to construct a rectangle of height y'_j, width $\Delta\Lambda_j$ and area $y'_j \Delta\Lambda_j$. This is problematical since Λ_j may not in an interval and hence we may not have a natural value for the width of Λ_j. Being continuous f is a Borel function. It follows that the inverse image Λ_j under f is a Borel set. We

[27] Capiński and Kopp pp. 9–10. Kingman and Taylor p. 100.

[28] Kingman and Taylor pp. 124–126. Capiński and Kopp pp. 54–55. Roman Vol. 1 pp. 367–368.

[29] A continuous and bounded function defined on a closed and bounded interval reaches its upper and lower bounds.

[30] For a diagrammatic illustration, *see* Roman Vol. 1 p. 368.

can assign a numerical size to this Borel set using the Lebesgue measure $\mathcal{M}_l$. In other words, we can identify the width of Λ_j with $\mathcal{M}_l(\Lambda_j)$, giving an area of $y'_j\mathcal{M}_l(\Lambda_j)$. This enables us to obtain an approximation to the area under the curve $y = f(x)$ to be

$$\sum_{j=1}^{n} y'_j\mathcal{M}_l(\Lambda_j) = \sum_{j=1}^{n} y'_j\mathcal{M}_l\big(f^{-1}(\Lambda'_j)\big). \tag{4.38}$$

The limit of the sum as the width $\Delta\Lambda'_j$ of every intervals Λ'_j tends to zero is defined to be the *Lebesgue integral* of f over the interval $[a, b]$, provided the limit exists.[31] The function is then said to be *Lebesgue-integrable* over $[a, b]$. This integral is denoted by

$$\int_{[a,b]} f(x)d\mathcal{M}_l(x). \tag{4.39}$$

The above idea of integration is easily extended to the integration of discontinuous simple functions. Consider a simple function f_s in Eq. (4.6). The range of f_s is the finite set $\{a_\ell, \ell = 1, 2, \cdots, n\}$. The inverse image of a singleton set $\{a_\ell\}$ under f_s is Λ_ℓ, i.e., $f_s^{-1}(\{a_\ell\}) = \Lambda_\ell$. In keeping with the spirit of Eq. (4.38) we can define the *Lebesgue integral* of f_s over the Borel set Λ to be the following sum:[32]

$$\begin{aligned}\int_\Lambda f_s(x)d\mathcal{M}_l(x) &:= \sum_{\ell=1}^{n} a_\ell\mathcal{M}_l\big(f_s^{-1}(\{a_\ell\})\big)\\ &= \sum_{\ell=1}^{n} a_\ell\mathcal{M}_l(\Lambda_l).\end{aligned} \tag{4.40}$$

For a piecewise-constant function as depicted in Fig. 3.1, this integral agrees with its Riemann integral. Generally Lebesgue integrals of continuous functions would agree with their Riemann integrals. However, the Lebesgue method of integration can be applied to a larger class of functions some of which are not Riemann integrable. An example is the following extremely discontinuous function D defined on the interval $[0, 1]$ by

$$D(x) := \begin{cases} 1 \text{ if } x \text{ is a rational number} \\ 0 \text{ if } x \text{ is an irrational number} \end{cases}. \tag{4.41}$$

[31] Roman Vol. 1 p. 367. Capiński and Kopp p. 73.
[32] Kingman and Taylor p. 110. *See* Eq. (4.43).

This is known as the *Dirichlet function*. It is discontinuous at every point in [0, 1]. Intuitively this is because there is an irrational number next to every rational number. This function is not Riemann integrable since its upper Riemann integral is 1 while its lower Riemann integral is 0.[33] To see if this function is Lebesgue integrable over an interval $[a, b]$ we first observe that this function has a range $\{0, 1\}$, i.e., the range consists of only two values 0 and 1. The inverse image $D^{-1}(\{1\})$ of the singleton set $\{1\}$ is the set of rational numbers which is a set of measure zero, i.e., $\mathcal{M}_\ell(D^{-1}(\{1\})) = 0$, and the inverse image $D^{-1}(\{0\})$ of the singleton set $\{0\}$ is the set of irrational numbers which has a measure $b-a$, i.e., $\mathcal{M}_\ell(D^{-1}(\{0\})) = b-a$. The sum in Eq. (4.40) is zero, i.e.,

$$1 \times \mathcal{M}_l(D^{-1}(\{1\})) + 0 \times \mathcal{M}_l(D^{-1}(\{0\})) = 0. \tag{4.42}$$

It follows that the Dirichlet function is Lebesgue integrable over any interval with an integral of value 0.

An alternative approach to Lebesgue integration is to establish such integrals in terms of simple functions, rather than through the construction in Eq. (4.38):[34]

(1) Let f_s be a simple function defined on the Borel set Λ, as given in Eq. (4.6). Define the Lebesgue integral of f_s over Λ to be

$$\int_\Lambda f_s(x) d\mathcal{M}_l(x) := \sum_{\ell=1}^{n} a_\ell \, \mathcal{M}_l(\Lambda_\ell). \tag{4.43}$$

This can be extended to be over any Borel set $\Lambda_{\text{any}} \subset \Lambda$ by

$$\int_{\Lambda_{\text{any}}} f_s(x) d\mathcal{M}_l(x) := \sum_{\ell=1}^{n} a_\ell \, \mathcal{M}_l(\Lambda_\ell \cap \Lambda_{\text{any}}). \tag{4.44}$$

(2) Since any Borel function can be expressed as the limit of an appropriate sequence of simple functions, a result mentioned earlier in §4.1.1 in relation to simple functions, we can define the Lebesgue integral of an arbitrary Borel function as the limit

[33] Byron-Fuller Vol. 1 p. 215.

[34] Kingman and Taylor pp. 110–114. Capiński and Kopp pp. 71–97. Roman Vol. 1 pp. 339, 359, 367.

of the integrals of an appropriate sequence of simple functions. We can also have functions $f(x)$ integrable over the entire real line, i.e., there are functions such that

$$\int_{I\!R} f(x) d\mathcal{M}_l(x) \tag{4.45}$$

is defined and has a finite value.

(3) If a function f is integrable over the entire real line its integral over any given Borel set Λ is defined by

$$\int_{\Lambda} f(x) d\mathcal{M}_l(x) := \int_{I\!R} f(x) \chi_{\Lambda}(x)\, d\mathcal{M}_l(x), \tag{4.46}$$

where $\chi_{\Lambda}(x)$ is the characteristic function of the Borel set Λ.

Lebesgue integral is a generalisation of Riemann integral. If a function f is Riemann integrable on $[a, b]$, then it can be shown to be also Lebesgue integrable, and its Lebesgue integral agrees with its Riemann integral. This happens in many practical applications where the functions are well behaved, e.g., they may be continuous or continuous almost everywhere with a finite number of discontinuities. It is often convenient to write down the Lebesgue integral in the form of a Riemann integral by replacing $d\mathcal{M}_l(x)$ by dx. Lebesgue integrals can be over an open interval or a half-open interval. Since the Lebesgue measure of a single point is zero the integral will give the same value whether it is over the closed interval $[a, b]$ or half-open intervals $[a, b)$, $(a, b]$, (a, b).[35] We can write down a Lebesgue integral over an interval, whether it is open, half-open or closed as

$$\int_a^b f(x) d\mathcal{M}_l(x). \tag{4.47}$$

Lebesgue integrals overcome the problem of lack of convergence of Riemann integrals. In other words, if an appropriate sequence of functions tends to another function pointwise, then the Lebesgue integrals of this sequence of functions would converge to the Lebesgue integral of the limit function.[36]

[35] Kingman and Taylor p. 124. Burril p. 113.
[36] Roman Vol. 1 p. 359. Kingman and Taylor p. 125.

A non-negative function f on $[a, b]$, i.e., $f(x) \geq 0 \quad \forall x \in [a, b]$, with a zero Lebesgue integral over $[a, b]$ does not imply that f is the zero function on $[a, b]$, i.e.,

$$\int_{[a,b]} f(x)\, d\mathcal{M}_l(x) = 0 \quad \not\Rightarrow \quad f(x) = 0 \ \forall x \in [a, b]. \tag{4.48}$$

The Dirichlet function is an example. Another example is the function f in Eq. (4.13). This function would have a zero Lebesgue integral over any interval $[a, b]$ containing the point $x = 0$, since the values of f on a set of Lebesgue measure zero do not contribute to the integral. In many applications which involve Lebesgue integration we can simply identify such functions with the zero function on $[a, b]$. Similarly we can deem two function to be equal even though they are only equal almost everywhere.

It should be pointed out that the notion of Lebesgue integrals hinges on the partition of the range of the function f into intervals Λ'_j and on the Lebesgue measure $\mathcal{M}_l(\Lambda_j)$ of the inverse image $\Lambda_j = f^{-1}(\Lambda'_j)$ of each interval Λ'_j in the range of the function. This requires all those inverse images to be Borel sets since the Lebesgue measure is defined only for Borel sets. It follows that Lebesgue integrals are definable only for Borel functions.

Lebesgue integrals are also related the differentiability of functions. The relationship is more complicated than those for Riemann integrals given in Eq. (4.33). Let us first consider the relationship between continuity and differentiability. As an example consider the function $f(x) = |x|$. This function is continuous everywhere. It is differentiable for $x < 0$ and $x > 0$ but it is not differentiable at $x = 0$ since its derivatives from the left and from the right of the point $x = 0$ are not equal, i.e., $f'(+0) = 1 \neq f'(-0) = -1$. The function $f(x) = |x|$ is continuous but it is differentiable only almost everywhere. A continuous function may not be differentiable at all. There are continuous functions which are nowhere differentiable.[37] In order to be differentiable almost everywhere a function needs to be more than just being continuous.

There is a class of functions defined below, known as *absolutely continuous functions*, which are differentiable almost everywhere.

[37] Spiegel (1) p. 64. Rieze and Nagy pp. 3–4.

For general applications where a function is not necessarily defined on the real line we shall denote the independent variable by τ rather than by x in what follows.

Definition 4.2.2(1)[38] *A real-valued function F defined on a closed and bounded interval $[a, b]$ is said to be* **absolutely continuous** *on $[a, b]$ if F is expressible in the form of a Lebesgue integral of a Lebesgue-integrable function f defined on $[a, b]$, i.e., if*

$$F(\tau) = F(a) + \int_a^\tau f(\tau)\, d\mathcal{M}_l(\tau). \tag{4.49}$$

A function F on $\mathbb{R}$ is absolutely continuous if it is absolutely continuous on every finite interval $[a, b]$ in $\mathbb{R}$. A complex-valued function is absolutely continuous if and only if its real and imaginary parts are both absolutely continuous.

If $[a, b]$ contains the origin, we can rewrite Eq. (4.49) as [39]

$$F(\tau) = F(0) + \int_0^\tau f(\tau)\, d\mathcal{M}_l(\tau). \tag{4.50}$$

Absolutely continuous functions on $\mathbb{R}^+$ and on $\mathbb{R}^-$ are defined similarly.

For absolutely continuous functions, the following results can be established:

R4.2.2(1) An absolutely continuous function is continuous. The converse is not always true.

R4.2.2(2) An absolutely continuous function $F(\tau)$ is differentiable almost everywhere and its derivative $F'(\tau)$ is integrable. For the function $F(\tau)$ in Eq. (4.49) its derivative $F'(\tau)$ equal to $f(\tau)$ almost everywhere, i.e., we have[40]

$$F'(\tau) = f(\tau) \quad \text{for almost every } \tau \text{ in } [a, b]. \tag{4.51}$$

This is similar to the result given in Eq. (4.35). We can rewrite the defining Eq. (4.49) as

$$F(\tau) = F(a) + \int_a^\tau F'(\tau)\, d\mathcal{M}_l(\tau), \tag{4.52}$$

[38] Roman Vol. 2 p. 545. There are other equivalent definitions.
[39] Weidmann p. 379.
[40] Kingman and Taylor p. 232. Phillips E.R. pp. 267–269.

or more conveniently as[41]

$$F(\tau) = F(a) + \int_a^\tau F'(\tau)\, d\tau. \tag{4.53}$$

This shows that an absolutely continuous function is the indefinite integral of its derivative, a highly desirable relationship between differentiation and integration. Note that $f(\tau)$ in Eq. (4.49) is not required to be continuous. The derivative $F'(\tau)$ is undefined at a discontinuity of $f(\tau)$. Consider the antisymmetric simple function $g_{\text{as}}(\tau)$ in Eq. (4.8). Define a function G on $[a, b]$ by

$$G(\tau) := G(0) + \int_0^\tau g_{\text{as}}(\tau)\, d\tau, \quad G(0) \in I\!R \tag{4.54}$$

$$\Rightarrow G(\tau) = G(0) + |\tau|. \tag{4.55}$$

This function is absolutely continuous. Equation (4.51) is satisfied. The derivative $G'(\tau)$ is undefined at $\tau = 0$.[42]

R4.2.2(3) An absolutely continuous function on $[a, b]$ whose derivative is zero almost everywhere is a constant. This result does not necessarily apply if a function is continuous but not absolutely continuous. There are functions, known as *singularly continuous functions*, which violate this result. These are real-valued functions on $I\!R$ which are continuous and differentiable almost everywhere with their derivatives being zero almost everywhere and yet they are not a constant. These functions are difficult to visualise but their existence can be proved.[43] Fortunately one rarely encounters such singular functions in physical applications.

R4.2.2(4) The sum and difference of two absolutely continuous functions are absolutely continuous.

---

[41] Writing $d\mathcal{M}_l(\tau)$ as $d\tau$.

[42] Papoulis p. 94. We can artificially assign a value to $G'(\tau)$ at $\tau = 0$ so that $G'(\tau)$ can formally have a value for all τ. Such a value does not have the usual significance.

[43] Riesz and Nagy p. 48. Phillips E.R. p. 267. Kingman and Taylor pp. 49–50, p. 238. An often quoted example of such singular functions is the *Cantor function*. The Cantor function is a function defined on the interval [0, 1] with values rises monotonically from 0 to 1. It is continuous and differentiable almost everywhere with their derivatives being zero almost everywhere. Since it is not constant it is not absolutely continuous.

R4.2.2(5) Absolutely continuous functions are crucial in defining differential operators as we shall see in §17.3 and §19.2.

Absolutely continuous functions are not the only functions which are differentiable almost everywhere. Non-decreasing functions introduced in Definition 3.6 (2) are also differentiable almost everywhere. However, a non-decreasing function is not necessarily continuous, a probability distribution function on a discrete sample space depicted in Fig. 3.1 being an example. These functions have properties given by the following theorem.

Theorem 4.2.2(1) on non-decreasing functions *Let F be a real-valued non-decreasing function defined on the interval $[a, b]$. Then we have the following*:[44]

(1) *The set of discontinuities of the function is countable.*

(2) *The function F is differentiable almost everywhere on $[a, b]$, i.e., its derivative $F'(\tau)$ exists almost everywhere.*

(3) *The derivative $F'(\tau)$ is Lebesgue integrable on $[a, b]$, with the integral related to the function F by*

$$F(b) - F(a) \geq \int_a^b F'(\tau)\, d\mathcal{M}_l(\tau). \tag{4.56}$$

As a result of Eq. (4.56) an arbitrary non-decreasing function is not necessarily expressible as an integral in the form of Eq. (4.49). This makes them unsuitable for defining differential operators.[45]

4.2.3 Riemann–Stieltjes Integrals

Riemann integrals are with respect to an independent variable. We can also define integrals with respect to a function, such as a non-decreasing function. The resulting integrals are known as *Riemann–Stieltjes integrals* or simply *Stieltjes integrals*.[46]

[44] Phillips E.R. pp. 253–260.
[45] *See* §17.3 and §19.2.
[46] Rieze and Nagy p. 105. Smirnov pp. 4–8. *See* Smirnov pp. 22–23 for a physical interpretation.

Let $[a, b]$ be a closed and bounded interval. Let τ_j be a set of points within $[a, b]$ such that

$$a = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_n = b. \tag{4.57}$$

We can divide $[a, b]$ into the following n smaller intervals:

$$[\tau_0, \tau_1],\ [\tau_1, \tau_2],\ [\tau_2, \tau_3],\ \cdots,\ [\tau_{n-1}, \tau_n]. \tag{4.58}$$

Let $g(\tau)$ be a non-decreasing function defined on $[a, b]$ which is not necessarily continuous, and let $f(\tau)$ be a continuous function defined on $[a, b]$. Then the *Riemann–Stieltjes integral of* $f(\tau)$ *with respect to the function* $g(\tau)$ *over the interval* $[a, b]$ is defined to be the following limit:

$$\lim_{\Delta\tau_j \to 0} \sum_{j=1}^{n} f(\tau_j')\Delta g_j, \tag{4.59}$$

where

$$\Delta g_j := g(\tau_j) - g(\tau_{j-1}),\quad \tau_j' \in [\tau_{j-1}, \tau_j]. \tag{4.60}$$

Compared with Eq. (4.16) we may interpret Δg_j as the increment of $g(\tau)$ over the subinterval $[\tau_{j-1}, \tau_j]$. The sum involved here is similar to the sum in Eq. (4.29) for Riemann integrals. The difference here is the use of the increment Δg_j of the function rather than the increment $\Delta\tau_j = \tau_j - \tau_{j-1}$ of τ over the subinterval. The function being non-decreasing ensures that the increment Δg_j is positive, just like $\Delta\tau_j$. We shall denote the integral by[47]

$$\int_{[a,b]} f(\tau)dg(\tau) \quad \text{or} \quad \int_a^b f(\tau)dg(\tau). \tag{4.61}$$

A wider class of functions defined on an interval, known as *functions of bounded variation*, which are functions expressible as the difference of two non-decreasing functions,[48] can also serve as $g(x)$. Moreover, we can also include integrands which are not continuous.

[47] Parzen p. 233 for a definition over a half-open interval.
[48] Kingman and Taylor p. 226.

The new integrals introduced here are more flexible and is applicable to many situations, continuous or otherwise. It is especially useful in dealing with discontinuities, hence rendering it possible to have a more uniform treatment of both continuous and discrete cases with many applications in probability theory.

The following examples illustrate the working of Riemann–Stieltjes integrals when $f(\tau)$ is continuous while $g(\tau)$ is not necessarily continuous.

E4.2.3(1) If $g(\tau)$ is absolutely continuous on an interval $[a, b]$ the Riemann–Stieltjes integral reduces to a Riemann integral, i.e., we have

$$\int_{[a,b]} f(\tau) dg(\tau) = \int_{[a,b]} f(\tau) \frac{dg(\tau)}{d\tau} d\tau. \tag{4.62}$$

When $f(\tau) = 1$ we have

$$\int_{[a,b]} dg(\tau) = g(b) - g(a). \tag{4.63}$$

E4.2.3(2) Let $g_s(\tau)$ be the step function in Eq. (4.9). This function has a discontinuous jump by an amount $\Delta g_1 = g_1 - g_0$ at $\tau = a_1$. For any continuous function $f(\tau)$ and any interval $[a, b]$ containing a_1, we have

$$\begin{aligned} \int_{[a,b]} f(\tau)\, dg_s(\tau) &= \lim_{\Delta\tau_j \to 0} \sum_{j=1}^{n} f(\tau_j') \Big(g_s(\tau_j) - g_s(\tau_{j-1}) \Big) \\ &= f(a_1) \Delta g_1, \end{aligned} \tag{4.64}$$

since $g_s(\tau_j) - g_s(\tau_{j-1}) = 0$ for every subinterval except for the one containing a_1.[49]

A special case is when g_s is replaced by the unit step function g_{us} in Eq. (4.7). We get

$$\int_{[a,b]} f(\tau)\, dg_{us}(\tau) = f(a_1). \tag{4.65}$$

E4.2.3(3) There is a link between the above Riemann–Stieltjes integrals and Dirac delta functions.[50] Equation (4.65) can be

[49] Smirnov p. 19. The partition points τ_j in Eq. (4.57) should not be chosen to coincide with the point of discontinuity of $g(x)$.

[50] Dirac (1902–1984) was a British theoretical physicist considered to be one of the most significant physicists of the 20th century. Dirac made fundamental contributions to quantum theory, especially to relativistic quantum mechanics. He shared the 1933 Nobel Prize in Physics with Schrödinger.

symbolically written in terms of an integral of a Dirac delta function, i.e.,[51]

$$\int_{[a,b]} f(\tau)\, dg_{\text{us}}(\tau) = \int_a^b f(\tau)\delta(\tau - a_1)d\tau. \tag{4.66}$$

The expression in Eq. (4.64) can be similarly written, i.e.,

$$\int_{[a,b]} f(\tau)\, dg_{\text{s}}(\tau) = \int_a^b f(\tau)\Big(\Delta g_1 \delta(\tau - a_1)\Big)d\tau. \tag{4.67}$$

One can go further with such symbolic manipulations by formally differentiating the unit step function in Eq. (4.7) and expressing its derivative as a Dirac delta function, i.e.,[52]

$$\frac{dg_{\text{us}}(\tau)}{d\tau} = \delta(\tau). \tag{4.68}$$

Following Eq. (4.62) we have

$$\begin{aligned}\int_{[a,b]} f(\tau)\, dg_{\text{us}}(\tau) &= \int_{[a,b]} f(\tau)\, \frac{dg_{\text{us}}(\tau)}{d\tau}\, d\tau \\ &= \int_a^b f(\tau)\delta(\tau - a_1)\, d\tau.\end{aligned} \tag{4.69}$$

We can similarly differentiate the step function $g_{\text{s}}(\tau)$ in Eq. (4.9) to get

$$\frac{dg_{\text{s}}(\tau)}{d\tau} = \Delta g_1\, \delta(\tau - a_1). \tag{4.70}$$

Again following Eq. (4.62) we have

$$\begin{aligned}\int_{[a,b]} f(\tau)\, dg_{\text{s}}(\tau) &= \int_a^b f(\tau)\, \frac{dg_{\text{s}}(\tau)}{d\tau}\, d\tau \\ &= \int_a^b f(\tau)\Big(\Delta g_1 \delta(\tau - a_1)\Big)d\tau.\end{aligned} \tag{4.71}$$

The discontinuity at $\tau = a_1$ effectively generates a Dirac delta function when one differentiates the step function.

E4.2.3(4) The above result can be extended to non-decreasing step functions $g_{\text{nd}}(\tau)$ in Eq. (4.10) with a finite set of discontinuities

[51] Zettili pp. 629–631 for properties of the Dirac delta function.
[52] Friedman pp. 141–142. Papoulis p. 97, p. 103.

at $\tau = a_1, a_2, \cdots, a_n$ within the interval $[a, b]$. The function undergoes a discontinuous jump in value by the amount $\Delta g_\ell = g_\ell - g_{\ell-1}$ at $\tau = a_\ell$. This is similar to the probability distribution function depicted in Fig. 3.1. We have[53]

$$\int_{[a,b]} f(\tau)\, dg_{\rm nd}(\tau) = \sum_{j=1}^{n} f(\tau_j)\Big(g_{\rm nd}(\tau_j) - g_{\rm nd}(\tau_{j-1})\Big) = f(a_1)\Delta g_1 + f(a_2)\Delta g_2 + \cdots . \tag{4.72}$$

since $g_{\rm nd}(\tau_j) - g_{\rm nd}(\tau_{j-1}) = 0$ for all subintervals except for the ones containing a value of a_ℓ. The integration process is seen to be equivalent to

(1) picking up the value $f(a_\ell)$ of the integrand and the corresponding increment Δg_ℓ at the discontinuity of $g_{\rm nd}(\tau)$ at $\tau = a_\ell$,
(2) multiplying them together to form the product $f(a_\ell)\Delta g_\ell$, and
(3) adding the products $f(a_\ell)\Delta g_\ell$ at all the discontinuities.

E4.2.3(5) The integral in Eq. (4.72) can be written in terms of Dirac delta functions, i.e.,

$$\begin{aligned}\int_{[a,b]} f(\tau)\, dg_{\rm nd}(\tau) &= \int_a^b f(\tau)\, \frac{dg_{\rm nd}(\tau)}{d\tau}\, d\tau \\ &= \int_a^b f(\tau)\Big(\Delta g_1 \delta(\tau - a_1) \\ &\qquad + \Delta g_2 \delta(\tau - a_2) + \cdots\Big)\, d\tau,\end{aligned} \tag{4.73}$$

with the derivative of $g_{\rm nd}(\tau)$ formally written as

$$\frac{dg_{\rm nd}(\tau)}{d\tau} = \Delta g_1\, \delta(\tau - a_1) + \Delta g_2\, \delta(\tau - a_2) + \cdots . \tag{4.74}$$

Some comments are required to clarify the situation when $g(x)$ is discontinuous. To avoid ambiguity partition points τ_j in Eq. (4.57) which divide the interval $[a, b]$ should not be chosen to coincide with the points of discontinuity of $g(x)$. In example E4.2.3(2) we choose $\tau_j \neq a_1$. A similar condition applies to example E4.2.3(3). A complication arises when $g(x)$ is discontinuous the end points a

[53] Smirnov pp. 6–7. Here $\tau_1 = a_1$ and $\tau_0 = a$.

and b of the interval, since we have $\tau_0 = a$ and $\tau_n = b$. The integral then has to take the jumps of $g(x)$ at a and b into consideration.[54]

4.2.4 Lebesgue–Stieltjes Integrals

This is an extension of the idea of Lebesgue integration to form integrals with respect to a Lebesgue–Stieltjes measure $\mathcal{M}_{\mathrm{ls,g}}$ generated by a non-decreasing and right continuous function $g(\tau)$ such as a distribution function, i.e., by replacing the Lebesgue measure $\mathcal{M}_{\mathrm{l}}$ in Eq. (4.38) by the Lebesgue–Stieltjes measure $\mathcal{M}_{\mathrm{ls,g}}$.[55] We denote such an integral of f over a closed and bounded interval $[a, b]$ by

$$\int_{[a,b]} f(\tau) d\mathcal{M}_{\mathrm{ls,g}}(\tau). \tag{4.75}$$

Similar to Lebesgue integrals a Lebesgue–Stieltjes integral can be defined over any Borel set Λ in the same way with the resulting integral written as

$$\int_{\Lambda} f(\tau) d\mathcal{M}_{\mathrm{ls,g}}(\tau). \tag{4.76}$$

In many applications Lebesgue–Stieltjes integrals agree with the corresponding Riemann–Stieltjes integrable.[56] This makes it easier to evaluation the Lebesgue–Stieltjes integrals. The examples of Riemann–Stieltjes integrals in the preceding section which remain valid as Lebesgue–Stieltjes integrals serve to illustrate this. It is often convenient to rewrite the integral simply as

$$\int_{\Lambda} f(\tau) dg(\tau). \tag{4.77}$$

Unlike Lebesgue integrals a Lebesgue–Stieltjes integral over a closed interval $[a, b]$ may differ from the integral over (a, b), or $[a, b)$, or $(a, b]$ due to possible discontinuities of the function $g(\tau)$ at a and b. It would then be ambiguous to use the notation in Eqs. (4.47) and (4.61) for the limits of integration.[57]

---

[54] Smirnov pp. 6–7, p. 19.

[55] Roman Vol. 1 p. 367.

[56] Roman Vol. 1 p. 370. Care has to be taken when $g(\tau)$ is discontinuous at a and b.

[57] Kingman and Taylor p. 125. Pitt p. 31. When comparing a Lebesgue–Stieltjes integral with the corresponding Riemann–Stieltjes the discontinuities of $g(\tau)$ at a and b should be taken into account.

4.3 Probability Distributions and Measures

4.3.1 Probability Distribution Functions

We have introduced probability measures in §4.1.2 as a means of describing probability distributions on a continuous sample space. The idea is to assign probabilities to events rather than to individual outcomes.[58] Another method would be to make use of the concept of probability distribution functions introduced in Definition 3.6(1). We can define probability distribution functions in terms of distribution functions introduced in Definition 3.6(2).

Definition 4.3.1(1)

(1) *A distribution function is referred to as a* **probability distribution function** *when it is applied to describe a probability distribution.*

(2) *A probability distribution function is said to be*

(a) *discrete, denoted by* $\mathcal{F}_{\mathrm{d}}(\tau)$*, if it is piecewise-constant as shown in Fig. 3.1 with possibly countably infinite number of discontinuities,*

(b) *absolutely continuous, denoted by* $\mathcal{F}_{\mathrm{ac}}(\tau)$*, if it is an absolutely continuous function of* τ*, and*

(c) *singularly continuous, denoted by* $\mathcal{F}_{\mathrm{sc}}(\tau)$*, if it is a singularly continuous function of* τ*.*

(3) *If a probability distribution function* $\mathcal{F}(\tau)$ *is differentiable with respect to* τ*, then its derivative* $w(\tau) = d\mathcal{F}(\tau)/d\tau$ *is called a* **probability density function***.*

For an absolutely continuous probability distribution function $\mathcal{F}_{\mathrm{ac}}(\tau)$, a probability density function w exists almost everywhere. An absolutely continuous probability distribution function and its

[58] A general event contains many outcomes, e.g., it may contain a continuous set of outcomes. Such an event would corresponds to a Borel set Λ of $I\!R$. An elementary event which contains a single outcome τ_0 corresponds to the singleton set $\{\tau_0\}$. As seen in Eq. (4.87) the probability of an elementary event, i.e., an individual outcome, is zero for continuous sample spaces.

probability density function are related by Eq. (4.53), i.e.,

$$\mathcal{F}_{\mathrm{ac}}(\tau) = \int_{-\infty}^{\tau} w(\tau)\, d\tau. \tag{4.78}$$

It can be shown that any probability distribution function $\mathcal{F}(\tau)$ can be decomposed into a combination of a discrete one $\mathcal{F}_{\mathrm{d}}(\tau)$, an absolutely continuous one $\mathcal{F}_{\mathrm{ac}}(\tau)$ and a singularly continuous one $\mathcal{F}_{\mathrm{ac}}(\tau)$, i.e., we have[59]

$$\mathcal{F}(\tau) = c_1\mathcal{F}_{\mathrm{d}}(\tau) + c_2\mathcal{F}_{\mathrm{ac}}(\tau) + c_3\mathcal{F}_{\mathrm{sc}}(\tau), \tag{4.79}$$

where $c_1, c_2, c_3 \in [0, 1]$ and $c_1 + c_2 + c_3 = 1$. From now on we shall ignore singularly continuous probability distributions, since in the great majority of practical applications we do not encounter singular probability distribution functions. We shall call absolutely continuous probability distribution simply as a continuous probability distribution for short.

Continuous probability distribution functions are often given in terms of their density functions. For example, we have the *Gaussian distribution density function*[60]

$$w_{\mathrm{G}}(\tau) := \frac{1}{\sqrt{2\pi\, a^2}}\, e^{-(\tau-b)^2/2a^2}, \quad a, b \in \mathbb{R}. \tag{4.80}$$

Another example is the *uniform distribution* given by the density function

$$w_{\mathrm{U}}(\tau) := \begin{cases} (b-a)^{-1} & a \le \tau \le b \\ 0 & \tau \notin [a, b] \end{cases}. \tag{4.81}$$

The expectation value and uncertainty defined in §4.3.3 may not exist, i.e., they may be infinite. An example is the *Cauchy distribution density function*[61]

$$w_{\mathrm{C}}(\tau) := \frac{1}{\pi\,(1+\tau^2)}. \tag{4.82}$$

The following integrals

$$\int_{-\infty}^{0} \tau w_{\mathrm{C}}(\tau)\, d\tau \quad \text{and} \quad \int_{0}^{\infty} \tau w_{\mathrm{C}}(\tau)\, d\tau \tag{4.83}$$

[59] Parzen p. 174. Kingman and Taylor p. 294.
[60] Lipschutz pp. 106–107. Gauss (1777–1855) was a German mathematician.
[61] Kingman and Taylor p. 310. Cauchy (1789–1857) was a French mathematician and a pioneer of analysis.

diverge and it is not possible to give the integral of $\tau w_C(\tau)$ over the real line an unambiguous finite value.[62]

4.3.2 Distribution Functions and Measures

Probability distribution functions are related one-to-one to probability measures:[63] A given probability measure $\mathcal{M}_p$ generates a probability distribution function $\mathcal{F}$ by

$$\mathcal{F}(\tau) := \mathcal{M}_p\big((-\infty, \tau]\big). \tag{4.84}$$

Conversely, a given probability distribution function $\mathcal{F}(\tau)$ can generate a probability measure. This is done by realizing that $\mathcal{F}(\tau)$, being non-decreasing and right continuous, can generate a Lebesgue–Stieltjes measure $\mathcal{M}_{ls,\mathcal{F}}$ by Eqs. (4.16) and (4.17). The resulting measures possess the following properties:

P4.3.2(1) $\mathcal{M}_{ls,\mathcal{F}}$ is a probability measure since it has the characteristic feature of probability measures in Eq. (4.24), i.e.,

$$\mathcal{M}_{ls,\mathcal{F}}(I\!R) = \mathcal{F}(\infty) - \mathcal{F}(-\infty) = 1. \tag{4.85}$$

P4.3.2(2) For a singleton set $\{\tau_0\}$, we have

$$\mathcal{M}_{ls,\mathcal{F}}\big(\{\tau_0\}\big) = \mathcal{F}(\tau_0) - \mathcal{F}(\tau_0 - 0). \tag{4.86}$$

This yields the value of 0 if $\mathcal{F}$ is continuous at τ_0. It follows that for a continuous sample space the probability of an elementary event is zero, i.e.,

$$\mathcal{M}_{ls,\mathcal{F}}\big(\{\tau_0\}\big) = 0. \tag{4.87}$$

If the probability of an individual outcome is not zero the total probability, i.e., $\mathcal{M}_{ls,\mathcal{F}}\big(\{I\!R\}\big)$ would add up to be infinite. The physics of this mathematical result will be discussed at the end of this subsection.

[62] The odd integrand does not guarantee a zero integral here because of divergence of the integrals in Eq. (4.83) above.

[63] Kingman and Taylor p. 291. Papoulis pp. 92–98.

P4.3.2(3) For intervals the values of the measure are given by Eqs. (4.18) to (4.23), e.g.,

$$\mathcal{M}_{\mathrm{ls},\mathcal{F}}\big((\tau_1,\,\tau_2]\,\big) = \mathcal{F}(\tau_2) - \mathcal{F}(\tau_1). \tag{4.88}$$

P4.3.2(4) For a general Borel set Λ, we have, following Eq. (4.76),[64]

$$\mathcal{M}_{\mathrm{ls},\mathcal{F}}\big(\Lambda\big) = \int_{\Lambda} d\mathcal{M}_{\mathrm{ls},\mathcal{F}} = \int_{\Lambda} d\mathcal{F}(\tau). \tag{4.89}$$

P4.3.2(5) A general probability distribution function is of the form given in Eq. (4.79). The function can be discrete or continuous, i.e., the treatment in terms of probability distribution functions and measures apply both to discrete and continuous cases.[65]

Probability distribution functions and their associated measures can describe the probability distribution of a statistical experiment with the following interpretation:

(1) The value $\mathcal{F}(\tau)$ is the probability of occurrence of the event corresponding to the interval $(-\infty,\,\tau\,]$, i.e., the probability of an outcome lying in the range $(-\infty,\,\tau]$. In other words, $\mathcal{F}(\tau)$ is *the probability of an outcome being less than or equal to* τ.

(2) The value $\mathcal{M}_{\mathrm{ls},\mathcal{F}}\big((\tau_1,\,\tau_2]\big)$ is the probability of occurrence of the event corresponding to the interval $(\tau_1,\,\tau_2]$, i.e., the probability of an outcome lying in the interval $(\tau_1,\,\tau_2]$.

(3) For a continuous distribution function, the *probability of an individual outcome is zero*. We can appreciate the physics of such a result. Any physical measurement would incur errors. We may be able to render the measurement errors small enough to enable us to distinguish a discrete set of values. But we will not be able to single out a precise value in a continuum. This corresponds to having a zero probability of identifying an individual outcome in a continuous sample space.

[64] Kingman and Taylor p. 291.

[65] We ignore singularly continuous functions. A continuous function is meant to be absolutely continuous.

The mathematics required for a rigorous treatment of observables of a quantum system having a continuous set of values turns out to be complicated. In Part II on the mathematical framework for quantum mechanics, we shall present the mathematics necessary for the treatment of such observables in quantum mechanics.

4.3.3 Expectation Values and Uncertainties

Associated with a probability distribution function $\mathcal{F}$, we can define the expectation value, variance and uncertainty in terms of Lebesgue–Stieltjes integrals, provided the integrals exist.

(1) The *expectation value* $\mathcal{E}(\mathcal{F})$ is defined by

$$\mathcal{E}(\mathcal{F}) = \int_{\mathbb{R}} \tau \, d\mathcal{F}(\tau). \tag{4.90}$$

(2) The *variance* $\mathrm{Var}\,(\mathcal{F})$ is defined by

$$\mathrm{Var}\,(\mathcal{F}) = \int_{\mathbb{R}} \Big(\tau - \mathcal{E}(\mathcal{F})\Big)^2 \, d\mathcal{F}(\tau). \tag{4.91}$$

This can be evaluated in terms of

$$\mathrm{Var}\,(\mathcal{F}) = \int_{\mathbb{R}} \tau^2 \, d\mathcal{F}(\tau) - \mathcal{E}(\mathcal{F})^2. \tag{4.92}$$

(3) The *uncertainty* $\Delta(\mathcal{F})$ is defined to be the square root of the variance, i.e.,

$$\Delta(\mathcal{F}) = \sqrt{\mathrm{Var}\,(\mathcal{F})}. \tag{4.93}$$

For a continuous probability distribution function $\mathcal{F}_{\mathrm{ac}}$, the integrals for the expectation value and for the variance can be calculated in terms of its associated probability density function $w(\tau)$, i.e., we have

$$\mathcal{E}(\mathcal{F}_{\mathrm{ac}}) = \int_{\mathbb{R}} \tau \, w(\tau) \, d\tau. \tag{4.94}$$

$$\begin{aligned} \mathrm{Var}\,(\mathcal{F}_{\mathrm{ac}}) &= \int_{\mathbb{R}} \Big(\tau - \mathcal{E}(\mathcal{F}_{\mathrm{ac}})\Big)^2 \, w(\tau) \, d\tau \\ &= \int_{\mathbb{R}} \tau^2 \, w(\tau) \, d\tau - \mathcal{E}(\mathcal{F}_{\mathrm{ac}})^2, \end{aligned} \tag{4.95}$$

$$\Delta(\mathcal{F}_{\mathrm{ac}}) = \sqrt{\mathrm{Var}\,(\mathcal{F}_{\mathrm{ac}})}. \tag{4.96}$$

For a discrete probability distribution function $\mathcal{F}_{\mathrm{d}}$ given by Eq. (3.38) with a corresponding probability mass function $\wp$, the integrals for the expectation value and for the variance become a sum, i.e., we have[66]

$$\mathcal{E}(\mathcal{F}_{\mathrm{d}}) = \sum_{\ell=1}^{n} \tau_\ell \, \wp(a_\ell). \tag{4.97}$$

$$\begin{aligned} \mathrm{Var}\,(\mathcal{F}_{\mathrm{d}}) &= \sum_{\ell=1}^{n} \Big(\tau_\ell - \mathcal{E}(\mathcal{F}_{\mathrm{d}})\Big)^2 \, \wp(a_\ell) \\ &= \sum_{\ell=1}^{n} \tau_\ell^2 \, \wp(a_\ell) - \mathcal{E}(\mathcal{F}_{\mathrm{d}})^2, \end{aligned} \tag{4.98}$$

$$\Delta(\mathcal{F}_{\mathrm{d}}) = \sqrt{\mathrm{Var}\,(\mathcal{F}_{\mathrm{d}})}. \tag{4.99}$$

These results agree with those in §3.5.

For an arbitrary probability distribution function, the expectation value may not exist, i.e., the integral may diverge, and if it exists the variance may not exist. For physical applications we would restrict ourselves to probability distribution functions with finite expectation values and finite variances. We shall return to this important restriction in our discussion of quantum mechanics in Part III.

Exercises and Problems

Q4(1) For the characteristic function χ_Λ show that the inverse image of every Borel set is a Borel set.

Q4(2) Prove Eq. (4.12).

Q4(3) Show that the Lebesgue measure of the set of irrational numbers in $[a, b]$ is equal to $b - a$.

Q4(4) Prove Eqs. (4.18) to (4.23).

[66] This is due to the results in E4.2.3(4). In our applications, Lebesgue–Stieltjes integrals agree with Riemann–Stieltjes integrals in calculating expectation values, variances and uncertainties.

Q4(5) Explain the main differences between Riemann and Lebesgue integrals.

Q4(6) Explain why Lebesgue integrals are only defined for Borel functions.

Q4(7) Explain the main differences between Riemann–Stieltjes and Lebesgue–Stieltjes integrals.

Q4(8) Explain the meaning of Eq. (4.68) which expresses the Dirac delta function $\delta(\tau)$ formally as the derivative of the unit step function $g_{\mathrm{us}}(\tau)$

Q4(9) Show that the coefficients c_1, c_2, c_3 in Eq. (4.79) must satisfy the condition $c_1 + c_2 + c_3 = 1$.

Q4(10) Show that expectation value is additive in the sense that for the distribution function in Eq. (4.79) we have

$$\begin{aligned} &\mathcal{E}(c_1\mathcal{F}_{\mathrm{d}} + c_2\mathcal{F}_{\mathrm{ac}} + c_3\mathcal{F}_{\mathrm{sc}}) \\ &= c_1\mathcal{E}(\mathcal{F}_{\mathrm{d}}) + c_2\mathcal{E}(\mathcal{F}_{\mathrm{ac}}) + c_3\mathcal{E}(\mathcal{F}_{\mathrm{sc}}). \end{aligned} \tag{4.100}$$

Q4(11) Show that the Gaussian probability distribution function given by the density function in Eq. (4.80) by

$$\mathcal{F}_{\mathrm{G}}(\tau) = \int_{-\infty}^{\tau} w_{\mathrm{G}}(\tau)\, d\tau \tag{4.101}$$

satisfies the defining properties of probability distribution functions. Verify that the expectation value and the uncertainty are given respectively by

$$\mathcal{E}(\mathcal{F}_{\mathrm{G}}) = b \quad \text{and} \quad \Delta(\mathcal{F}_{\mathrm{G}}) = a. \tag{4.102}$$

Q4(12) Find the probability distribution function $\mathcal{F}_{\mathrm{U}}(\tau)$ given by the density function $w_{\mathrm{U}}(\tau)$ in Eq. (4.81).

Q4(13) Discuss the mathematical and physical differences between discrete and continuous probability distributions.

Chapter 5

Quantum Mechanical Systems

Quantum systems range from the traditional microscopic ones like electron and atoms to macroscopic ones like superconductors. There are also quantum fields. In this book, we shall study mainly traditional non-relativistic quantum systems. Some general properties of these systems are set out in this chapter.

5.1 Experimental Measurability

Although not exclusively so quantum systems are typically microscopic and beyond the direct reach of our senses. As a result, we do not have any intuition as to how they would behave. Sophisticated experiments have to be performed to study their behaviour. We learn from these experiments that quantum systems may behave very differently from classical systems. Our intuition gained from our knowledge of classical objects may or may not apply. So, we have to treat with great caution any argument based on intuition. This is a characteristic feature of modern physics. Special Relativity is full of examples of this.

Without intuition we have to rely on experiments. Meaningful things to talk about are those which can be experimentally

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

observed. The phenomenon of electron diffraction shows that even our classical intuition of particles and waves does not apply. Many intuitively defined quantities in classical mechanics become meaningless when applied to electrons. Take the example of the concept of *instantaneous speed* $v = dx/dt$ for a particle moving to the right along the x-axis. To ascertain if this is a meaningful quantity of an electron we have to see if the defining expression dx/dt is measurable. For a classical particle, the measurement of dx/dt can be obtained in the following manner:

(1) Measure position x at time t.
(2) Measure position $x + \Delta x_1$ at time $t + \Delta t_1$.
(3) Take the ratio $v_1 = \Delta x_1/\Delta t_1$.
(4) Prepare the particle in the same state in position x again. Repeat the measurements above with a smaller time interval $\Delta t_2 < \Delta t_1$, and a smaller displacement Δx_2 to obtain a ratio $v_2 = \Delta x_2/\Delta t_2$.
(5) Repeat the above procedure with an even smaller time interval Δt_3 and an even smaller displacement Δx_3 to obtain the ratio $v_3 = \Delta x_3/\Delta t_3$ and so on to obtain a sequence of such ratios v_n with decreasing time interval Δt_n.
(6) We say that the particle in the given state at time t has an instantaneous speed v if the sequence $v_n = \Delta x_n/\Delta t_n$ tends to a limiting value v as $\Delta t_n \to 0$ and we write $v = dx/dt$.

For a classical particle, such a sequence does converge to give a value for the instantaneous velocity with arbitrary accuracy.[1] For a quantum particle, the situation is entirely different. If we try to reduce Δx_n by reducing Δt_n we are in effect trying to impose a smaller spatial confinement of the particle. As we know from the uncertainty relation in quantum mechanics, this would give rise to an increase in uncertainty in the momentum. In other words, the motion of the particle becomes more erratic as one tries to reduce its spatial uncertainty. The result is that the ratio $v_n = \Delta x_n/\Delta t_n$ will be erratic and will not converge. In other words, we cannot measure the

[1] Even though we cannot actually obtain the limiting value in an experiment we can approach the limiting value with arbitrary accuracy.

instantaneous velocity. It is therefore not sensible to base a theory on such quantities.[2]

5.2 Observable Values

It is accepted that a quantum system, say an electron, has an objective existence. At the outset we assume that a measurement operation exists to produce a value of any given observable of the system. This assumption is built into the concept of observables which are physical quantities which can be measured to produce a value. An observable of a classical system in an arbitrary state possesses a value at any instance of time. If we wish to know what this value is, we can perform a measurement which will *reveal* the value of the observable at that time. A measurement can be executed within a short period of time without disturbing the system significantly so that the value revealed by a measurement is the same as the value the observable has *before, during and after* the measurement with arbitrary accuracy. We say that classical observables have *objective values* independent of measurement.

The situation for quantum systems is fundamentally different. There are two cases here:

Special case This is similar to the classical situation. The measured value of an observable of a quantum system in certain states is the value the observable possessed *before* the measurement. The measurement serves only to reveal the value.

General case A quantum observable may not have a value in an arbitrary state, i.e., a measured value does not tell us what value the observable possessed *before* the measurement. It does not even tell us whether the observable has a value at all *before* the measurement.[3] As an example consider an electron. The state of the election is describable by a wave function. The wave function can only give us the probability distribution of possible positions of the electron. It does not determine a definite position. In other

[2] Landau and Lifshitz Chapter 1 §1.
[3] Classical continuous systems discussed in §2.2 have a similar property.

words, the electron described by a wave function does not have a definite position. A position measurement will yield a position of the electron, but this position is not what the electron had before the measurement. A repetition of the measurement for the electron described by the same wave function may yield a different position.

Since a quantum particle in a general state may not have an *objective value* of a given observable independent of measurement it is useful to distinguish two kinds of values:

Possessed values For a given quantum system, it may well be that there is a state ϕ^s in which a given observable A has a definite value a.[4] We say that observable A *possesses* the value a in state ϕ^s, or state ϕ^s determines the value a of observable A. When an observable possesses a value in a given state this value can be revealed by measurement.

Greek letters are widely used in quantum mechanics. They are often used to mean different things. A Greek letter could mean a function, a vector or a state. For clarity we shall use a letter on its own or with an independent variable, e.g., ϕ or $\phi(\tau)$, to mean a function. We shall attach a superscript s to a letter, e.g., ϕ^s, to signify a state. Finally we shall attach an overhead arrow, i.e., $\vec{\phi}$, when it is used to denote a vector.

Measured values A value of an observable obtained by a measurement is called a *measured value*. If an observable possesses a value in a given state, then the measured value would coincide with the possessed value. When an observable does not possess a value in a given state a measurement would still produce a value. However, a repetition of the measurement in the same state would generally produce a different value. The measured value is obviously not the value of the observable before the measurement. Furthermore such a measurement would significantly disturb the state. The state after the measurement can be quite different from the initial state. An example is an electron's position mentioned earlier. Such a change of state presents a host of problems loosely known as *measurement problems* which is a source of great controversy in quantum theory.[5]

[4] There may be more than one state in which observable A has the value a.
[5] *See* §34.7 for a discussion of the measurement problem.

5.3 Observables, States and Probabilistic Behaviour

We shall discuss some general properties of observables and states of quantum systems in this section. These properties are summarised into four statements, i.e., QMP5.3(1) to QMP5.3(4) given below.

QMP5.3(1) *Not all observables are simultaneously measurable.*

This statement means that we cannot make simultaneous measurements in order to obtain a value for every observable of the system at the same time. In fact we cannot make simultaneous measurements of two arbitrary observables to obtain their values at the same time. For example, we cannot simultaneously measure the x and z components of the spin angular momentum of a spin-$\frac{1}{2}$ particle. Observables which can be simultaneously measured are said to be **compatible** and observables which cannot be simultaneously measured in any state are said to be **incompatible**.[6]

Starting from any chosen observable A, we may be able to find another observable A' which is compatible with A but independent of A. If we continue the process we will come to a maximum set $\mathcal{A}_{\text{cc}} = \{A, A', A'', \cdots\}$ of mutually compatible observables. Such a set is called a **complete set of compatible observables** of the system, or **a complete set of observables** for short, and it is not unique. For the spatial motion of a spinless particle the x, y, and z components of its linear momentum form a complete set of observables. An alternative set would be the x, y, and z components of its position vector. Starting with a different initial observable we may well end up with a different complete set.

A complete set can contain only one observable. Two familiar examples are

E5.3(1) For one-dimensional motion along the x-axis, e.g., a one-dimensional harmonic oscillator, the position x constitutes a complete set of observables.[7]

[6] *See* Definition 28.4.1(1) for a more detailed definition of these terms.
[7] The harmonic oscillator is discussed in detail in Chapter 35.

E5.3(2) For electron spin the z-component spin S_z constitutes a complete set of spin observables. The x component of spin S_x is *incompatible* with S_z, i.e., S_x and S_z cannot be measured simultaneously in any state. Once we have a measured value of S_z we cannot demand a simultaneous value of S_x. The y components of spin S_y is also incompatible with S_z.

Some observables possess only a discrete set of values. We call them **discrete observables**. In contrast, observables possessing a continuous set of values are called **continuous observables**. There are also observables having part continuous and part discrete values. Continuous observables are more complicated, both physically and mathematically. The discussion on Eq. (4.87) tells us that there is zero probability of a measurement of a continuous observable producing an exact value. Generally a measurement would produce a value in an interval. This would make it difficult to associate states with individual measured values. For simplicity and for definiteness we shall confine our discussion to discrete observables in what follows.

A set of simultaneously measured values $a, a', a'', \cdots$ of a complete set $\mathcal{A}_{\text{cc}}$ of discrete observables at any particular instant of time t contain a maximum amount of information about the system at t. We can use this information to characterise the **state** $\phi^s_{a,a',\ldots}$ of the system at that instant of time. In other words, when the system is in state $\phi^s_{a,a',\ldots}$ observables $A, A', A'', \cdots$ possess simultaneous values $a, a', a'', \cdots$. A measurement of A in that state would yield the value a, a measurement of A' in that state would yield the value a' and so on. We can re-phrase QMP5.3(1) into the statement that

> *there does not exist a state in which all observables can possess a value.*

On the other hand the nature of states as defined above does mean that there are states in which a given observable possesses a value. We can illustrate the situation with the case of electron spin. Suppose the z-component spin S_z is measured resulting in the value $\hbar/2$. The z-component spin being a complete set this value determines a state α^s_z for the electron spin. A value $-\hbar/2$ of S_z would determine a different state β^s_z. Conversely the electron in state

α_z^s possesses the value $\hbar/2$ of S_z and it would possesses the value $-\hbar/2$ of S_z in state β_z^s. We can summarise the above discussion in a statement.

QMP5.3(2)[8] *For a discrete observable, there are states in which the observable possesses a definite value, which can be revealed by measurement, and conversely each measured value would correspond to one or more states.*

While the state $\phi^s_{a,a',\ldots}$ determines the values of observables in a complete set $\mathcal{A}_{\text{cc}}$ it cannot determine the value of an observable which is incompatible with observables in the complete set $\mathcal{A}_{\text{cc}}$. To be specific consider a discrete observable B which is incompatible with observables in the complete set $\mathcal{A}_{\text{cc}}$. Taken on its own B has a set of values $b_1, b_2, \cdots$. However, it is impossible to have a simultaneous assignment of a set of values $a, a', a'', \cdots$ for the observables in $\mathcal{A}_{\text{cc}}$ and a value for B. This means that a state $\phi^s_{a,a',\ldots}$ determined by a set of values $a, a', a'', \cdots$ of $\mathcal{A}_{\text{cc}}$ cannot determine a definite value of B. Such a state is said to be **incompatible** with observable B and vice versa. In contrast the state $\phi^s_{a,a',\ldots}$ is said to be **compatible** with observables in $\mathcal{A}_{\text{cc}}$.

For electron spin the state α_z^s determined by the value $\hbar/2$ of S_z is compatible with S_z. Since S_z is incompatible with S_x the spin state α_z^s is incompatible with S_x, i.e., the state α_z^s does not determine the value of S_x. This means that a measurement of S_x in state α_z^s may yield the value $\hbar/2$ but a repetition of the measurement in the same state α_z^s may result in the value $-\hbar/2$.

Let us examine generally what can happen if we make a measurement of an observable B when the system is in state $\phi^s_{a,a',\ldots}$ incompatible with B:

(1) In a single measurement we will get a value, say b_1, of B. A repetition of the measurement with the system again in state $\phi^s_{a,a',\ldots}$ may well produce a different value, say b_2, since the

[8] Observables satisfying this property are said to be *sharp*. It is possible to extend the family of observables to include *unsharp* observables which do not satisfy this property (*see* Busch pp. 9–10, Wan p. 404). In standard quantum theory, all observables are assumed to be sharp.

state $\phi^S_{a,a',\ldots}$ cannot tell us that a measurement would definitely produce the value b_1.

(2) This unpredictability turns out to obey the probabilistic theory described in Chapters 3 and 4. The reason for this unpredictability is not due to a lack of knowledge of the system. It is a fundamental nature of quantum systems that we simply cannot predict with certainty the outcome of any individual measurement of an arbitrary observable in a given state. A state can only predict probabilistically the measured values of an arbitrary observables. We call such behaviour **intrinsically probabilistic**.

The above discussion is summarised in the following statement:

QMP5.3(3) *A state cannot determine with certainty the value of an arbitrary observable. A measurement would generate a value which may not be the value possessed by the observable before the measurement. However, a state can predict the probability distribution of the measured values of an arbitrary observable.*

We can use electron spin to illustrate the intrinsic probabilistic behaviour of quantum systems. The spin state α^S_z does not determine the value of S_x but it can determine the probability distribution of the values of S_x, i.e., there will be a probability of $1/2$ of a measured value of S_x to be $\hbar/2$ and a probability of $1/2$ of a measured value of S_x to be $-\hbar/2$.[9]

It is not all chaotic. As in a statistical experiment we can determine the expectation value of an arbitrary observable in a given state. The last statement in QMP5.3(3) is so characteristic of the properties of states that it can be reversed to produce a general definition of a quantum states as a probability measure on an appropriate set of basic observables known as *propositions* from which a formulation of quantum theory, known as *quantum logic* approach, can be established.[10]

While on the subject of states we should discuss briefly as to how one would prepare a state. Suppose we measure observables

[9] *See* Chapter 14 for more details.
[10] Mackey pp. 56–81, Jauch pp. 67–110. Beltrametti and Gassinelli Chapter 10.

A, A', A'', ... in a complete set $\mathcal{A}_{cc}$ and obtain a set of values a, a', a'', $\cdots$. This set of values would characterise a state $\phi^s_{a,a',\ldots}$. Intuitively we would hope that the state of the system immediately after the measurement would be $\phi^s_{a,a',\ldots}$, independent of the state before the measurement. As a result, the observables A, A', A'', ... would possess the values a, a', $a'' \cdots$ immediately after the measurement. In many cases this turns out to be true. Measurements satisfying this property are called **measurements of the first kind** or an **ideal measurements**.[11] For measurements of this kind we can conclude that an immediate repetition of the measurement would yield the same results. Measurement of the first kind can serve to prepare a quantum state, e.g., the spin state α^s_z can be prepared in this way. However, quantum measurements are not all of this kind, i.e., the state immediately after the measurement may not be the state corresponding to the measured values of the observables being measured. In the extreme case a quantum system, e.g., a photon, can even be destroyed after its detection.[12] Generally we do have to distinguish measurement process and state preparation process. A great deal of research has gone into various theories and models of measurement and state preparation. Here we shall assume that physical processes exist in principle to prepare a system in any desired states.[13]

Finally we should point out that quantum states obey what is known as the *superposition principle*:

QMP5.3(4) *It is possible to superpose two or more quantum states to produce a new state which contains more information than the individual constituent states taken separately can provide.*

For example, the expectation value in this new state of an arbitrary observable is not equal to the sum of the expectation values of the observable in each individual constituent state. Instead it is equal to the sum plus additional terms. These additional terms would depend on the relationship between different constituent

[11] Isham p. 134. Jauch p. 165.
[12] Jauch p. 164. Isham p. 134.
[13] *See* Wan §3.5 and §3.6 for model theories.

states and the observable. These terms are referred to *correlation* or *interference* terms.

5.4 Structure of Quantum Mechanics

The general structure of physical theories set out before should apply here. This means that there should be similarities to the structure of classical mechanics as well as differences:

1. *Basic mathematical framework*[14] The starting point is the choice an appropriate **state space** for a given quantum system. Following classical mechanics one expect the state space to be a vector space. However, we would expect this vector space to be different from that for a classical mechanical system.

2. *Description of states* Quantum states are described by elements of the state space, i.e., by vectors since the state space is a vector space. These vectors can be added to produce new vectors. This may provide the combination of states required by the superposition principle.

3. *Description of observables* Unlike classical mechanics observables of a quantum system cannot all be described by real-valued functions of the state. If this were the case a state would determine the values of all the observables, contradicting fundamental properties QMP5.3(1) and QMP5.3(3) of quantum systems stated in the preceding section. We have to explore all the mathematical quantities associated with the state space to find something suitable. The quantities chosen to represent observables must give us

(1) the correct values of the observables, and
(2) together with the state these quantities must be able to produce the probability distributions of measured values of all observables in accordance with QMP5.3(3).

[14] In the quantum logic approach we would start with a mathematical structure of a suitably chosen set of observables rather than starting with a mathematical structure for the state space.

4. *Dynamics* We expect the time evolution of a quantum system to involve the time dependence of the state of the system. Since observables are no longer functions of the state we may well have to examine possible time dependence of observables separately from that of the state.

Exercises and Problems

Q5(1) Explain why the values of a classical observable are deemed to be objective while the values of a quantum observable are generally regarded as non-objective.

Q5(2) Give a brief account of the relationship between states and possessed values of discrete observables of a quantum system.

Q5(3) Explain why quantum observables cannot be related to the state in the same way kinematic observables of a classical system are related to the state.

Q5(4) Discuss the effect of QMP5.3(1) on the specification of states.

Q5(5) Explain what is meant by the behaviour of quantum systems being intrinsically probabilistic.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Section II

Mathematical Framework

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 6

Three-Dimensional Real Vectors

Vectors are employed extensively in physics. Conceptually we all know what a vector is: a *vector* is a quantity which has a *length* (*magnitude* or *norm*) as well as a *direction*. Vectors are denoted by letters with an overhead arrow, e.g., $\vec{u}$, $\vec{v}$, $\vec{w}$. We can tell in practice which quantity is a vector and which is not. In contrast to vectors, our familiar real or complex numbers are called *scalars*.

We know the usefulness of vectors. It is hard to imagine how we would formulate classical mechanics and electromagnetism without vectors. Here we are talking about three-dimensional vectors which corresponds to the three-dimensional physical space we live in. These vectors have three real components along the x, y and z axes of a Cartesian coordinate system and they form the basis of Euclidean geometry.[1] We shall denote the set of all these three-dimensional vectors by $\vec{I\!E}^3$. The concept of three-dimensional vectors can be generalised to produce higher dimensional vectors to formulate modern quantum physics. We want these higher dimensional vectors to possess many of the properties of three-dimensional vectors. So, we shall start with a summary of the properties of three-dimensional vectors in this chapter. We shall study

[1] Euclid (300 BC) was a Greek mathematician in ancient times, well known for his work on geometry.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

operations on these vectors and their mathematical representations as operators in Chapters 8 and 9. This study is preceded by a chapter on matrices which provides the intuition for the various concepts and operations of vectors. In Chapter 10 we shall examine how it is possible to generate probability distributions in terms of vectors and operators in order to gain an insight into the mathematical framework for the formulation of quantum mechanics.

We shall summarise the properties of three-dimensional vectors in §6.1, §6.2, §6.3 and §6.4. The discussion is organised in a way that can be directly generalised to real and complex vectors in higher as well as lower dimensions. Chapters 11 to 12 will be devoted to a study of these generalisations.

6.1 Properties 1: Algebraic Properties

6.1.1 Addition

The sum $\vec{u} + \vec{v}$ of every pair of vectors $\vec{u}$, $\vec{v}$ in $\vec{I\!E}^3$ is defined by the parallelogram law to be another vector in $\vec{I\!E}^3$. This addition operation possesses the following properties:

A6.1.1(1) *Commutative*	$\vec{u} + \vec{v} = \vec{v} + \vec{u}$.
A6.1.1(2) *Associative*	$(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
A6.1.1(3) *Zero vector*	There is a unique vector in $\vec{I\!E}^3$, the zero vector $\vec{0}$, such that $\vec{u} + \vec{0} = \vec{u} \quad \forall \vec{u} \in \vec{I\!E}^3$.
A6.1.1(4) *Inverse*	Each $\vec{u} \in I\!E^3$ possesses a unique inverse $\vec{u}^{-1}$ such that $\vec{u} + \vec{u}^{-1} = \vec{0}$.

The zero vector $\vec{0}$ is often denoted by 0 without an overhead arrow.

6.1.2 Scalar Multiplication

For all $\vec{u}, \vec{v} \in \vec{I\!E}^3$ and $a, b \in I\!R$ the products $a\vec{v}$ and $b\vec{v}$ are defined as vectors in $\vec{I\!E}^3$. This scalar multiplication possesses the following properties:

SM6.1.2(1) *Distributive*	$a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$, $(a + b)\vec{u} = a\vec{u} + b\vec{u}$.
SM6.1.2(2) *Associative*	$(ab)\vec{u} = a(b\vec{v})$.
SM6.1.2(3) *Multiplication by* 1	$1\,\vec{u} = \vec{u}$.

The following results follow:

$$\vec{v}_1 + \vec{u} = \vec{v}_2 + \vec{u} \quad \Rightarrow \quad \vec{v}_1 = \vec{v}_2\,; \tag{6.1}$$

$$0\,\vec{u} = \vec{0}, \qquad a\,\vec{0} = \vec{0}, \tag{6.2}$$

$$\vec{u}^{\,-1} = (-1)\,\vec{u}. \tag{6.3}$$

We can equate the inverse $\vec{u}^{\,-1}$ with $-\vec{u}$, i.e., we have

$$\vec{u} - \vec{u} = \vec{u} + (-1)\vec{u} = \vec{u} + \vec{u}^{\,-1} = \vec{0}. \tag{6.4}$$

The set of vectors in $\vec{I\!E}^3$ is said to constitute a **vector space**, i.e., $\vec{I\!E}^3$ is a vector space. Generally a set of quantities endowed with the algebraic properties given in §6.1.1 and §6.1.2 is said to form a *vector space*.[2] Since only real numbers are involved the vector space is said to be *real*, i.e., $\vec{I\!E}^3$ is a *real vector space*.

6.2 Properties 2: Dimensions and Bases

Any number of vectors $\vec{u}_\ell$, $\ell = 1, 2, \cdots, n$, can be added up to produce new vectors. A sum of the form

$$\sum_{\ell=1}^{n} c_\ell \vec{u}_\ell = c_1\,\vec{u}_1 + c_2\,\vec{u}_2 + \cdots + c_n\,\vec{u}_n, \quad c_\ell \in I\!R \tag{6.5}$$

is called a *linear combination of the set of vectors* $\vec{u}_\ell$ *with coefficients* c_ℓ.

6.2.1 Linear Dependence and Independence

Definition 6.2.1(1) *A finite set* $\{\vec{u}_\ell : \ell = 1, 2, \cdots, n\}$ *of n vectors in* $\vec{I\!E}^3$ *is said to be*

(1) *linearly dependent if there exists a set of scalars* $c_\ell \in I\!R$*, not all zero, such that the linear combination of* $\vec{u}_\ell$ *with coefficients* c_ℓ *is zero*, i.e.,

$$\sum_{\ell=1}^{n} c_\ell\,\vec{u}_\ell = \vec{0}, \tag{6.6}$$

[2] As remarked before a "space" generally means a set endowed with certain properties. It should not be confused with the *physical space* we live in.

(2) *linearly independent if no such set of coefficients exists, i.e., if*

$$\sum_{\ell=1}^{n} c_\ell \, \vec{u}_\ell = \vec{0} \quad \Rightarrow \quad c_\ell = \vec{0} \;\; \forall \ell. \tag{6.7}$$

If a set is linear dependent, then a member of the set may be expressed as a linear combination of the others of the set. To appreciate this let us suppose $c_1 \neq 0$ in Eq. (6.6). Then

$$c_1 \vec{u}_1 + \sum_{\ell=2}^{n} c_\ell \vec{u}_\ell = \vec{0} \quad \Rightarrow \quad \vec{u}_1 = \sum_{\ell=2}^{n} c'_\ell \vec{u}_\ell, \quad c'_\ell = -\frac{c_\ell}{c_1}. \tag{6.8}$$

The set is linear independent if no member of the set can be expressed as a linear combination of the others in the set.

As examples we can see that three vectors all lying in the x-y plane are linear dependent while a vector along the z-axis is linear independent to vectors lying in the x-y plane.

A very important property of a linear independent set of vectors is given by the following theorem.

Theorem 6.2.1(1) *If a vector $\vec{v}$ is given as a linear combination of a linearly independent set of vectors $\vec{u}_\ell$, i.e.,*

$$\vec{v} = \sum_{\ell=1}^{n} c_\ell \, \vec{u}_\ell, \quad c_\ell \in I\!R, \tag{6.9}$$

then the set of coefficients $\{c_\ell\}$ is uniquely related to $\vec{v}$.

6.2.2 Dimensions, Bases and Complete Sets

Definition 6.2.2(1) *The dimension of $\vec{I\!E}^3$ is the maximum number of linearly independent vectors which can be found in $\vec{I\!E}^3$.*

The number turns out to be 3 and this is why vectors in $\vec{I\!E}^3$ are called three-dimensional. We also call $\vec{I\!E}^3$ a *three-dimensional real vector space*. We can choose three vectors aligned along the positive directions of the x, y and the z axes of a chosen Cartesian coordinate system to form a linear independent set. To make things simpler we can choose these three vectors to have a unit length. A vector is called a *unit vector* if it has a unit length. Unit vectors can be employed to specify a direction. For example, the three unit vectors along the x, y, z axes, shown in the diagram below and denoted

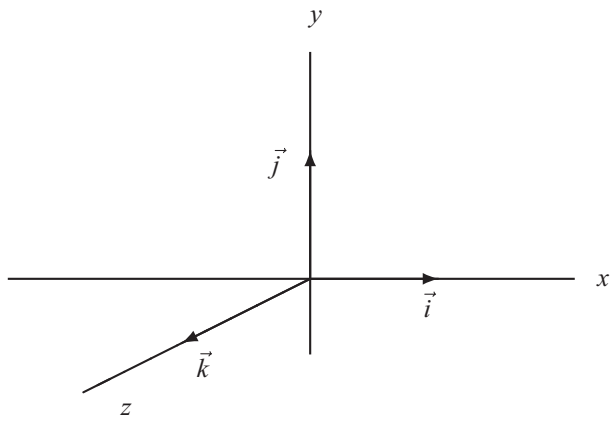


Figure 6.1 Unit vectors along x, y and z axes.

by $\vec{i},\ \vec{j},\ \vec{k}$, specify the positive directions along the three coordinate axes.

These three unit vectors $\vec{i},\ \vec{j},\ \vec{k}$ form a maximum set of linear independent vectors. Let $\vec{v}$ be an arbitrary vector different from the three unit vectors. Then the set $\vec{v}, \vec{i},\ \vec{j}, \vec{k}$ of four vectors is a linearly dependent set, e.g., the vector $\vec{v}$ is dependent on the three unit vectors. It follows that $\vec{v}$ can be written as

$$\vec{v} = v_x\vec{i} + v_y\vec{j} + v_z\vec{k}, \tag{6.10}$$

where the coefficients v_x, v_y, v_z are uniquely related to $\vec{v}$. As an example the vector $\vec{x}$ which specifies the position of a particle, known as the *position vector* of the particle, is often written as

$$\vec{x} = x\vec{i} + y\vec{j} + z\vec{k}. \tag{6.11}$$

For easy visualisation we shall always assume that vectors, e.g., the position vector, would extend from the coordinate origin outward, unless otherwise is stated.

The importance of a set of three linearly independent vectors $\vec{e}_\ell,\ \ell = 1, 2, 3$, in $\vec{I\!E}^3$ is that the set $\{\vec{e}_\ell\}$ **spans** $\vec{I\!E}^3$ in the sense that *every vector* $\vec{v}$ *in* $\vec{I\!E}^3$ *may be expressed uniquely as a linear*

combination of $\vec{e}_\ell$, i.e.,

$$\vec{v} = \sum_{\ell=1}^{3} c_\ell \vec{e}_\ell. \tag{6.12}$$

This leads to the notion of *basis, basis vectors* and *complete sets*.

Definition 6.2.2(2) *A basis for or a complete set in* $\vec{I\!E}^3$ *is a set of linearly independent vectors which spans* $\vec{I\!E}^3$*. Vectors in the basis are called basis vectors.*

A basis in $\vec{I\!E}^3$ must consists of three linearly independent vectors. An example is the three unit vectors $\vec{i}$, $\vec{j}$, $\vec{k}$. There are many different bases in $\vec{I\!E}^3$, e.g., if we rotate a given Cartesian coordinate axes we obtain three new coordinate axes and the unit vectors along these new axes will form a linearly independent set of unit vectors.

6.3 Properties 3: Scalar Product

6.3.1 Scalar Product

There are two traditional ways to multiply two vectors in $\vec{I\!E}^3$:

(1) We can multiply two vectors $\vec{u}$ and $\vec{v}$ to produce a new vector. The new vector, denoted by $\vec{u} \times \vec{v}$, is called the *vector product* of $\vec{u}$ and $\vec{v}$. An example is given in Eq. (2.3) which is defined by Eq. (2.7).
(2) We can multiply $\vec{u}$ and $\vec{v}$ to produce a scalar. The scalar, denoted by $\vec{u} \cdot \vec{v}$, is called the *scalar product* of $\vec{u}$ and $\vec{v}$.

We are interested in the scalar product here. For later convenience we shall adopt the *Dirac notation* for the scalar product, i.e., we shall denote the scalar product of $\vec{u}$ and $\vec{v}$ by $\langle \vec{u} \mid \vec{v} \rangle$.[3] The actual value of the scalar product is explicitly defined to be $uv \cos\theta$, where u and v are respectively, the lengths of $\vec{u}$ and $\vec{v}$ and θ is the angle between the two vectors, i.e.,

$$\langle \vec{u} \mid \vec{v} \rangle := uv \cos\theta, \tag{6.13}$$

which can be positive or negative depending on the angle θ.

[3] Here we do not attach any separate meaning to the symbols $\langle \vec{u} \mid$ and $\mid \vec{v} \rangle$.

The scalar product is an assignment of a real number to every pair of vectors $\vec{u}$ and $\vec{v}$ in $\vec{I\!E}^3$ satisfying the following properties[4]:

SP6.3.1(1) *Commutative* $\langle \vec{u} \mid \vec{v} \rangle = \langle \vec{v} \mid \vec{u} \rangle$.

SP6.3.1(2) *Distributive* $\langle \vec{u} \mid (a_1 \vec{v}_1 + a_2 \vec{v}_2) \rangle = a_1 \langle \vec{u} \mid \vec{v}_1 \rangle + a_2 \langle \vec{u} \mid \vec{v}_2 \rangle$.

SP6.3.1(3) *Positive definite* $\langle \vec{u} \mid \vec{u} \rangle \geq 0$, $\langle \vec{u} \mid \vec{u} \rangle = 0 \Rightarrow \vec{u} = \vec{0}$.

The following examples are of particular interest:

E6.3.1(1) *Perpendicular vectors* The scalar product of two perpendicular vectors, i.e., when $\theta = \pi/2$, is zero. For example, we have

$$\langle \vec{i} \mid \vec{j} \rangle = \langle \vec{i} \mid \vec{k} \rangle = \langle \vec{j} \mid \vec{k} \rangle = 0. \tag{6.14}$$

E6.3.1(2) *Parallel vectors* The scalar product of two parallel vectors, i.e., when $\theta = 0$, is equal to the product of their lengths, i.e., $\langle \vec{u} \mid \vec{v} \rangle = uv$. An example is the scalar product of a vector with itself, i.e., $\langle \vec{v} \mid \vec{v} \rangle = v^2$. In particular we have

$$\langle \vec{i} \mid \vec{i} \rangle = \langle \vec{j} \mid \vec{j} \rangle = \langle \vec{k} \mid \vec{k} \rangle = 1. \tag{6.15}$$

6.3.2 Orthonormality

Being numerical values scalar product is tangible. It enables us to introduce a number of very useful quantities:

(1) The *norm* $\|\vec{v}\|$ of a vector $\vec{v}$ is defined in terms of scalar product to be

$$\|\vec{v}\| := \sqrt{\langle \vec{v} \mid \vec{v} \rangle}. \tag{6.16}$$

This agrees with our usual notion of length or magnitude of a vector in $\vec{I\!E}^3$, i.e., we have $||\vec{v}|| = v$. This formal definition in terms of scalar product is important when we discuss higher dimensional vectors. We shall use the term *norm* in preference to the term *length* from now on.

[4] Here $a, b \in I\!R$.

(2) A vector is said to be *normalised* if its norm is equal to 1. A normalised vector is previously referred to as a *unit vector*.

(3) A vector $\vec{v}$ has associated with it a unit vector $\vec{v}^{\,(u)}$, i.e.,

$$\vec{v}^{\,(u)} := \vec{v}/||\vec{v}\,||, \tag{6.17}$$

which signifies the direction of the vector. We shall call $\vec{v}^{\,(u)}$ the *unit directional vector* of $\vec{v}$.

(4) Two vectors $\vec{v}_1, \vec{v}_2$ are said to be *orthogonal* (perpendicular) if their scalar product vanishes. They are *orthonormal* if they are also normalised, i.e., $\vec{v}_1, \vec{v}_2$ are orthonormal if [5]

$$\langle \vec{v}_1 \mid \vec{v}_2 \rangle = \delta_{\ell\ell'}, \quad \ell, \ell' = 1, 2. \tag{6.18}$$

(5) Two orthogonal vectors $\vec{v}_1, \vec{v}_2$ are linearly independent. A set of three mutually orthogonal vectors can serve as a basis for the $\vec{I\!E}^3$.

6.3.3 Orthonormal Bases and Complete Sets

Definition 6.3.3(1) *A basis is an orthonormal basis if the basis vectors are orthonormal. An orthonormal basis is also referred to as a complete orthonormal set.*

The three unit vectors $\vec{i}$, $\vec{j}$, $\vec{k}$ along x, y, z axes constitute an orthonormal basis for $\vec{I\!E}^3$. There are many advantages in employing orthonormal basis, rather than an arbitrary basis with non-orthonormal basis vectors. One such advantage is in the calculation of the coefficients for the expression of an arbitrary vector $\vec{v}$ in an orthonormal basis $\{\vec{e}_\ell\}$, i.e., we can prove the following result:

$$\vec{v} = \sum_{\ell=1}^{3} v_\ell \, \vec{e}_\ell, \quad v_\ell = \langle \vec{e}_\ell \mid \vec{v} \, \rangle. \tag{6.19}$$

We call the coefficients v_ℓ the *components* of $\vec{v}$ on the basis vectors $\vec{e}_\ell$, e.g., we call v_1 the *component of* $\vec{v}$ *on* $\vec{e}_1$. These components are numerical values which can be positive or negative. Further advantages can be seen in the *Pythagoras theorem* discussed in the next section.

[5] Here $\delta_{\ell\ell'}$ is the Kronecker delta.

6.3.4 Pythagoras Theorem

We can calculate the norm of a vector and the scalar product of two vectors in terms of their components on the basis vectors of an orthonormal basis. Take the orthonormal basis $\{\vec{i},\ \vec{j},\ \vec{k}\}$ for example. When a vector $\vec{v}$ is expressed as a linear combination of $\{\vec{i},\ \vec{j},\ \vec{k}\}$ as in Eq. (6.10) the components are given, in accordance with Eq. (6.19), by

$$v_x = \langle\, \vec{i} \mid \vec{v}\, \rangle, \quad v_y = \langle\, \vec{j} \mid \vec{v}\, \rangle, \quad v_z = \langle \vec{k} \mid \vec{v}\, \rangle. \tag{6.20}$$

Using the orthonormality properties of the basis vectors $\vec{i},\ \vec{j},\ \vec{k}$ shown in Eqs. (6.14) and (6.15) we get

$$\begin{aligned} \langle \vec{v} \mid \vec{v}\, \rangle &= \langle (v_x\vec{i} + v_y\vec{j} + v_z\vec{k}) \mid (v_x\vec{i} + v_y\vec{j} + v_z\vec{k}) \rangle \\ &= v_x^2 + v_y^2 + v_z^2. \end{aligned} \tag{6.21}$$

The norm $||\vec{v}\,||$ of $\vec{v}$ is given in terms of its components by

$$||\vec{v}\,|| = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{6.22}$$

This result is known in geometry as the *Pythagoras theorem*. Furthermore, the scalar product of two vectors can also be calculated in terms of its components, i.e., we have

$$\begin{aligned} \langle \vec{u} \mid \vec{v}\, \rangle &= \langle (u_x\vec{i} + u_y\vec{j} + u_z\vec{k}) \mid (v_x\vec{i} + v_y\vec{j} + v_z\vec{k}) \rangle \\ &= u_x v_x + u_y v_y + u_z v_z. \end{aligned} \tag{6.23}$$

Similar results apply in a general orthonormal basis, i.e., when expressing any two vectors $\vec{u}$ and $\vec{v}$ in an orthonormal basis $\{\vec{e}_\ell\}$ we have

$$\vec{u} = \sum_{\ell=1}^{3} u_\ell \vec{e}_\ell, \quad u_\ell = \langle \vec{e}_\ell \mid \vec{u}\, \rangle, \tag{6.24}$$

$$\vec{v} = \sum_{\ell=1}^{3} v_\ell\, \vec{e}_\ell, \quad v_\ell = \langle \vec{e}_\ell \mid \vec{v}\, \rangle, \tag{6.25}$$

$$||\vec{v}\,|| = \left(\sum_{\ell=1}^{3} v_\ell^2 \right)^{1/2} = \sqrt{v_1^2 + v_2^2 + v_3^2}, \tag{6.26}$$

$$\langle \vec{u} \mid \vec{v}\, \rangle = \sum_{\ell=1}^{3} u_\ell v_\ell = u_1 v_1 + u_2 v_2 + u_3 v_3, \tag{6.27}$$

It follows that

$$\vec{u},\ \vec{v} \text{ are orthogonal} \quad \Leftrightarrow \quad u_1v_1 + u_2v_2 + u_3v_3 = 0. \tag{6.28}$$

We shall refer to the result in Eq. (6.27) also as *Pythagoras theorem* for easy reference later. The equation can be written explicitly in terms of the basis vectors as

$$\langle \vec{u} \mid \vec{v} \rangle = \sum_{\ell=1}^{3} u_\ell v_\ell = \sum_{\ell=1}^{3} \langle \vec{u} \mid \vec{e}_\ell \rangle \langle \vec{e}_\ell \mid \vec{v} \rangle. \tag{6.29}$$

Equations (6.21) and (6.23) become special cases of Eq. (6.27).

The following results are easily verified:

(1) No vector, apart from the zero vector, can be orthogonal to all basis vectors $\vec{e}_\ell$, i.e.,

$$\langle \vec{e}_\ell \mid \vec{v} \rangle = 0 \ \ \forall \ell \quad \Rightarrow \quad \vec{v} = \vec{0}. \tag{6.30}$$

It follows from this that

$$\langle \vec{e}_\ell \mid \vec{u} \rangle = \langle \vec{e}_\ell \mid \vec{v} \rangle \ \ \forall \ell \quad \Rightarrow \quad \vec{u} = \vec{v}. \tag{6.31}$$

(2) The components of a normalised vector $\vec{u}$ in any orthonormal basis satisfies the following *normalisation condition*:

$$u_1^2 + u_2^2 + u_3^2 = 1. \tag{6.32}$$

(3) Any vector $\vec{v}$ can be *normalised* by a *normalisation constant* c in accordance with Eq. (6.17), i.e., $\vec{v}^{\,(u)} = c\,\vec{v}$ is normalised with the normalisation constant c given by

$$c = \frac{1}{||\vec{v}\,||} = \frac{1}{\sqrt{(v_1^2 + v_2^2 + v_3^2)}}. \tag{6.33}$$

6.3.5 Gram-Schmidt Orthogonalisation

Let $\vec{v}_\ell$, $\ell = 1, 2, 3$, be a set of linearly independent vectors, not necessarily orthonormal. We can construct an orthonormal set of vectors $\vec{u}_\ell$ in terms of $\vec{v}_\ell$ by the Gram-Schmidt orthogonalisation process[6]:

[6] Gram (1850–1916) was a Danish mathematician and Schmidt (1876–1959) was an Estonian mathematician.

(1) Introducing a set of vectors $\vec{u}_{\ell'}$ by

$$\vec{u}\,'_1 := \vec{v}_1, \tag{6.34}$$

$$\vec{u}\,'_2 := \vec{v}_2 - \frac{\langle \vec{u}\,'_1 \mid \vec{v}_2 \rangle}{\langle \vec{u}\,'_1 \mid \vec{u}\,'_1 \rangle}\, \vec{u}\,'_1, \tag{6.35}$$

$$\vec{u}\,'_3 := \vec{v}_3 - \frac{\langle \vec{u}\,'_1 \mid \vec{v}_3 \rangle}{\langle \vec{u}\,'_1 \mid \vec{u}\,'_1 \rangle}\, \vec{u}\,'_1 - \frac{\langle \vec{u}\,'_2 \mid \vec{v}_3 \rangle}{\langle \vec{u}\,'_2 \mid \vec{u}\,'_2 \rangle}\, \vec{u}\,'_2. \tag{6.36}$$

These vectors are orthogonal, i.e., $\langle \vec{u}\,'_\ell \mid \vec{u}\,'_{\ell'} \rangle = 0 \quad \text{if} \quad \ell \neq \ell'$.

(2) Normalizing $\vec{u}_{\ell'}$ to arrive at $\vec{u}_\ell = \vec{u}\,'_\ell / \left\| \vec{u}\,'_{\ell'} \right\|$ which can be used to form an orthonormal basis for $\vec{I\!E}^3$.

6.3.6 Inequalities on Scalar Product and Norm

The following inequalities familiar in geometry can be established:

1. *The Schwarz Inequality*[7] The scalar product is less then or at most equal to the product of the norms:

$$|\langle \vec{u} \mid \vec{v} \rangle| = |uv \cos\theta| \leq uv = \|\vec{u}\| \, ||\vec{v}||. \tag{6.37}$$

2. *Triangle Inequalities* In a triangle the sum of the lengths of any two sides is greater than the length of the third side and the difference of the lengths of any two sides is less than the length of the third side. In terms of vectors we have

$$||\vec{u} + \vec{v}|| \leq \|\vec{u}\| + ||\vec{v}||. \tag{6.38}$$

$$\big|\, ||\vec{u}|| - ||\vec{v}|| \,\big| \leq \|\vec{u} - \vec{v}\|. \tag{6.39}$$

6.4 Properties 4: Scalar Product and Projections

The scalar product defined by Eq. (6.13) has a clear geometric interpretation in terms of an intuitive concept of *projection* of a vector onto another vector as we shall now discuss.

6.4.1 Projection onto $\vec{i}$

A vector lying in the x-y plane is expressible as

$$\vec{v} = v_x \vec{i} + v_y \vec{j}, \quad v_x = \langle \vec{i} \mid \vec{v} \rangle, \;\; v_y = \langle \vec{j} \mid \vec{v} \rangle. \tag{6.40}$$

[7] Schwarz (1843–1921) was a German mathematician.

The number v_x is the component of $\vec{v}$ on the basis vector $\vec{i}$.

Definition 6.4.1(1) *The projection of the vector $\vec{v}$ onto $\vec{i}$ is defined to be the vector $\vec{v}_{\vec{i}}$ given by*

$$\vec{v}_{\vec{i}} := v_x \vec{i}. \tag{6.41}$$

Figure 6.2 below serves to illustrate the projection of vectors.

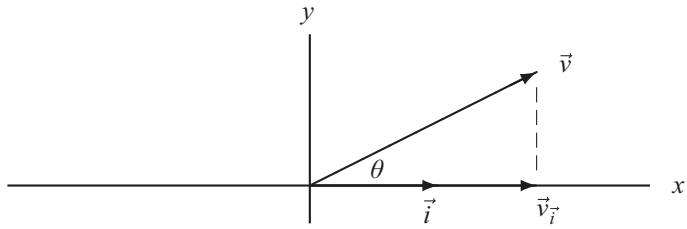


Figure 6.2 A vector and its projection.

It must be emphasised that while the component of $\vec{v}$ on $\vec{i}$ is a number, hence the notation v_x, the projection of $\vec{v}$ onto $\vec{i}$ is a vector, and hence the notation $\vec{v}_{\vec{i}}$. The projection of $\vec{v}$ onto $\vec{i}$ is also known as the *projection of $\vec{v}$ onto the x-axis*. Since the component v_x is equal to the scalar product of $\vec{i}$ and $\vec{v}$ the projection can be written down explicitly as

$$\vec{v}_{\vec{i}} = \langle \vec{i} \mid \vec{v} \rangle \vec{i}. \tag{6.42}$$

The projections of two vectors $\vec{v}_1$ and $\vec{v}_2$ onto $\vec{i}$ are shown in Fig. 6.3 below. These projections have the following properties:

P6.4.1(1) For a vector $\vec{v}_1$ in the first quadrant of the x-y plane we have

$$\vec{v}_1 = v_{1x}\vec{i} + v_{1y}\vec{j}, \quad \text{where} \quad v_{1x},\ v_{1y} > 0. \tag{6.43}$$

Its projection onto $\vec{i}$ is $\vec{v}_{1\vec{i}} = v_{1x}\vec{i} = \langle \vec{i} \mid \vec{v}_1 \rangle \vec{i}$.

P6.4.1(2) For a vector $\vec{v}_2$ in the second quadrant we have

$$\vec{v}_2 = v_{2x}\vec{i} + v_{2y}\vec{j} \quad v_{2x} < 0,\ \ v_{2y} > 0. \tag{6.44}$$

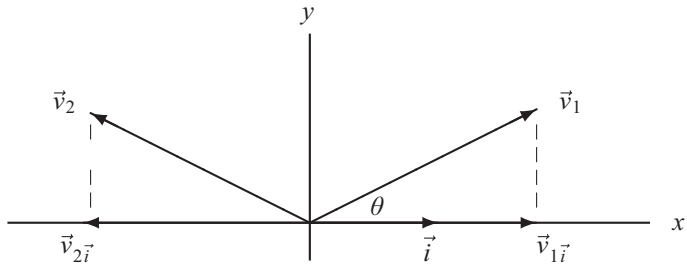


Figure 6.3 Two vectors and their projections onto $\vec{i}$.

Its projection onto $\vec{i}$ is $\vec{v}_{2\vec{i}} = v_{2x}\,\vec{i} = \langle \vec{i} \mid \vec{v}_2 \rangle\, \vec{i}$. The angle between $\vec{v}_2$ and $\vec{i}$ is bigger than $\pi/2$ so that the scalar product $\langle \vec{i} \mid \vec{v}_2 \rangle$ is negative. As a result the projection of $\vec{v}_2$ onto $\vec{i}$ is pointing in the negative direction, as shown in Fig. 6.3.

6.4.2 Projection onto Arbitrary Unit Vector $\vec{e}$

When written in the form of Eq. (6.10) a vector $\vec{v}$ has components v_x, v_y and v_z given by Eq. (6.20). We can define its projections onto $\vec{i}$, $\vec{j}$ and $\vec{k}$ by

$$\vec{v}_{\vec{i}} := \langle \vec{i} \mid \vec{v} \rangle\, \vec{i}, \quad \vec{v}_{\vec{j}} := \langle \vec{j} \mid \vec{v} \rangle\, \vec{j}, \quad \vec{v}_{\vec{k}} := \langle \vec{k} \mid \vec{v} \rangle\, \vec{k}. \tag{6.45}$$

All this can be extended to an arbitrary orthonormal basis $\{\vec{e}_\ell\}$. Based on Eq. (6.19) we can define the *projection* of $\vec{v}$ onto the unit basis vector $\vec{e}_\ell$ as

$$\vec{v}_{\vec{e}_\ell} := v_\ell\, \vec{e}_\ell, \quad v_\ell = \langle \vec{e}_\ell \mid \vec{v} \rangle. \tag{6.46}$$

An orthonormal basis is not necessary. As the figure below shows we can project any vector $\vec{v}$ onto an arbitrary unit vector $\vec{e}$.

Definition 6.4.2(1) *The projection of $\vec{v}$ onto a unit vector $\vec{e}$ is defined to be the vector*

$$\vec{v}_{\vec{e}} := \langle \vec{e} \mid \vec{v} \rangle\, \vec{e}. \tag{6.47}$$

The projection $\vec{v}_{\vec{e}}$ has a norm equal to $|\langle \vec{e} \mid \vec{v} \rangle| = |v \cos\theta|$ and is directed along $\vec{e}$.

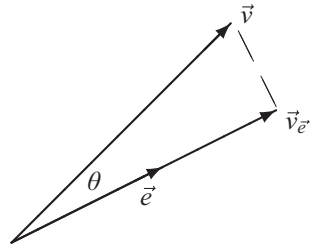


Figure 6.4 Projection onto $\vec{e}$.

6.4.3 Planes and Projection onto Planes

Consider a plane in $\vec{I\!E}^3$ passing through the origin, and let $\vec{e}_1$ and $\vec{e}_2$ be two orthonormal vectors lying on the plane.[8] Then all the vectors lying on the plane can be written as a linear combination of $\vec{e}_1$ and $\vec{e}_2$, i.e., the plane is spanned by $\vec{e}_1$ and $\vec{e}_2$. For example, the x-y plane is spanned by $\vec{i}$ and $\vec{j}$, the x-z plane is spanned by $\vec{i}$ and $\vec{k}$ and the y-z plane is spanned by $\vec{j}$ and $\vec{k}$. Every of these planes are closed under addition and scalar multiplication. It is intuitively obvious that we can project a vector $\vec{v}$ onto a plane in $\vec{I\!E}^3$. For example, we can define the projection of $\vec{v}$ in Eq. (6.10) onto the x-y plane to be the vector $\vec{v}_{xy}$ given by

$$\vec{v}_{xy} := v_x\vec{i} + v_y\vec{j}. \tag{6.48}$$

The concepts of planes and projection onto a plane in $\vec{I\!E}^3$ can be generalised to *subspaces* in more general vector spaces, e.g., higher dimensional ones. To facilitate such generalisation we shall devote the following section to a formal discussion of subspaces in $\vec{I\!E}^3$.

6.4.4 Subspaces and Projection onto Subspaces

Definition 6.4.4(1) *A subset $\vec{\mathcal{S}}$ of $\vec{I\!E}^3$ is a subspace of $\vec{I\!E}^3$ if it is closed under addition and scalar multiplication*, i.e.,

$$\vec{v}_1,\ \vec{v}_2\ \in \vec{\mathcal{S}} \quad \Rightarrow \quad a_1\vec{v}_1 + a_2\vec{v}_2\ \in \vec{\mathcal{S}} \quad \forall\, a_1, a_2\ \in\ I\!R. \tag{6.49}$$

A subspace possesses the algebraic properties given in §6.1 and therefore it may be regarded as a vector space in its own right.

[8] Intuitively vectors can be imagined as arrows originating from the origin.

Definition 6.4.4(2)

(1) *Two subspaces $\vec{\mathcal{S}}_1$ and $\vec{\mathcal{S}}_2$ are said to be orthogonal if*

$$\langle \vec{v}_1 \mid \vec{v}_2 \rangle = 0 \quad \forall\, \vec{v}_1 \in \vec{\mathcal{S}}_1 \ \text{ and } \ \forall\, \vec{v}_2 \in \vec{\mathcal{S}}_2. \tag{6.50}$$

(2) *The orthogonal complement of a subspace $\vec{\mathcal{S}}$, denoted by $\vec{\mathcal{S}}^{\perp}$, is the set of all the vectors in $\vec{I\!E}^3$ which are orthogonal to every vector in $\vec{\mathcal{S}}$, i.e.,*

$$\vec{\mathcal{S}}^{\perp} := \left\{ \vec{u} \in \vec{I\!E}^3 : \langle \vec{u} \mid \vec{v} \rangle = 0, \quad \vec{v} \in \vec{\mathcal{S}} \right\}. \tag{6.51}$$

The orthogonal complement of a subspace is again a subspace. We can illustrate the concepts of subspaces and their orthogonal complements by the following examples:

E6.4.4(1) A subspace $\vec{\mathcal{S}}$ in $\vec{I\!E}^3$ is one-dimensional if there is only one linearly independent vector in $\vec{\mathcal{S}}$, i.e., there is a unit vector $\vec{e}$ such that

$$\vec{\mathcal{S}}_{\vec{e}} := \{ \vec{v} = v_e \vec{e} : \quad v_e \in I\!R \}. \tag{6.52}$$

The subspace is said to be *spanned* by the unit vectior $\vec{e}$. A subspace is two-dimensional if it is spanned by two orthonormal vectors.[9] A three-dimensional subspace would be identical to $\vec{I\!E}^3$ itself. The following are examples:

E6.4.4(1)(a) The x-axis and the z-axis defined by

$$\vec{\mathcal{S}}_{\vec{i}} := \{ \vec{v} = v_x \vec{i} : \quad v_x \in I\!R \}, \tag{6.53}$$

$$\vec{\mathcal{S}}_{\vec{k}} := \{ \vec{v} = v_z \vec{k} : \quad v_z \in I\!R \}, \tag{6.54}$$

are one-dimensional subspaces which are conveniently written as $\vec{\mathcal{S}}_x$ and $\vec{\mathcal{S}}_z$. They are orthogonal to each other. But $\vec{\mathcal{S}}_x$ is not the orthogonal complement to $\vec{\mathcal{S}}_z$, since $\vec{\mathcal{S}}_x$ does not contain all the vectors orthogonal to $\vec{\mathcal{S}}_z$.

[9] A two-dimensional subspace can be spanned by two linearly independent vectors, not necessarily orthonormal. But we can always employ two orthonormal vectors obtained by the Gram-Schmidt orthogonalisation process to span the subspace.

E6.4.4(1)(b) Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be an orthonormal basis. Using the basis vectors we can span subspaces of different dimensions, e.g.,

$$\vec{\mathcal{S}}_{12} := \{\vec{v}_{12} = v_2\vec{e}_1 + v_3\vec{e}_2 : \quad v_1,\ v_2 \in I\!R\}, \tag{6.55}$$

$$\vec{\mathcal{S}}_3 := \{\vec{v}_3 = v_3\vec{e}_3 : \quad v_3 \in I\!R\}, \tag{6.56}$$

where $\vec{\mathcal{S}}_{12}$ is two-dimensional and $\vec{\mathcal{S}}_3$ is one-dimensional.

E6.4.4(2) The orthogonal complement of a one-dimensional subspace spanned by a unit vector $\vec{e}$ is necessarily two-dimensional which can be spanned by two orthonormal unit vectors orthogonal to $\vec{e}$. Conversely a two-dimensional subspace has a one-dimensional orthogonal complement. The following are examples:

E6.4.4(2)(a) The orthogonal complement of $\vec{\mathcal{S}}_z$ is the x-y plane $\vec{\mathcal{S}}_{xy}$ which contains all the vectors in $I\!\vec{E}^3$ which are orthogonal to $\vec{\mathcal{S}}_z$. The orthogonal complement of the x-y plane is the z-axis, i.e.,

$$\vec{\mathcal{S}}_z^{\perp} = \vec{\mathcal{S}}_{xy}, \quad \vec{\mathcal{S}}_{xy}^{\perp} = \vec{\mathcal{S}}_z. \tag{6.57}$$

E6.4.4(2)(b) The subspaces in Eqs. (6.55) and (6.56) are the orthogonal complements to each other, i.e.,

$$\vec{\mathcal{S}}_{12}^{\perp} = \vec{\mathcal{S}}_3, \quad \vec{\mathcal{S}}_3^{\perp} = \vec{\mathcal{S}}_{12}. \tag{6.58}$$

The importance of a subspace $\vec{\mathcal{S}}$ and its orthogonal complement $\vec{\mathcal{S}}^{\perp}$ lies in the fact that an arbitrary vector $\vec{v}$ can be expressed as a unique linear combination of a vector $\vec{v}_{\vec{\mathcal{S}}}$ in $\vec{\mathcal{S}}$ and a vector $\vec{v}_{\vec{\mathcal{S}}^{\perp}}$ in $\vec{\mathcal{S}}^{\perp}$, i.e., for every vector $\vec{v}$ the following decomposition[10]:

$$\vec{v} = \vec{v}_{\vec{\mathcal{S}}} + \vec{v}_{\vec{\mathcal{S}}^{\perp}}, \quad \text{where} \quad \vec{v}_{\vec{\mathcal{S}}} \in \vec{\mathcal{S}}, \quad \vec{v}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{S}}^{\perp}, \tag{6.59}$$

exists and is unique. In the case of $\vec{\mathcal{S}}_z$ and $\vec{\mathcal{S}}_z^{\perp} = \vec{\mathcal{S}}_{xy}$ the decomposition becomes

$$\vec{v} = \vec{v}_{\vec{\mathcal{S}}_z} + \vec{v}_{\vec{\mathcal{S}}_z^{\perp}} \quad \text{with} \quad \vec{v}_{\vec{\mathcal{S}}_z} = v_z\vec{k}, \quad \vec{v}_{\vec{\mathcal{S}}_z^{\perp}} = v_x\vec{i} + v_y\vec{j}. \tag{6.60}$$

In the orthonormal basis $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ we can express an arbitrary vector $\vec{v}$ uniquely in accordance with Eq. (6.19) as

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3. \tag{6.61}$$

Clearly we have

$$v_3\vec{e}_3 \in \vec{\mathcal{S}}_3, \quad v_2\vec{e}_2 + v_3\vec{e}_3 \in \vec{\mathcal{S}}_{12} = \vec{\mathcal{S}}_3^{\perp}. \tag{6.62}$$

[10] *See* Eq. (13.15) and its proof in the solution to Q13(5).

Following the idea of projection onto a plane we can define the *projections of* $\vec{v}$ *onto the subspaces* $\vec{\mathcal{S}}$ *and* $\vec{\mathcal{S}}^{\perp}$ *to be* $\vec{v}_{\vec{\mathcal{S}}}$ *and* $\vec{v}_{\vec{\mathcal{S}}^{\perp}}$, respectively. Projections and the operation of projection will be seen later to be of crucial importance in the formulation of quantum mechanics.

Exercises and Problems

Q6(1) Prove Eqs. (6.1) to (6.3).

Q6(2) Prove Theorem 6.2.1(1).

Q6(3) Verify that the expression for $\langle \vec{u} \mid \vec{v} \rangle$ in Eq. (6.13) satisfies properties SP6.3.1(1), SP6.3.1(2) and SP6.3.1(3) of scalar product.

Q6(4) Show that two orthogonal vectors are linearly independent.

Q6(5) Prove Eq. (6.19).

Q6(6) Prove the Pythagoras theorem in the forms of Eqs. (6.22) and (6.27).

Q6(7) Prove Eqs. (6.30) and (6.31).

Q6(8) Verify the Gram-Schmidt orthogonalisation procedure given in §6.3.5.

Q6(9) Prove triangle inequalities (6.38) and (6.39).

Q6(10) Let $\{\vec{e}_\ell\}$ be an orthonormal basis. Show that any vector $\vec{v}$ is expressible as a sum of the projections $\vec{v}_{\vec{e}_\ell}$ of $\vec{v}$ onto the basis vectors $\vec{e}_\ell$, i.e., $\vec{v} = \vec{v}_{\vec{e}_1} + \vec{v}_{\vec{e}_2} + \vec{v}_{\vec{e}_3}$.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 7

Matrices and their Relations with Vectors

Matrices can provide a tangible and easily understood description of many abstract mathematical quantities, including vectors and operators on vectors. In this chapter we shall present a review of matrices. The review would include studies of the close relationship between matrices and vectors. In order to motivate later studies of operators the discussion in this chapter is presented in a way which can be readily generalised to similar discussion on operators.

7.1 Basic Definitions

There are many situations in mathematics and physics in which we have to manipulate rectangular arrays of real or complex numbers. We can display the numbers in an array, i.e., in rows and columns. For example, a 2 by 3 (written as 2×3) array is of the form

$$\boldsymbol{M} := \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \end{pmatrix}. \tag{7.1}$$

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

with 2 rows and 3 columns. We can also have $1 \times n$ arrays which consist of a row of n numbers, e.g., a 1×3 array is of the form

$$\boldsymbol{R} := \begin{pmatrix} R_{11} & R_{12} & R_{13} \end{pmatrix}, \tag{7.2}$$

and $n \times 1$ arrays which consist a column of n numbers, e.g., a 3×1 array is of the form

$$\boldsymbol{C} := \begin{pmatrix} C_{11} \\ C_{21} \\ C_{31} \end{pmatrix}. \tag{7.3}$$

For clarity it is sometimes useful to denote an $m \times n$ array, i.e., an array with m rows and n columns, explicitly by the symbol

$$(M_{ij})_{m\times n} \quad \text{where } i = 1, 2, \cdots, m \text{ and } j = 1, 2, \cdots, n, \tag{7.4}$$

rather than $\boldsymbol{M}$.

These arrays are called *matrices* with the numbers M_{ij} known as their *elements* when they obey certain rules of manipulation to be defined below. First let us introduce a few useful definitions before setting out the rules of manipulations of matrices:

(1) *Order of a matrix* A matrix of m rows and n columns is said to be a matrix of *order* $m \times n$.

(2) *Square matrices* An $n \times n$ matrix, i.e., a matrix of n rows and n columns, is called a *square matrix* of order n.

   (a) *Diagonal matrices* A square matrix $(M_{ij})_{n\times n}$ is said to be *diagonal* if $M_{ij} = 0$ when $i \neq j$. In other words, all the *off-diagonal elements*, i.e., M_{ij}, $i \neq j$, are zero. A diagonal matrix has only non-zero *diagonal elements* M_{ii}.

   (b) *Identity matrices* The $n \times n$ *identity matrix*, denoted by $(I_{ij})_{n\times n}$ or $\boldsymbol{I}_{n\times n}$, or simply by $\boldsymbol{I}$, is a diagonal matrix with all its diagonal elements equal to 1.

(3) *Row matrices* A matrix consisting only a row of elements, e.g., a $1 \times n$ matrix $(R_{1j})_{1\times n}$, is a *row matrix* of order n.

(4) *Column matrices* A matrix consisting only a column of elements, e.g., an $n \times 1$ matrix $(C_{i1})_{n\times 1}$, is a *column matrix* of order n.

(5) *Zero matrices* A matrix of order $m \times n$ with all its elements equal to zero is called the *zero matrix* of order $m \times n$, to be denoted by $\mathbf{0}_{m\times n}$. The zero row and column matrices of order n are denoted by $\mathbf{0}_{1\times n}$ and $\mathbf{0}_{n\times 1}$, respectively. For brevity a zero matrix is simply denoted by $\mathbf{0}$.

For a set of arrays to be matrices they must obey the following rules of addition, scalar multiplication and matrix multiplication:

1. Addition Matrices having the same numbers of rows and columns can be added by adding the corresponding elements, i.e.,

$$(M_{ij})_{m\times n} + (N_{ij})_{m\times n} := (M_{ij} + N_{ij})_{m\times n}. \tag{7.5}$$

2. Scalar multiplication When multiplying a matrix by a scalar $a \in \mathbb{C}$ we multiply every element of the matrix by the scalar, i.e., the new matrix $a(M_{ij})_{m\times n}$ has elements aM_{ij}, i.e.,

$$a(M_{ij})_{m\times n} := (aM_{ij})_{m\times n}. \tag{7.6}$$

For example, we have

$$aM := \begin{pmatrix} aM_{11} & aM_{12} & aM_{13} \\ aM_{21} & aM_{22} & aM_{23} \end{pmatrix}. \tag{7.7}$$

This is consistent with addition operation defined above, e.g.,

$$2\,(M_{ij})_{m\times n} = (M_{ij})_{m\times n} + (M_{ij})_{m\times n}. \tag{7.8}$$

To illustrate the addition and scalar multiplication operations consider the following 2×2 matrices

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{7.9}$$

known as the **Pauli matrices**.[1] It is a standard convention for Pauli matrices to be denoted by low case letters σ_x, σ_y and σ_z. The matrix σ_y contains the imaginary number $i = \sqrt{-1}$. These matrices can be

[1] Pauli (1900–1958) is an Austrian theoretical physicist, well known for the Pauli exclusion principle and Pauli matrices and his theory of spin.

multiplied by scalars and added together, e.g., given $a_x, a_y, a_z \in \mathbb{R}$ we have

$$a_x\sigma_x + a_y\sigma_y + a_z\sigma_z = \begin{pmatrix} a_z & a_x - ia_y \\ a_x + ia_y & -a_z \end{pmatrix}. \tag{7.10}$$

The resulting matrix has both real and complex elements.

3. Matrix multiplication The problem is to see how we can multiply two matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ with elements M_{ij} and N_{kl}.

(1) *Multiplication of a row matrix by a column matrix* Consider a 1×3 row matrix $(R_{1j})_{1\times 3}$ and a 3×1 column matrix $(C_{k1})_{3\times 1}$. With only three elements these matrices resemble vectors in $\vec{I\!E}^3$ which also have three components.[2] We can multiply a vector by another vector to form their scalar product by multiplying their corresponding components as given by the Pythagoras theorem in Eq. (6.27). So, a natural way to multiply a row matrix $(R_{1j})_{1\times 3}$ to a column matrix $(C_{k1})_{3\times 1}$ is to multiply their corresponding elements, i.e.,

$$\begin{aligned} \boldsymbol{R}\cdot\boldsymbol{C} &= \begin{pmatrix} R_{11} & R_{12} & R_{13} \end{pmatrix} \cdot \begin{pmatrix} C_{11} \\ C_{21} \\ C_{31} \end{pmatrix} \\ &:= R_{11}C_{11} + R_{12}C_{21} + R_{13}C_{31} = \sum_{j=1}^{3} R_{1j}C_{j1}. \end{aligned} \tag{7.11}$$

This results in a scalar. For brevity we often denote the elements of a column matrix $\boldsymbol{C}$ by C_1, C_2, C_3 and the elements of a row matrix $\boldsymbol{R}$ are often denoted by R_1, R_2, R_3. Then we can rewrite Eq. (7.11) as

$$\begin{aligned} \boldsymbol{R}\cdot\boldsymbol{C} &= \begin{pmatrix} R_1 & R_2 & R_3 \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} \\ &:= R_1C_1 + R_2C_2 + R_3C_3 = \sum_{j=1}^{3} R_jC_j. \end{aligned} \tag{7.12}$$

[2] Details on column matrices and vectors will be discussed in §7.5.

(2) *Multiplication of square matrices* Consider two 3×3 square matrices $(M_{ij})_{3\times 3}$ and $(N_{k\ell})_{3\times 3}$, i.e.,

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}, \quad \begin{pmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \\ N_{31} & N_{32} & N_{33} \end{pmatrix}. \tag{7.13}$$

To get an intuition of how one would multiply $(M_{ij})_{3\times 3}$ and $(N_{k\ell})_{3\times 3}$ we can relate these matrices to row and column matrices:

(a) The matrix $(M_{ij})_{3\times 3}$ may be considered to be formed by three row matrices, i.e.,

$$\boldsymbol{R}_1 := (M_{1j})_{1\times 3}, \quad \boldsymbol{R}_2 := (M_{2j})_{1\times 3}, \quad \boldsymbol{R}_3 := (M_{3j})_{1\times 3}.$$

(b) The matrix $(N_{k\ell})_{3\times 3}$ may be considered to be formed by three column matrices, i.e.,

$$\boldsymbol{C}_1 := (N_{k1})_{3\times 1}, \quad \boldsymbol{C}_2 := (N_{k2})_{3\times 1}, \quad \boldsymbol{C}_3 := (N_{k3})_{3\times 1}.$$

(c) Using Eq. (7.11) we can multiply these row matrices to the column matrices to obtain nine scalars $L_{il} =: \boldsymbol{R}_i \cdot \boldsymbol{C}_l$:

$$\boldsymbol{R}_1 \cdot \boldsymbol{C}_1 = \sum_{j=1}^{3} M_{1j} N_{j1}, \quad \boldsymbol{R}_1 \cdot \boldsymbol{C}_2 = \sum_{j=1}^{3} M_{1j} N_{j2}, \tag{7.14}$$

$$\boldsymbol{R}_1 \cdot \boldsymbol{C}_3 = \sum_{j=1}^{3} M_{1j} N_{j3}, \tag{7.15}$$

$$\boldsymbol{R}_2 \cdot \boldsymbol{C}_1 = \sum_{j=1}^{3} M_{2j} N_{j1}, \quad \boldsymbol{R}_2 \cdot \boldsymbol{C}_2 = \sum_{j=1}^{3} M_{2j} N_{j2}, \tag{7.16}$$

$$\boldsymbol{R}_2 \cdot \boldsymbol{C}_3 = \sum_{j=1}^{3} M_{2j} N_{j3}, \tag{7.17}$$

$$\boldsymbol{R}_3 \cdot \boldsymbol{C}_1 = \sum_{j=1}^{3} M_{3j} N_{j1}, \quad \boldsymbol{R}_3 \cdot \boldsymbol{C}_2 = \sum_{j=1}^{3} M_{3j} N_{j2}, \tag{7.18}$$

$$\boldsymbol{R}_3 \cdot \boldsymbol{C}_3 = \sum_{j=1}^{3} M_{3j} N_{j3}. \tag{7.19}$$

We can use these nine scalars to construct a new matrix $(L_{i\ell})_{3\times 3}$ with elements $L_{i\ell} := \boldsymbol{R}_i \cdot \boldsymbol{C}_\ell$. This new matrix is naturally defined to be the product of $(M_{ij})_{3\times 3}$ and $(N_{k\ell})_{3\times 3}$, i.e., we have

$$(M_{ij})_{3\times 3} \cdot (N_{k\ell})_{3\times 3} := (L_{i\ell})_{3\times 3}, \tag{7.20}$$

where

$$L_{i\ell} = \sum_{j=1}^{3} M_{ij} N_{j\ell}. \tag{7.21}$$

(3) *Generalisations* The definition in Eqs. (7.20) and (7.21) applies to the multiplication of any two $n \times n$ square matrices. We can further generalise the above multiplication rules to matrices with matching rows and columns, i.e., a matrix $(M_{ij})_{m\times r}$ having r columns can be multiplied by a matrix $(N_{k\ell})_{r\times n}$ having r rows. The product matrix $(L_{i\ell})_{m\times n}$ has m rows and n columns with its elements defined by

$$(M_{ij})_{m\times r} \cdot (N_{k\ell})_{r\times n} := (L_{i\ell})_{m\times n}, \tag{7.22}$$

where

$$L_{i\ell} = \sum_{j=1}^{r} M_{ij} N_{j\ell}. \tag{7.23}$$

The following examples serve as illustrations:

E7.1(1) *Multiplication of an $n \times n$ matrix $\boldsymbol{M}_{n\times n}$ and an $n \times 1$ column matrix $\boldsymbol{C}_{n\times 1}$* The resulting matrix $\boldsymbol{C}'$ is an $n \times 1$ column matrix. Denoting the elements of $\boldsymbol{C}$ and $\boldsymbol{C}'$ by $C_1, C_2, \cdots, C_n$ and $C'_1, C'_2, \cdots, C'_n$, respectively, we have

$$\boldsymbol{M} \cdot \boldsymbol{C} = \boldsymbol{C}' \quad \Leftrightarrow \quad \sum_{j=1}^{n} M_{ij} C_j = C'_i. \tag{7.24}$$

For example, we have

$$\boldsymbol{M} \cdot \boldsymbol{C} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} C'_1 \\ C'_2 \end{pmatrix}, \tag{7.25}$$

or explicitly

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} M_{11}C_1 + M_{12}C_2 \\ M_{21}C_1 + M_{22}C_2 \end{pmatrix}. \tag{7.26}$$

E7.1(2) *Multiplication of a column matrix* $(C_{i1})_{3\times 1}$ *by a row matrix* $(R_{1j})_{1\times 3}$ The product $\boldsymbol{C}\cdot\boldsymbol{R}$ is an $n\times n$ square matrix:

$$\begin{aligned}\boldsymbol{C}\cdot\boldsymbol{R} &= \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}\cdot\begin{pmatrix} R_1 & R_2 & R_3 \end{pmatrix} \\ &= \begin{pmatrix} C_1R_1 & C_1R_2 & C_1R_3 \\ C_2R_1 & C_2R_2 & C_2R_3 \\ C_3R_1 & C_3R_2 & C_3R_3 \end{pmatrix}. \end{aligned} \tag{7.27}$$

E7.1(3) *Multiplication of row matrix* $(R_{1j})_{1\times 3}$ by a column matrix $(C_{i1})_{3\times 1}$ The product $\boldsymbol{R}\cdot\boldsymbol{C}$ is a square matrix of order 1 which can be identified with a scalar, i.e.,

$$\begin{aligned}\boldsymbol{R}\cdot\boldsymbol{C} &= \begin{pmatrix} R_1 & R_2 & R_3 \end{pmatrix}\cdot\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} \\ &= \begin{pmatrix} C_1R_1 + C_1R_2 + C_1R_3 \end{pmatrix}. \end{aligned} \tag{7.28}$$

Matrix multiplication possesses the following associative and distributive properties:

$$\boldsymbol{L}\cdot(\boldsymbol{M}\cdot\boldsymbol{N}) = (\boldsymbol{L}\cdot\boldsymbol{M})\cdot\boldsymbol{N}, \tag{7.29}$$

$$\boldsymbol{L}\cdot(\boldsymbol{M}+\boldsymbol{N}) = \boldsymbol{L}\cdot\boldsymbol{M} + \boldsymbol{L}\cdot\boldsymbol{N}. \tag{7.30}$$

7.2 Square Matrices

For square matrices we can introduce the following important quantities:

1. Commutators The products $\boldsymbol{M}\cdot\boldsymbol{N}$ and $\boldsymbol{N}\cdot\boldsymbol{M}$ of two square matrices $\boldsymbol{M}=(M_{ij})_{n\times n}$ and $\boldsymbol{N}=(N_{k\ell})_{n\times n}$ are well-defined.[3] However, they are generally not equal. We call their difference, denoted in a bracket, i.e.,

$$[\boldsymbol{M},\boldsymbol{N}] := \boldsymbol{M}\cdot\boldsymbol{N} - \boldsymbol{N}\cdot\boldsymbol{M}, \tag{7.31}$$

[3] For matrices of different order the product $(N_{k\ell})_{r\times n}\cdot(M_{ij})_{m\times r}$ is not defined if $m\neq n$.

the *commutator* of $\boldsymbol{M}$ and $\boldsymbol{N}$. Two $n \times n$ matrices are said to *commute* if their commutator is equal to the $n \times n$ *zero matrix* $\mathbf{0}$. In other words, $\boldsymbol{M}$ and $\boldsymbol{N}$ commute if

$$[\boldsymbol{M}, \boldsymbol{N}] = \mathbf{0} \quad \text{or} \quad \boldsymbol{M} \cdot \boldsymbol{N} = \boldsymbol{N} \cdot \boldsymbol{M}. \tag{7.32}$$

2. Anticommutators Two $n \times n$ matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ are said to *anticommute* if their *anticommutator*

$$\{\boldsymbol{M}, \boldsymbol{N}\} := \boldsymbol{M} \cdot \boldsymbol{N} + \boldsymbol{N} \cdot \boldsymbol{M} \tag{7.33}$$

vanishes, i.e.,

$$\boldsymbol{M} \cdot \boldsymbol{N} + \boldsymbol{N} \cdot \boldsymbol{M} = \mathbf{0}. \tag{7.34}$$

3. Trace The sum of the diagonal elements of a square matrix $\boldsymbol{M}=(M_{ij})_{n\times n}$ is called the *trace* of the matrix which is denoted by $\mathrm{tr}\,(\boldsymbol{M})$, i.e.,

$$\mathrm{tr}\,(\boldsymbol{M}) := M_{11} + M_{22} + \cdots + M_{nn}. \tag{7.35}$$

For two square matrices $\boldsymbol{M}=(M_{ij})_{n\times n}$ and $\boldsymbol{N}=(N_{k\ell})_{n\times n}$ of the same order we have

$$\mathrm{tr}\,\big(\boldsymbol{M} + \boldsymbol{M}\big) = \mathrm{tr}\,(\boldsymbol{M}) + \mathrm{tr}\,(\boldsymbol{N}), \tag{7.36}$$

$$\mathrm{tr}\,\big(\boldsymbol{M} \cdot \boldsymbol{N}\big) = \mathrm{tr}\,\big(\boldsymbol{N} \cdot \boldsymbol{M}\big). \tag{7.37}$$

These results are valid whether the two matrices commute or not. This concept of trace of a square matrix has important physical applications, as will be seen in Chapter 32.

4. Determinant We can associate a value to an $n \times n$ square matrix $\boldsymbol{M}$. This value is calculated from the elements of the matrix according to certain rules. This value is known as the *determinant* of the matrix and is denoted by $\det(\boldsymbol{M})$. The rules for calculating the determinant of 2×2 and 3×3 matrices are

$$\det\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} := M_{11}M_{22} - M_{12}M_{21}. \tag{7.38}$$

$$\begin{aligned} \det\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} &:= M_{11} \det\begin{pmatrix} M_{22} & M_{23} \\ M_{32} & M_{33} \end{pmatrix} \\ -M_{12} \det\begin{pmatrix} M_{21} & M_{23} \\ M_{31} & M_{33} \end{pmatrix} &+ M_{13} \det\begin{pmatrix} M_{21} & M_{22} \\ M_{31} & M_{32} \end{pmatrix}. \end{aligned} \tag{7.39}$$

The determinant of an identity matrix is equal to 1 and the determinant of the product of two $n \times n$ matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ is equal to the product of the determinants of $\boldsymbol{M}$ and $\boldsymbol{N}$, i.e.,

$$\det(\boldsymbol{M} \cdot \boldsymbol{N}) = \det(\boldsymbol{M}) \det(\boldsymbol{N}). \tag{7.40}$$

5. Pauli matrices These simple but important matrices can serve to illustrate some of the properties discussed above.

(1) Pauli matrices do not commute, i.e., multiplication of Pauli matrices is not commutative. We can verify that

$$[\sigma_x, \sigma_y] = \sigma_x \cdot \sigma_y - \sigma_y \cdot \sigma_x = 2i\sigma_z, \tag{7.41}$$
$$[\sigma_z, \sigma_x] = \sigma_z \cdot \sigma_x - \sigma_x \cdot \sigma_z = 2i\sigma_y, \tag{7.42}$$
$$[\sigma_y, \sigma_z] = \sigma_y \cdot \sigma_z - \sigma_z \cdot \sigma_y = 2i\sigma_x. \tag{7.43}$$

(2) Pauli matrices do anticommute i.e.,

$$\{\sigma_x, \sigma_y\} = \sigma_x \cdot \sigma_y + \sigma_y \cdot \sigma_x = \mathbf{0}_{2\times 2}, \tag{7.44}$$
$$\{\sigma_z, \sigma_x\} = \sigma_z \cdot \sigma_x + \sigma_x \cdot \sigma_z = \mathbf{0}_{2\times 2}, \tag{7.45}$$
$$\{\sigma_y, \sigma_z\} = \sigma_y \cdot \sigma_z + \sigma_z \cdot \sigma_y = \mathbf{0}_{2\times 2}. \tag{7.46}$$

(3) The square of a Pauli matrix is equal to the 2×2 identity matrix $\boldsymbol{I}_{2\times 2}$, i.e.,

$$\sigma_x \cdot \sigma_x = \boldsymbol{I}_{2\times 2}, \tag{7.47}$$
$$\sigma_y \cdot \sigma_y = \boldsymbol{I}_{2\times 2}, \tag{7.48}$$
$$\sigma_z \cdot \sigma_z = \boldsymbol{I}_{2\times 2}. \tag{7.49}$$

(4) Pauli matrices have zero trace.

7.3 Transpose and Adjoint of a Matrix

Given a matrix $\boldsymbol{M}_{m\times n}$, not necessarily square, we can associate another matrix with it, i.e., its *transpose*. For a square matrix $\boldsymbol{M}_{n\times n}$ we can introduce two more matrices to associate with it, i.e., its *adjoint* and its *inverse*.[4]

[4] Zettili pp. 104–111 for explicit examples.

7.3.1 The Transpose of a Matrix

Definition 7.3.1(1) *The transpose* $\boldsymbol{M}^T$ *of an* $m \times n$ *matrix* $\boldsymbol{M}$ *is an* $n \times m$ *matrix with elements* $M_{ij}^T := M_{ji}$.

The following examples serve to illustrate the transpose operation:

E7.3.1(1) For a square matrix we have

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}^T := \begin{pmatrix} M_{11} & M_{21} & M_{31} \\ M_{12} & M_{22} & M_{32} \\ M_{13} & M_{23} & M_{33} \end{pmatrix}. \tag{7.50}$$

This example shows us that in the transpose operation the first row of $\boldsymbol{M}$ is converted into the first column of $\boldsymbol{M}^T$, the second row of $\boldsymbol{M}$ is converted into the second column of $\boldsymbol{M}^T$ and so on, and the first column of $\boldsymbol{M}$ is converted into the first row of $\boldsymbol{M}^T$, the second column of $\boldsymbol{M}$ is converted into the second row of $\boldsymbol{M}^T$ and so on. The transpose operation is also seen to swop the diagonally opposite elements of the matrix.

E7.3.1(2) Pauli matrices have the following transposes:

$$\sigma_x^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_x, \tag{7.51}$$

$$\sigma_y^T = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \neq \sigma_y, \tag{7.52}$$

$$\sigma_z^T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma_z. \tag{7.53}$$

E7.3.1(3) The transpose of a row matrix is a column matrix, e.g.,

$$\boldsymbol{R}^T = \begin{pmatrix} R_1 & R_2 & R_3 \end{pmatrix}^T = \begin{pmatrix} R_1^T \\ R_2^T \\ R_3^T \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}. \tag{7.54}$$

E7.3.1(4) The transpose of a column matrix is a row matrix, e.g.,

$$\boldsymbol{C}^T = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}^T = \begin{pmatrix} C_1^T & C_2^T & C_3^T \end{pmatrix} = (C_1 \quad C_2 \quad C_3). \tag{7.55}$$

E7.3.1(5) Using Eq. (7.11) or Eq. (7.12) we can multiply $\boldsymbol{C}^T$ by $\boldsymbol{C}$ to produce a number, e.g., for a 3×1 column matrix we have

$$\boldsymbol{C}^T \cdot \boldsymbol{C} = C_1^2 + C_2^2 + C_3^2. \tag{7.56}$$

On the other hand multiplying $\boldsymbol{C}$ by $\boldsymbol{C}^T$ produces a 3×3 matrix, i.e., $\boldsymbol{C} \cdot \boldsymbol{C}^T$ is a square matrix given by Eq. (7.27).

For complex matrices it is more useful to introduce the concept of their adjoints.

7.3.2 The Adjoint of a Square Matrix

Definition 7.3.2(1)[5] *The complex conjugate of the transpose of a square matrix $\boldsymbol{M}$, denoted by $\boldsymbol{M}^\dagger$, is called the adjoint of $\boldsymbol{M}$, i.e., the adjoint has elements $M_{ij}^\dagger$ given by*[6]

$$M_{ij}^\dagger := M_{ji}^* \tag{7.57}$$

The adjoint operation has the following properties:

P7.3.2(1) The adjoint of a real square matrix is equal to the transpose of the matrix.

P7.3.2(2) The adjoint operation satisfy the following equations:

$$(\boldsymbol{M}^\dagger)^\dagger = \boldsymbol{M}, \tag{7.58}$$
$$(a\boldsymbol{M})^\dagger = a^* \boldsymbol{M}^\dagger, \tag{7.59}$$
$$(\boldsymbol{M} + \boldsymbol{N})^\dagger = \boldsymbol{M}^\dagger + \boldsymbol{N}^\dagger, \tag{7.60}$$
$$(\boldsymbol{M} \cdot \boldsymbol{N})^\dagger = \boldsymbol{N}^\dagger \cdot \boldsymbol{M}^\dagger. \tag{7.61}$$

Here $\boldsymbol{M}$ and $\boldsymbol{N}$ are $n \times n$ matrices, possibly complex, and a is a complex number. Note that the adjoint of the product of two square matrices is equal to the product of the adjoints of the two matrices taken in the reverse order. The transpose operation also possesses

[5] Some authors use the term *conjugate matrix* instead, with the term *adjoint matrix* used to mean something quite different (*see* Hohn p. 85). Our definition of the term (*see* Finkbeiner II pp. 154, 271, 282) is motivated by a similar definition for operators

[6] The complex conjugate of a number a is denoted by a^*, e.g., M_{ji}^* is the complex conjugate of M_{ji}, and a is real if $a = a^*$.

these properties, except for Eq. (7.59) which is replaced by $(a\boldsymbol{M})^T = a\boldsymbol{M}^T$.

P7.3.2(3) Some complex matrices can remain unchanged under the adjoint operation, the Pauli matrices being examples, i.e.,

$$\sigma_x^\dagger = \sigma_x, \quad \sigma_y^\dagger = \sigma_y, \quad \sigma_z^\dagger = \sigma_z. \tag{7.62}$$

P7.3.2(4) The adjoint operation transform row matrices into and column matrices and vice versa.

(1) The adjoint of row matrices are column matrices, e.g.,

$$\boldsymbol{R}^\dagger = \begin{pmatrix} R_1 & R_2 & R_3 \end{pmatrix}^\dagger = \begin{pmatrix} R_1^\dagger \\ R_2^\dagger \\ R_3^\dagger \end{pmatrix} = \begin{pmatrix} R_1^* \\ R_2^* \\ R_3^* \end{pmatrix}. \tag{7.63}$$

(2) The adjoint of column matrices are row matrices, e.g.,

$$\begin{aligned} \boldsymbol{C}^\dagger = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}^\dagger &= \begin{pmatrix} C_1^\dagger & C_2^\dagger & C_3^\dagger \end{pmatrix} \\ &= (C_1^* \quad C_2^* \quad C_3^*). \end{aligned} \tag{7.64}$$

(3) Multiplying $\boldsymbol{C}$ by $\boldsymbol{C}^\dagger$ produces a $n \times n$ matrix, e.g., for a 3×1 column matrix we have, in accordance with Eq. (7.27),

$$\begin{aligned} \boldsymbol{C} \cdot \boldsymbol{C}^\dagger &= \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} \cdot (C_1^* \quad C_2^* \quad C_3^*) \\ &= \begin{pmatrix} C_1C_1^* & C_1C_2^* & C_1C_3^* \\ C_2C_1^* & C_2C_2^* & C_2C_3^* \\ C_3C_1^* & C_3C_2^* & C_3C_3^* \end{pmatrix}. \end{aligned} \tag{7.65}$$

(4) Multiplying $\boldsymbol{C}^\dagger$ by $\boldsymbol{C}$ produces a number, e.g., for a 3×1 column matrix we have

$$\boldsymbol{C}^\dagger \cdot \boldsymbol{C} = C_1^*C_1 + C_2^*C_2 + C_3^*C_3. \tag{7.66}$$

(5) When $\boldsymbol{C}$ contains complex elements we have

(a) The product $\boldsymbol{C}^T \cdot \boldsymbol{C}$ in Eq. (7.56) is generally complex.

(b) The product $\boldsymbol{C}^\dagger \cdot \boldsymbol{C}$ in Eq. (7.66) is *real and positive*. This product vanishes if and only if $\boldsymbol{C}$ is the zero column matrix. *This is what makes the adjoint operation and the product* $\boldsymbol{C}^\dagger \cdot \boldsymbol{C}$ *essential for physical applications*.

P7.3.2(5) The adjoint of a matrix can be characterised by its action on column vectors. This is discussed in relation to Eq. (7.81).

P7.3.2(6) The determinant of the transpose of a matrix is equal to the determinant of the matrix, i.e.,

$$\det(\boldsymbol{M}^T) = \det(\boldsymbol{M}). \tag{7.67}$$

7.4 The Inverse of a Matrix

Definition 7.4(1) *Given an* $n \times n$ *square matrix* $\boldsymbol{M}_{n\times n}$ *there may be another* $n \times n$ *matrices* $\boldsymbol{M}^{-1}_{n\times n}$ *such that*

$$\boldsymbol{M}_{n\times n} \cdot \boldsymbol{M}^{-1}_{n\times n} = \boldsymbol{M}^{-1}_{n\times n} \cdot \boldsymbol{M}_{n\times n} = \boldsymbol{I}_{n\times n}. \tag{7.68}$$

We call $\boldsymbol{M}^{-1}_{n\times n}$*, if it exists, the inverse of* $\boldsymbol{M}_{n\times n}$ *and the matrix* $\boldsymbol{M}_{n\times n}$ *is said to be invertible.*[7]

The following properties are well-known:

P7.4(1) The inverse of a square matrix, if it exists, is unique.[8]

P7.4(2) A square matrix possesses an inverse if and only if its determinant is not zero. There is also a standard formula for the construction of the inverse matrix.[9]

P7.4(3) A square matrix $\boldsymbol{M}_{n\times n}$ is not invertible if a non-zero column matrix $\boldsymbol{C}_{n\times 1}$ exists such that $\boldsymbol{M}_{n\times n}\boldsymbol{C}_{n\times 1} = \boldsymbol{0}_{n\times 1}$, i.e.,

$$\boldsymbol{M}_{n\times n} \cdot \boldsymbol{C}_{n\times 1} = \boldsymbol{0}_{n\times 1} \quad \not\Rightarrow \quad \boldsymbol{C}_{n\times 1} = \boldsymbol{0}_{n\times 1}. \tag{7.69}$$

[7] Here $\boldsymbol{I}_{n\times n}$ is the $n \times n$ identity matrix. Inverses are defined for square matrices.
[8] Meyer, p. 116.
[9] Hohn p. 93. Since we do not need to do explicit calculation in terms of inverses we shall not provide a formula for inverses here.

The matrix is invertible if no such column matrix exists, i.e., if

$$\boldsymbol{M}_{n\times n}\cdot\boldsymbol{C}_{n\times 1}=\boldsymbol{0}_{n\times 1}\quad\Rightarrow\quad\boldsymbol{C}_{n\times 1}=\boldsymbol{0}_{n\times 1}. \tag{7.70}$$

P7.4(4) The following result holds for $n\times n$ matrices[10]:

$$\boldsymbol{N}\cdot\boldsymbol{M}=\boldsymbol{I}\quad\Rightarrow\quad\boldsymbol{M}\cdot\boldsymbol{N}=\boldsymbol{I}. \tag{7.71}$$

The condition $\boldsymbol{N}\cdot\boldsymbol{M}=\boldsymbol{I}$ is sufficient for $\boldsymbol{N}$ to be the inverse of $\boldsymbol{M}$. This result is valid also if $\boldsymbol{M}\cdot\boldsymbol{N}=\boldsymbol{I}$.

P7.4(5) The inverse of an invertible matrix $\boldsymbol{M}$ is also invertible and we have $(\boldsymbol{M}^{-1})^{-1}=\boldsymbol{M}$.

P7.4(6) The inverse of the product of two $n\times n$ square matrices of the same order is equal to the product of the inverses of the two matrices taken in the reverse order, i.e.,

$$\Big(\boldsymbol{M}\cdot\boldsymbol{N}\Big)^{-1}=\boldsymbol{N}^{-1}\cdot\boldsymbol{M}^{-1}. \tag{7.72}$$

7.5 Matrix Representation of Vectors in $\vec{I\!E}^3$

7.5.1 Scalar Product of Column Matrices

Column matrices possess many of the properties of vectors, e.g., the algebraic properties given in and §6.1. We can also define norms and scalar products of column matrices in a way similar to Eqs. (6.26) and (6.28) for vectors. To illustrate the similarities we can introduce the following definitions on 3×1 column matrices[11]:

(1) Assign a number, possibly complex and denoted by $\langle\boldsymbol{C}\mid\boldsymbol{C}'\rangle$, to two column matrices $\boldsymbol{C}$ and $\boldsymbol{C}'$ by

$$\langle\boldsymbol{C}\mid\boldsymbol{C}'\rangle:=\boldsymbol{C}^{\dagger}\cdot\boldsymbol{C}'=C_1^*C_1'+C_2^*C_2'+C_3^*C_3'. \tag{7.73}$$

[10] Meyer, p. 117. Halmos p. 62. This does not hold generally for matrices of infinite order. We shall drop the subscripts $n\times n$ in what follows.

[11] All the column matrices cited, i.e., $\boldsymbol{C}$, $\boldsymbol{C}'$ and $\boldsymbol{C}''$, are 3×1 column matrices.

This assignment possesses the following properties[12]:

$$\langle \boldsymbol{C} \mid \boldsymbol{C}' \rangle = \langle \boldsymbol{C}' \mid \boldsymbol{C} \rangle^*. \tag{7.74}$$

$$\langle a\boldsymbol{C} \mid \boldsymbol{C}' \rangle = a^* \langle \boldsymbol{C}' \mid \boldsymbol{C} \rangle^*, \ a \in \mathbb{C}. \tag{7.75}$$

$$\langle \boldsymbol{C} \mid a\,\boldsymbol{C}' + b\,\boldsymbol{C}'' \rangle = a\langle \boldsymbol{C} \mid \boldsymbol{C}' \rangle + b\langle \boldsymbol{C} \mid \boldsymbol{C}'' \rangle, \ a, b \in \mathbb{C}. \tag{7.76}$$

$$\langle \boldsymbol{C} \mid \boldsymbol{C} \rangle \geq 0, \qquad \langle \boldsymbol{C} \mid \boldsymbol{C} \rangle = 0 \ \Rightarrow \ \boldsymbol{C} = \boldsymbol{0}. \tag{7.77}$$

This number is defined to be the **scalar product** of $\boldsymbol{C}$ and $\boldsymbol{C}'$.

(2) The **norm** $||\boldsymbol{C}||$ of a column matrix $\boldsymbol{C}$ is defined to be[13]

$$||\boldsymbol{C}|| := \sqrt{\langle \boldsymbol{C} \mid \boldsymbol{C} \rangle} = \sqrt{C_1^* C_1 + C_2^* C_2 + C_3^* C_3}. \tag{7.78}$$

A column matrix $\boldsymbol{C}$ is said to be *normalised* if $||\boldsymbol{C}|| = 1$.

(3) Two column matrices $\boldsymbol{C}$ and $\boldsymbol{C}'$ are said to be *orthogonal* if their scalar product vanishes, i.e., if

$$\langle \boldsymbol{C} \mid \boldsymbol{C}' \rangle = C_1^* C_1' + C_2^* C_2' + C_3^* C_3' = 0. \tag{7.79}$$

(4) Two column matrices $\boldsymbol{C}$ and $\boldsymbol{C}'$ are said to be *orthonormal* if they are normalised and orthogonal to each other.

The definitions listed above can be extended to $n \times 1$ column matrices in a straight forward manner.

Let $\boldsymbol{M}$ be a 3×3 square matrix and let $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ be two 3×1 column matrices. Then we have, by Eqs. (7.73) and (7.61),

$$\begin{aligned} \langle \boldsymbol{M}^\dagger \cdot \boldsymbol{C}_1 \mid \boldsymbol{C}_2 \rangle :&= \left(\boldsymbol{M}^\dagger \cdot \boldsymbol{C}_1\right)^\dagger \cdot \boldsymbol{C}_2 = \left(\boldsymbol{C}_1^\dagger \cdot \boldsymbol{M}\right) \cdot \boldsymbol{C}_2 \\ &= \boldsymbol{C}_1^\dagger \cdot (\boldsymbol{M} \cdot \boldsymbol{C}_2). \end{aligned} \tag{7.80}$$

This means that the scalar product of $\boldsymbol{M}^\dagger \cdot \boldsymbol{C}_1$ and $\boldsymbol{C}_2$ is equal to the scalar product of $\boldsymbol{C}_1$ and $\boldsymbol{M} \cdot \boldsymbol{C}_2$. We can express this result as

$$\boxed{\langle \boldsymbol{M}^\dagger \cdot \boldsymbol{C}_1 \mid \boldsymbol{C}_2 \rangle = \langle \boldsymbol{C}_1 \mid \boldsymbol{M} \cdot \boldsymbol{C}_2 \rangle.} \tag{7.81}$$

[12] For matrices with real elements these correspond to SP6.3.1(1), SP6.3.1(2) and SP6.3.1(3) in §6.3 for the scalar product of real vectors, hence the Dirac notation.

[13] This is why the adjoint, not transpose, is chosen in Eq. (7.73), i.e., we want $\langle \boldsymbol{C} \mid \boldsymbol{C} \rangle$ to be real.

This equation is so characteristic of adjoint matrices that it can be used to define the adjoint of any $n \times n$ matrix, i.e.,

> *the adjoint of an $n \times n$ square matrix $\boldsymbol{M}$ is the matrix $\boldsymbol{M}^\dagger$ which satisfies Eq. (7.81) for all $n \times 1$ column matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$.*

This is an important result since this definition is also used to define the adjoint of operators.[14]

7.5.2 Column Matrices and Column Vectors

Any vector in $\vec{I\!E}^3$ can be expressed uniquely as a unique linear combination of the basis vectors of any chosen orthonormal basis $\{\vec{e}_\ell\}$, i.e., we have

$$\vec{u} = \sum_{\ell=1}^{3} u_\ell \, \vec{e}_\ell, \quad u_\ell = \langle \vec{e}_\ell \mid \vec{u} \rangle, \tag{7.82}$$

$$\vec{v} = \sum_{\ell=1}^{3} v_\ell \, \vec{e}_\ell, \quad v_\ell = \langle \vec{e}_\ell \mid \vec{v} \rangle. \tag{7.83}$$

From Eqs. (6.20) to (6.28) we know that we can perform calculations on vectors in terms of their components. In fact we can represent vectors themselves together with all the calculations using their components. All this can be done formally in terms of matrices.

First we choose an orthonormal basis $\{\vec{e}_\ell\}$. All vectors in $\vec{I\!E}^3$ can be expressed in the form of Eqs. (7.82) and (7.83). Using the components in the linear combination we can construct a unique 3×1 column matrix for each vector, e.g.,

$$\boldsymbol{C}_{\vec{u}} := \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{C}_{\vec{v}} := \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \tag{7.84}$$

All the usual operations of vectors can be performed in terms of these column matrices:

(1) *Scalar multiplication*

$$\text{For} \quad a\vec{u} \quad \text{we have} \quad a\boldsymbol{C}_{\vec{u}} = \begin{pmatrix} au_1 \\ au_2 \\ au_3 \end{pmatrix}. \tag{7.85}$$

[14] *See* Eq. (8.33).

(2) *Addition*

$$\text{For} \quad \vec{u}+\vec{v} \quad \text{we have} \quad \boldsymbol{C}_{\vec{u}}+\boldsymbol{C}_{\vec{v}}=\begin{pmatrix} u_1+v_1 \\ u_2+v_2 \\ u_3+v_3 \end{pmatrix}. \tag{7.86}$$

(3) *Norm and scalar product* For the scalar product and norm of vectors we have, using Eqs. (7.73) and (7.78),

$$\langle \vec{u} \mid \vec{v} \rangle = \langle \boldsymbol{C}_{\vec{u}} \mid \boldsymbol{C}_{\vec{v}} \rangle, \tag{7.87}$$

$$||\vec{v}||^2 = \langle \boldsymbol{C}_{\vec{v}} \mid \boldsymbol{C}_{\vec{v}} \rangle = ||\boldsymbol{C}_{\vec{v}}||^2. \tag{7.88}$$

The conclusion is that vectors in $\vec{I\!E}^3$ can be represented by 3×1 column matrices in the sense of the following correspondences:

$$\vec{u} \quad \leftrightarrow \quad \boldsymbol{C}_{\vec{u}}, \qquad \vec{v} \quad \leftrightarrow \quad \boldsymbol{C}_{\vec{v}}, \tag{7.89}$$

$$a\,\vec{u} \quad \leftrightarrow \quad a\,\boldsymbol{C}_{\vec{u}}, \qquad \vec{u}+\vec{v} \quad \leftrightarrow \quad \boldsymbol{C}_{\vec{u}}+\boldsymbol{C}_{\vec{v}}, \tag{7.90}$$

$$||\vec{v}|| \quad \leftrightarrow \quad ||\boldsymbol{C}_{\vec{v}}||, \qquad \langle \vec{u} \mid \vec{v} \rangle \quad \leftrightarrow \quad \langle \boldsymbol{C}_{\vec{u}} \mid \boldsymbol{C}_{\vec{v}} \rangle. \tag{7.91}$$

These correspondences enable us to carry out operations and calculations in terms of column matrices. In addition to scalar product and norm the concepts of linear combination, linear dependence and independence presented in §6.2.1 can be applied to column matrices. We shall refer to column matrices as **column vectors** to emphasise their similarity with vectors.

It should be pointed out that the actual representative column matrix of a vector depends on the choice of the orthonormal basis. For example, if we choose the basis $\vec{i}, \vec{j}, \vec{k}$ of a given coordinate system the column vector for $\vec{u}$ in this basis is

$$\boldsymbol{C}_{\vec{u}} := \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}, \quad u_x = \langle \vec{i} \mid \vec{u} \rangle,\ u_y = \langle \vec{j} \mid \vec{u} \rangle,\ u_z = \langle \vec{k} \mid \vec{u} \rangle, \tag{7.92}$$

and the basis vectors $\vec{i}, \vec{j}, \vec{k}$ corresponds to column vectors

$$\boldsymbol{C}_{\vec{i}} := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{C}_{\vec{j}} := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{C}_{\vec{k}} := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{7.93}$$

We have

$$\boldsymbol{C}_{\vec{u}} = u_x \boldsymbol{C}_{\vec{i}} + u_y \boldsymbol{C}_{\vec{j}} + u_z \boldsymbol{C}_{\vec{k}}. \tag{7.94}$$

This result can be generalised to $n \times 1$ column matrices which are also referred to as $n \times 1$ column vectors. A set of n orthonormal $n \times 1$ column vectors $\boldsymbol{C}_\ell$, $\ell = 1, 2, \cdots, n$, can span the set of all $n \times 1$ column vectors $\boldsymbol{C}$, i.e., we have,

$$\boldsymbol{C} = \sum_{\ell=1}^{n} c_\ell \boldsymbol{C}_\ell, \quad c_\ell = \langle \boldsymbol{C}_\ell \mid \boldsymbol{C} \rangle. \tag{7.95}$$

Such a set of $n \times 1$ column vectors is said to constitute an *orthonormal basis in the set of all* $n \times 1$ *column vectors* $\boldsymbol{C}$.

7.6 Eigenvalue Problem for Matrices

Mathematicians have investigated what is known as the *eigenvalue problem* for a long time. Basically we start with an operation on a set of quantities, be it column vectors, vectors or functions. The problem is to find those quantities within the set which are least affected by the operation. Let $\boldsymbol{C}$ be an $n \times 1$ column vector with elements $c_1, c_2, \cdots, c_n$. An $n \times n$ matrix $\boldsymbol{M}$ can operate on $\boldsymbol{C}$ by matrix multiplication, i.e., we have

$$\boldsymbol{M} \cdot \boldsymbol{C} = \boldsymbol{C}', \tag{7.96}$$

or rewriting the multiplication without the *dot* for brevity

$$\boldsymbol{M}\boldsymbol{C} = \boldsymbol{C}'. \tag{7.97}$$

The column vector $\boldsymbol{C}$ is considered *least affected* if the action of $\boldsymbol{M}$ changes $\boldsymbol{C}$ only to the extend of a multiplicative constant, i.e., $\boldsymbol{M}$ acts like scalar multiplication on $\boldsymbol{C}$ so that the above equation becomes[15]

$$\boldsymbol{M}\boldsymbol{C} = \lambda \boldsymbol{C}, \quad \lambda \in \mathbb{C} \quad \text{or} \tag{7.98}$$

$$\sum_{j=1}^{n} M_{ij} C_j = \lambda C_i. \tag{7.99}$$

Then:

(1) Equation (7.98) is known as an **eigenvalue equation**.

[15] Both $\boldsymbol{M}$ and $\boldsymbol{C}$ may be complex. We exclude the trivial case of $\boldsymbol{C} = \boldsymbol{0}$. Zettili pp. 114–117 for explicit examples.

(2) The number λ is called an **eigenvalue** of the matrix $\boldsymbol{M}$.

(3) The column vector $\boldsymbol{C}$ is called an **eigenvector** of $\boldsymbol{M}$ corresponding to the eigenvalue λ.

(4) Eigenvalues and eigenvectors are generally not unique. A matrix can have many different eigenvalues. There may be many eigenvectors corresponding to the same eigenvalue. For example, if we multiply an eigenvector by a scalar the new column vector will also be an eigenvector corresponding to the same eigenvalue. However, we do not count this as a new eigenvector.

> *Two eigenvectors are deemed different only if they are linearly independent of each other, i.e., they are not related to each other by a multiplicative constant.*

(5) The **eigenvalue problem** for a given matrix $\boldsymbol{M}$ is to solve its eigenvalue equation, i.e., Eq. (7.98), for the eigenvalues λ and their corresponding eigenvectors $\boldsymbol{C}$.

For a 2×2 matrix $\boldsymbol{M}$ its eigenvalue equation is

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \lambda \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}. \tag{7.100}$$

We can re-arrange Eq. (7.100) as

$$\begin{pmatrix} M_{11} - \lambda & M_{12} \\ M_{21} & M_{22} - \lambda \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{7.101}$$

Carrying out the matrix multiplication we obtain two simultaneous equations involving the unknowns λ, C_1 and C_2:

$$(M_{11} - \lambda)C_1 + M_{12}C_2 = 0, \tag{7.102}$$

$$M_{21}C_1 + (M_{22} - \lambda)C_2 = 0. \tag{7.103}$$

Simultaneous equations of this kind are known to admit solutions if and only if the following determinant is equal to zero, i.e.,

$$\begin{vmatrix} M_{11} - \lambda & M_{12} \\ M_{21} & M_{22} - \lambda \end{vmatrix} = 0. \tag{7.104}$$

Expanding the determinant we obtain a polynomial in λ of degree 2, i.e.,

$$\lambda^2 - (M_{11} + M_{22})\lambda + (M_{11}M_{22} - M_{12}M_{21}) = 0. \tag{7.105}$$

We can solve for λ and then substitute the value into the simultaneous equations to solve for C_1 and C_2. A quadratic equation with possibly complex coefficients such as Eq. (7.105) has two solutions for λ in general. The solutions are not necessarily different. We can summarise the results as follows:

R7.6(1) A 2×2 matrix $\boldsymbol{M}$ has a maximum of two eigenvalues λ_1, λ_2 corresponding to two different eigenvectors $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$. The two eigenvalues may be identical but correspond to different eigenvectors.

R7.6(2) The eigenvalues and eigenvectors may be complex, even if the original matrix is real, and on the other hand a complex matrix can have real eigenvalues.[16]

The following examples illustrate the results listed above:

E7.6(1) *Diagonal matrices* The eigenvalues of a diagonal matrix are simply its diagonal elements. The corresponding eigenvectors have only one non-zero element, e.g.,

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{7.106}$$

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = b \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{7.107}$$

When $a = b$ the matrix would have a single eigenvalue but with two different eigenvectors to correspond to it.[17] Generally an $n \times n$ diagonal matrix would have n eigenvalues, not necessarily all different, corresponding to n orthonormal eigenvectors.

E7.6(2) *Degeneracy* An eigenvalue is said to be **nondegenerate** if it corresponds to only one eigenvector.[18] A single eigenvalue can correspond to different eigenvectors. The eigenvalue in Eqs. (7.106) and (7.107) when $a = b$ is an example. Such an eigenvalue is then said to be **degenerate**. The number of different eigenvectors corresponding to the same eigenvalue is called the *degeneracy*

[16] A real algebraic equation, e.g., $x^2 = -1$, can have complex solutions.
[17] The zero column vector is not counted as an eigenvector.
[18] Recall that eigenvectors related by a multiplicative constant are not regarded as different.

of that eigenvalue. A linear combination of two eigenvectors corresponding to the same eigenvalue is again an eigenvector corresponding to the same eigenvalue.

E7.6(3) *Real matrices with only one eigenvalue* The following real matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{7.108}$$

has only one nondegenerate real eigenvalue. One can check that the determinant in Eq. (7.104) gives rise to a polynomial $\lambda^2 - 2\lambda + 1 = 0$ which has only one root, i.e., $\lambda = 1$. The eigenvalue equation is

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{7.109}$$

There are no other eigenvectors and eigenvalues.

E7.6(4) *Real matrices with complex eigenvalues* The following matrix

$$\boldsymbol{R}(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in (0, 2\pi) \tag{7.110}$$

admits no real eigenvalues nor real eigenvectors. Equation (7.105) gives rise to two complex eigenvalues $\lambda_\pm = \exp(\pm i\theta) = \cos\theta \pm i \sin\theta$, together with two complex eigenvectors, i.e.,

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = e^{i\theta} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \tag{7.111}$$

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} = e^{-i\theta} \begin{pmatrix} 1 \\ i \end{pmatrix}. \tag{7.112}$$

E7.6(5) *Complex matrices with real eigenvalues* Complex matrices can have real eigenvalues with complex eigenvectors. This is illustrated by the Pauli matrix σ_y. For references we shall list the eigenvalues and eigenvectors of the three Pauli matrices here.

(1) The Pauli matrix σ_x possesses two real eigenvalues, i.e., ± 1, corresponding to two real eigenvectors:

$$\sigma_x \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \sigma_x \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1 \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{7.113}$$

(2) The Pauli matrix σ_y possesses two real eigenvalues, i.e., ± 1, corresponding to two complex eigenvectors:

$$\sigma_y \begin{pmatrix} 1 \\ i \end{pmatrix} = 1 \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \sigma_y \begin{pmatrix} 1 \\ -i \end{pmatrix} = -1 \begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{7.114}$$

(3) The Pauli matrix σ_z possesses two real eigenvalues, i.e., ± 1, corresponding to two real eigenvectors:

$$\sigma_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \sigma_z \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{7.115}$$

We can express the above eigenvectors in terms of our notation for normalised column vectors, i.e.,[19]

$$\boldsymbol{C}_{\vec{\alpha}_x} := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\beta}_x} := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{7.116}$$

$$\boldsymbol{C}_{\vec{\alpha}_y} := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\beta}_y} := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \tag{7.117}$$

$$\boldsymbol{C}_{\vec{\alpha}_z} := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\beta}_z} := \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{7.118}$$

The eigenvalue equations of Pauli matrices can be written as

$$\sigma_x \boldsymbol{C}_{\vec{\alpha}_x} = \boldsymbol{C}_{\vec{\alpha}_x}, \quad \sigma_x \boldsymbol{C}_{\vec{\beta}_x} = -\boldsymbol{C}_{\vec{\beta}_x}, \tag{7.119}$$

$$\sigma_y \boldsymbol{C}_{\vec{\alpha}_y} = \boldsymbol{C}_{\vec{\alpha}_y}, \quad \sigma_y \boldsymbol{C}_{\vec{\beta}_y} = -\boldsymbol{C}_{\vec{\beta}_y}, \tag{7.120}$$

$$\sigma_z \boldsymbol{C}_{\vec{\alpha}_z} = \boldsymbol{C}_{\vec{\alpha}_z}, \quad \sigma_z \boldsymbol{C}_{\vec{\beta}_z} = -\boldsymbol{C}_{\vec{\beta}_z}, \tag{7.121}$$

E7.6(6) *Matrices of higher order* We have similar results for the eigenvalue problem for matrices of higher order, e.g., a square matrix of order n has at least one eigenvalue and at most n different eigenvalues.

[19] The notation is related to the various states of electron spin, e.g., α_z^s introduced in §5.3. In Chapter 36 we will see that these spin states are described by vectors known as *spin state vectors*, e.g., $\vec{\alpha}_z$. So, $\boldsymbol{C}_{\vec{\alpha}_z}$ is the representation of the spin state vector $\vec{\alpha}_z$.

7.7 Special Matrices

7.7.1 Introduction

In many applications we desire square matrices with some particular properties.[20] For example, we may desire matrices with the following properties:

1. *Idempotence* A matrix is said to be *idempotent* if its square is equal to itself, i.e., $\boldsymbol{M}^2 = \boldsymbol{M}$.

2. *Symmetry* A matrix $\boldsymbol{M}$ is called *symmetric* if its transpose is equal to itself, i.e., $\boldsymbol{M}^T = \boldsymbol{M}$.

3. *Antisymmetry* A matrix $\boldsymbol{M}$ is called *antisymmetric* if its transpose is equal to the negative of itself, i.e., $\boldsymbol{M}^T = -\boldsymbol{M}$.

There are square matrices which are closely related to their inverses and adjoints. A study of these relations leads to a number of special matrices which are important in many applications. In this section we shall present a study of several of these matrices

7.7.2 Orthogonal Matrices

Definition 7.7.2(1) *A real square matrix is said to be orthogonal if it is invertible and its inverse is equal to its transpose, i.e.,* $\boldsymbol{R}^T = \boldsymbol{R}^{-1}$.

Orthogonal matrices are real square matrices satisfying the following condition:

$$\boldsymbol{R}^T \boldsymbol{R} = \boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}. \tag{7.122}$$

In terms of its elements we have

$$\sum_{j=1}^{n} R_{ij}^T R_{j\ell} = \delta_{i\ell}, \quad \text{or} \quad \sum_{j=1}^{n} R_{ji} R_{j\ell} = \delta_{i\ell}. \tag{7.123}$$

[20] We shall confine ourselves to square matrices through out this section.

The term *orthogonal matrices* arises from the fact that the column vectors

$$\boldsymbol{C}_1 = \begin{pmatrix} R_{11} \\ R_{21} \\ . \\ . \\ . \\ R_{n1} \end{pmatrix}, \quad \boldsymbol{C}_2 = \begin{pmatrix} R_{12} \\ R_{22} \\ . \\ . \\ . \\ R_{n2} \end{pmatrix}, \quad \cdots \quad \boldsymbol{C}_n = \begin{pmatrix} R_{1n} \\ R_{2n} \\ . \\ . \\ . \\ R_{nn} \end{pmatrix}, \tag{7.124}$$

which form the matrix $\boldsymbol{R}$ are orthogonal to each other, i.e., $\langle \boldsymbol{C}_i \mid \boldsymbol{C}_j \rangle = \boldsymbol{0}$. Each of these column vector is also normalised. Orthogonal matrices have the following properties:

P7.7.2(1) Their determinants have a value 1 or -1, since[21]

$$\det(\boldsymbol{R}^T \boldsymbol{R}) = \det(\boldsymbol{I}) = 1 \quad \Rightarrow \quad \big(\det(\boldsymbol{R})\big)^2 = 1. \tag{7.125}$$

P7.7.2(2) An $n \times n$ orthogonal matrix preserves the norms of $n \times 1$ real column vectors, i.e., if $\boldsymbol{C}' = \boldsymbol{R}\,\boldsymbol{C}$ then $||\boldsymbol{C}'||^2 = ||\boldsymbol{C}||^2$, since

$$\begin{aligned} ||\boldsymbol{C}'||^2 &= (\boldsymbol{R}\boldsymbol{C})^\dagger (\boldsymbol{R}\boldsymbol{C}) = (\boldsymbol{C}^\dagger \boldsymbol{R}^\dagger)(\boldsymbol{R}\boldsymbol{C}) \\ &= \boldsymbol{C}^\dagger (\boldsymbol{R}^\dagger \boldsymbol{R})\,\boldsymbol{C} = \boldsymbol{C}^\dagger (\boldsymbol{R}^T \boldsymbol{R})\,\boldsymbol{C} = ||\boldsymbol{C}||^2. \end{aligned} \tag{7.126}$$

This is such a characteristic feature of orthogonal matrices that it can be used to define orthogonal matrices, i.e.,

> *an $n \times n$ real matrix is orthogonal if and only if it preserves the norms of $n \times 1$ real column vectors.*

P7.7.2(3) An orthogonal matrix $\boldsymbol{R}$ can be used to change a column vector $\boldsymbol{C}$ to $\boldsymbol{C}'$ by

$$\boldsymbol{C}' := \boldsymbol{R}\boldsymbol{C}. \tag{7.127}$$

Such a change is called an **orthogonal transformation**. An orthogonal transformation preserves the scalar product of column vectors on account of Eq. (7.126).

P7.7.2(4) Consider the eigenvalue equation of an orthogonal matrix $\boldsymbol{R}$:

$$\boldsymbol{R}\boldsymbol{C}_r = r\boldsymbol{C}_r, \tag{7.128}$$

[21] Using Eqs. (7.40) and (7.67).

where r is an eigenvalue and $\boldsymbol{C}_r$ is a corresponding eigenvector. Since $\boldsymbol{R}$ preserves the norm of $\boldsymbol{C}_r$ we get

$$\langle \boldsymbol{R}\boldsymbol{C}_r \mid \boldsymbol{R}\boldsymbol{C}_r \rangle = \langle \boldsymbol{C}_r \mid \boldsymbol{C}_r \rangle \Rightarrow \quad |r|^2 = 1. \tag{7.129}$$

This does not mean that $r = \pm 1$ since r may be complex.[22]

P7.7.2(5) The product of two orthogonal matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ is again orthogonal, since

$$(\boldsymbol{R}_1\boldsymbol{R}_2)^T = \boldsymbol{R}_2^T\boldsymbol{R}_1^T = \boldsymbol{R}_2^{-1}\boldsymbol{R}_1^{-1} = (\boldsymbol{R}_1\boldsymbol{R}_2)^{-1}. \tag{7.130}$$

The geometric significance of orthogonal matrices of order 2 and order 3 can be seen in the following discussion.

1. Orthogonal matrices of order 2

(1) *Proper rotations* Consider a vector $\vec{u} = u_x\,\vec{i}$ lying on the x-axis with one end attached to the origin. We can rotate this vector anticlockwise on the x-y plane about the z-axis by an angle θ to arrive at a new vector $\vec{u}\,'$ without changing its norm, i.e., we have $||\vec{u}\,|| = ||\vec{u}\,'||$. The rotated vector can be shown to be related to $\vec{u}$ by[23]

$$\vec{u}\,' = u_x'\,\vec{i} + u_y'\,\vec{j}, \quad u_x' = u_x\cos\theta,\ u_y' = u_x\sin\theta. \tag{7.131}$$

Similarly we can rotate a vector $\vec{w} = w_y\,\vec{j}$ lying along the y-axis about z-axis by angle θ anticlockwise to obtain

$$\vec{w}\,' = w_x'\,\vec{i} + w_y'\,\vec{j}, \quad w_x' = -w_y\sin\theta,\ w_y' = w_y\cos\theta. \tag{7.132}$$

These rotations are shown in the Figure 7.1.

Combining Eqs. (7.131) and (7.132) we can see that when rotated about the z-axis by an angle θ an arbitrary vector $\vec{v} = v_x\,\vec{i} + v_y\,\vec{j}$ in the x-y plane will become $\vec{v}\,' = v_x'\,\vec{i} + v_y'\,\vec{j}$ with their components related by

$$v_x' = v_x\cos\theta - v_y\sin\theta,\ \ v_y' = v_x\sin\theta + v_y\cos\theta. \tag{7.133}$$

The norm of the vector is preserved for such a rotation, i.e.,

$$||\vec{v}\,|| = \sqrt{v_x^2 + v_y^2} = \sqrt{v_x'^2 + v_y'^2} = ||\vec{v}\,'||. \tag{7.134}$$

[22] As shown in Eq. (7.112), there is no guarantee that the eigenvalues and eigenvectors of a real matrix $\boldsymbol{R}$ are real.

[23] Ballentine pp. 132–133 for the concept of active and passive rotations.

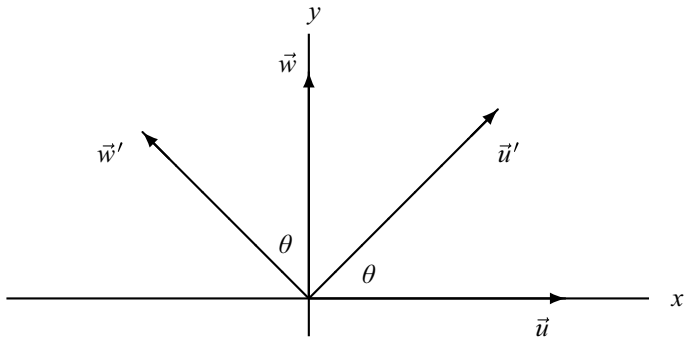


Figure 7.1 Vectors on the x-y plane and their rotations.

These rotations can be described in terms of column vectors and matrices:

(a) We can represent $\vec{v} = v_x \vec{i} + v_y \vec{j}$ and $\vec{v}' = v'_x \vec{i} + v'_y \vec{j}$ by column vectors

$$\boldsymbol{C}_{\vec{v}} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}, \quad \boldsymbol{C}_{\vec{v}'} = \begin{pmatrix} v'_x \\ v'_y \end{pmatrix} \tag{7.135}$$

in basis $\{\vec{i}, \vec{j}\}$.

(b) Introduce the following family of 2×2 matrices[24]

$$\boldsymbol{R}_p(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in [0, 2\pi]. \tag{7.136}$$

These matrices satisfy the condition in Eq. (7.122), and are hence orthogonal. We can see that the two column vectors which make up $\boldsymbol{R}_p(\theta)$ are orthonormal. These orthogonal matrices can be used to describe rotations of $\vec{v}$ in that the column vectors representing $\vec{v}$ and the rotated vector $\vec{v}\,'$ are related by

$$\boldsymbol{C}_{\vec{v}'} = \boldsymbol{R}_p(\theta)\boldsymbol{C}_{\vec{v}}. \tag{7.137}$$

Explicitly we have

$$\begin{pmatrix} v'_x \\ v'_y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix}. \tag{7.138}$$

[24] Croft and Davison p. 465. The subscript signifies proper rotation. These are the matrices in Eq. (7.110). Note that $\boldsymbol{R}_p(0) = \boldsymbol{R}_p(2\pi) = \boldsymbol{I}$, i.e., no rotation.

Rotations as described above are called *proper rotations*. The matrix for a proper rotation has determinant 1.

(2) *Improper rotations* It can be shown that there is just one other family of 2×2 orthogonal matrices. This family consists of the following matrices

$$\boldsymbol{R}_i(\theta) := \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}. \tag{7.139}$$

These matrices do not correspond to rotations as depicted in Figure 7.7.2. For $\theta = 0$ the above matrix reduces to

$$\boldsymbol{R}_{rx} := \boldsymbol{R}_i(0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.140}$$

For $\theta = \pi$ the matrix reduces to

$$\boldsymbol{R}_{ry} := \boldsymbol{R}_i(\pi) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7.141}$$

These matrices correspond to *reflections* of coordinate axes, i.e., $\boldsymbol{R}_{rx}$ represents a reflection about the x axis changing

$$\vec{v} = v_x\vec{i} + v_y\vec{j} \quad \text{to} \quad \vec{v}\,' = v_x\vec{i} - v_y\vec{j}, \tag{7.142}$$

and $\boldsymbol{R}_{ry}$ represents a reflection about the y axis changing

$$\vec{v} = v_x\vec{i} + v_y\vec{j} \quad \text{to} \quad \vec{v}\,' = -v_x\vec{i} + v_y\vec{j}. \tag{7.143}$$

Matrices in Eq. (7.141) are the products of a matrix for proper rotation $\boldsymbol{R}_p(\theta)$ and a matrix for coordinate reflection, i.e.,

$$\begin{aligned} \boldsymbol{R}_i(\theta) &:= \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{7.144}$$

This change due to matrix $\boldsymbol{R}_i(\theta)$ is called an *improper rotation*. Generally an improper rotation consists of a coordinate refection followed by a proper rotation. The matrix for an improper rotation has determinant -1.

(3) *The Pauli matrices* Being complex the Pauli matrix σ_y is not orthogonal. However, the other two Pauli matrices σ_x and σ_z are orthogonal, on account of Eqs. (7.51) to (7.53). They correspond to improper rotations, i.e., $\sigma_x = \boldsymbol{R}_i(\pi/2)$ which would swop the x and y axes, and $\sigma_z = \boldsymbol{R}_i(0) = \boldsymbol{R}_{rx}$.

(4) *Eigenvalues and eigenvectors* As shown in Eq. (7.112) the matrix $\boldsymbol{R}_p(\theta)$ for proper rotations does not possess any real eigenvalue or real eigenvectors. Geometrically this is because that no vector in the x-y plane is left unchanged by a rotation. On the other hand an improper rotation matrix, e.g., the Pauli matrices σ_x and σ_z, can have real eigenvalues and eigenvectors.

2. Orthogonal matrices of order 3

(1) *Proper rotations* When rotated by an angle θ_z about the z-axis a vector $\vec{v}$ changes to $\vec{v}'$, i.e.,

$$\vec{v} = v_x\vec{i} + v_y\vec{j} + v_z\vec{k} \quad \rightarrow \quad \vec{v}\,' = v'_x\vec{i} + v'_y\vec{j} + v'_z\vec{k}, \tag{7.145}$$

$$v'_x = v_x\cos\theta_z - v_y\sin\theta_z, \tag{7.146}$$

$$v'_y = v_x\sin\theta_z + v_y\cos\theta_z, \tag{7.147}$$

$$v'_z = v_z. \tag{7.148}$$

Such a rotation can be represented by the 3×3 matrix

$$\boldsymbol{R}_z(\theta_z) := \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{7.149}$$

in the sense of the following matrix equation

$$\boldsymbol{C}_{\vec{v}'} = \boldsymbol{R}_z(\theta_z)\boldsymbol{C}_{\vec{v}}. \tag{7.150}$$

Explicitly we have

$$\begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} v'_x \\ v'_y \\ v'_z \end{pmatrix}. \tag{7.151}$$

We can perform rotations about the x axis by an angle θ_x and the y axis by an angle θ_y in a similar manner, i.e., we have

$$\boldsymbol{C}_{\vec{v}''} = \boldsymbol{R}_x(\theta_x)\,\boldsymbol{C}_{\vec{v}}, \quad \boldsymbol{C}_{\vec{v}'''} = \boldsymbol{R}_y(\theta_y)\,\boldsymbol{C}_{\vec{v}}. \tag{7.152}$$

where the matrices are

$$\boldsymbol{R}_x(\theta_x) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix}, \tag{7.153}$$

$$\boldsymbol{R}_y(\theta_y) := \begin{pmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{pmatrix}. \tag{7.154}$$

These are the *proper rotations*. Rotations do not have to be about the three coordinate axes. We can have a rotation about any chosen axis. The determinants of all these rotation matrices are equal to 1. It can be shown that every 3×3 orthogonal matrix of determinant 1 generates a proper rotation about an axis. Coordinate axes can be chosen so that the matrix would appear in form of, say, $\boldsymbol{R}_x(\theta_x)$.[25]

(2) *Improper rotations* We also have improper rotations involving reflections about coordinate axes, e.g., the following matrix causes the reflection of all three coordinate axes:

$$\boldsymbol{R}_{rxyz} := \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{7.155}$$

Generally a 3×3 orthogonal matrix for an improper rotation consists of a proper rotation followed by coordinate reflection. These orthogonal matrices has determinant -1.

(3) *Eigenvalues and eigenvectors* A proper rotation matrix of order 3 can have a real eigenvalue. Geometrically we can see that a vector $\vec{v}_z$ lying in the z-axis is left unchanged by a rotation about the z-axis, i.e.,

$$\boldsymbol{R}_z(\theta_z)\boldsymbol{C}_{\vec{v}_z} = \boldsymbol{C}_{\vec{v}_z}. \tag{7.156}$$

This means that $\boldsymbol{R}_z(\theta_z)$ admits $\boldsymbol{C}_{\vec{v}_z}$ as an eigenvector corresponding to eigenvalue 1. In contrast the improper rotation matrix $\boldsymbol{R}_{rxyz}$ has eigenvalue -1.

[25] Fano p. 103.

7.7.3 Unitary Matrices

Definition 7.7.3(1) *A square matrix* $\boldsymbol{U}$ *is said to be unitary if it is invertible and its inverse is equal to its adjoint, i.e.,* $\boldsymbol{U}^{-1} = \boldsymbol{U}^{\dagger}$.

Unitary matrices are square matrices satisfying

$$\boldsymbol{U}^{\dagger}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^{\dagger} = \boldsymbol{I}. \tag{7.157}$$

On account of Eq. (7.71) this is reduced further to the statement that unitary matrices are square matrices satisfying[26]

$$\boldsymbol{U}^{\dagger}\boldsymbol{U} = \boldsymbol{I}. \tag{7.158}$$

Note that if $\boldsymbol{U}$ is unitary then $\boldsymbol{U}^{\dagger}$ is also unitary.

Unitary matrices are a generalisation of the concept of orthogonal matrices to complex matrices. Real unitary matrices are the same as orthogonal matrices since the adjoint of a real matrix is the same as its transpose. The examples below serve to illustrate our discussion:

E7.7.3(1) The following complex diagonal matrix is unitary:

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}, \quad \theta \in (0, 2\pi), \tag{7.159}$$

E7.7.3(2) Pauli matrices are unitary, on account of Eqs. (7.47) to (7.49) and Eq. (7.62). Being real σ_x and σ_z are also orthogonal while σ_y is not.

Unitary matrices preserve the scalar product and norm of complex column vectors, e.g., if $\boldsymbol{U}$ be an $n \times n$ unitary matrix, $\boldsymbol{C}$, $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are $n \times 1$ complex column vectors with their scalar products given by Eq. (7.73) then we have

$$\langle \boldsymbol{U}\boldsymbol{C} \mid \boldsymbol{U}\boldsymbol{C} \rangle = \langle \boldsymbol{C} \mid \boldsymbol{C} \rangle, \quad \langle \boldsymbol{U}\boldsymbol{C}_1 \mid \boldsymbol{U}\boldsymbol{C}_2 \rangle = \langle \boldsymbol{C}_1 \mid \boldsymbol{C}_2 \rangle. \tag{7.160}$$

We can introduce unitary transformations in analogy to orthogonal transformations:

(1) The column vector $\boldsymbol{C}' := \boldsymbol{U}\boldsymbol{C}$ is called the **unitary transform** of $\boldsymbol{C}$ *generated* by the unitary matrix $\boldsymbol{U}$. The properties in Eq. (7.160) are so characteristic of unitary matrices that it can be used to define unitary matrices, i.e., a complex matrix satisfying Eq. (7.160) is unitary.

[26] Hohn p. 251. Halmos p. 142.

(2) The matrix $\boldsymbol{M}' := \boldsymbol{U}\boldsymbol{M}\boldsymbol{U}^\dagger$ is called the **unitary transform** of an $n \times n$ matrix $\boldsymbol{M}$ *generated* by the unitary matrix $\boldsymbol{U}$. Multiplying $\boldsymbol{M}'$ on the left by $\boldsymbol{U}^\dagger$ and on the right by $\boldsymbol{U}$ we get $\boldsymbol{M} = \boldsymbol{U}^\dagger\boldsymbol{M}\boldsymbol{U}$.

(3) A **simultaneous unitary transformation** of both $\boldsymbol{C}$ and $\boldsymbol{M}$ preserves the value of $\langle \boldsymbol{C} \mid \boldsymbol{M}\boldsymbol{C} \rangle$, i.e.,

$$\boldsymbol{C}' := \boldsymbol{U}\boldsymbol{C}, \quad \boldsymbol{M}' := \boldsymbol{U}\boldsymbol{M}\boldsymbol{U}^\dagger \tag{7.161}$$

$$\Rightarrow \quad \langle \boldsymbol{C}' \mid \boldsymbol{M}'\boldsymbol{C}' \rangle = \langle \boldsymbol{C} \mid \boldsymbol{M}\boldsymbol{C} \rangle. \tag{7.162}$$

We conclude our discussion with the remark that the eigenvalues of a unitary matrix may be real or complex, e.g., the Pauli matrices have real eigenvalues ± 1 and the matrix in Eq. (7.159) has complex eigenvalues $\exp(i\theta)$ and $\exp(-i\theta)$.[27] The common feature is that an eigenvalue of a unitary matrix must have absolute value 1, a result followed from Eq. (7.160). It can be shown that a unitary matrix possesses n orthonormal eigenvectors which may be complex.[28]

7.7.4 Selfadjoint Matrices

Definition 7.7.4(1)[29] *A square matrix $\boldsymbol{M}$ is said to be selfadjoint if it is equal to its adjoint, i.e., $\boldsymbol{M} = \boldsymbol{M}^\dagger$.*

On account of Eq. (7.81) we see that a selfadjoint $n \times n$ matrix $\boldsymbol{M}$ satisfies the following equation

$$\langle \boldsymbol{C}_1 \mid \boldsymbol{M}\boldsymbol{C}_2 \rangle = \langle \boldsymbol{M}\boldsymbol{C}_1 \mid \boldsymbol{C}_2 \rangle, \tag{7.163}$$

for all $n \times 1$ column vectors $\boldsymbol{C}_1, \boldsymbol{C}_2$. We call this a **selfadjointness condition** since this condition can be used to define selfadjoint matrices, i.e., a matrix satisfying this condition is selfadjoint.

An important property of selfadjoint matrices, obvious on account of Eq. (7.74), is stated in the following theorem.

Theorem 7.7.4(1) *Let $\boldsymbol{M}$ be an $n \times n$ selfadjoint matrix and let $\boldsymbol{C}$ be a possibly complex $n \times 1$ column vector. Then the scalar product of*

[27] Recall E7.6(1) which says the eigenvalues of a diagonal matrix are its diagonal elements.

[28] Fano pp. 96–97.

[29] Selfadjoint matrices are often referred to as *Hermitian matrices*. For real matrices the selfadjointness requirement reduces to the symmetry condition.

$\boldsymbol{C}$ and $\boldsymbol{MC}$ is real, i.e.,

$$\langle \boldsymbol{C} \mid \boldsymbol{MC} \rangle \in I\!R. \tag{7.164}$$

Selfadjoint matrices do not have the geometric meaning of orthogonal and unitary matrices, but they are indispensable in physical applications on account of the nature of their eigenvalues and eigenvectors. The eigenvalues of orthogonal and unitary matrices are not very useful since they may be complex and they all have absolute value 1. The eigenvalues and the eigenvectors of a selfadjoint matrix possess some very distinctive and useful properties. We shall list these properties in the theorems below.

Theorem 7.7.4(2)[30]

(1) *The eigenvalues of a selfadjoint matrix are real.*
(2) *The eigenvectors of a selfadjoint matrix corresponding to different eigenvalues are orthogonal to each other.*

Proof Let $\boldsymbol{M}$ be an $n \times n$ selfadjoint matrix, and let λ_ℓ be an eigenvalue corresponding to normalised eigenvector $\boldsymbol{C}_\ell$.[31] Then we have the eigenvalue equation $\boldsymbol{MC}_\ell = \lambda_\ell \boldsymbol{C}_\ell$. Forming the scalar product of this equation and $\boldsymbol{C}_\ell$ from the left we get

$$\langle \boldsymbol{C}_\ell \mid \boldsymbol{MC}_\ell \rangle = \langle \boldsymbol{C}_\ell \mid \lambda_\ell \boldsymbol{C}_\ell \rangle = \lambda \langle \boldsymbol{C}_\ell \mid \boldsymbol{C}_\ell \rangle. \tag{7.165}$$

Since $\langle \boldsymbol{C}_\ell \mid \boldsymbol{MC}_\ell \rangle$ is real by Theorem 7.7.4 (1) we can conclude that λ is real. Taking scalar product of the eigenvalue equation and another eigenvector $\boldsymbol{C}_{\ell'}$ corresponding to eigenvalue $\lambda_{\ell'}$ on both sides we get

$$\text{RHS} = \langle \boldsymbol{C}_{\ell'} \mid \boldsymbol{MC}_\ell \rangle = \lambda_\ell \langle \boldsymbol{C}_{\ell'} \mid \boldsymbol{C}_\ell \rangle, \tag{7.166}$$

$$\text{LHS} = \langle \boldsymbol{C}_{\ell'} \mid \boldsymbol{MC}_\ell \rangle = \langle \boldsymbol{MC}_{\ell'} \mid \boldsymbol{C}_\ell \rangle = \lambda_{\ell'} \langle \boldsymbol{C}_{\ell'} \mid \boldsymbol{C}_\ell \rangle, \tag{7.167}$$

$$\text{LHS} = \text{RHS} \;\Rightarrow\; \langle \boldsymbol{C}_{\ell'} \mid \boldsymbol{C}_\ell \rangle = 0 \quad \text{if } \lambda_\ell \neq \lambda_{\ell'}. \tag{7.168}$$

QED

Theorem 7.7.4(3)[32] *Unitary transformations preserve the eigenvalues of selfadjoint matrices.*

[30] This theorem applies to real as well as complex selfadjoint matrices.
[31] We can always choose these eigenvectors to be normalised.
[32] Hohn p. 296.

Proof Let λ be an eigenvalue of an $n \times n$ selfadjoint matrix $\boldsymbol{M}$ corresponding to eigenvector $\boldsymbol{C}$. Let $\boldsymbol{M}'$ and $\boldsymbol{C}'$ be the unitary transforms of $\boldsymbol{M}$ and $\boldsymbol{C}$ generated by an $n \times n$ unitary matrix $\boldsymbol{U}$. Then:

$$\boldsymbol{M}'\boldsymbol{C}' = \left(\boldsymbol{U}\,\boldsymbol{M}\boldsymbol{U}^\dagger\right)\left(\boldsymbol{U}\,\boldsymbol{C}\right) = \boldsymbol{U}\,\boldsymbol{M}\boldsymbol{C} = \lambda\boldsymbol{C}'.$$

It follows that an eigenvalue of $\boldsymbol{M}$ is an eigenvalue of $\boldsymbol{M}'$. **QED**

Theorem 7.7.4(4)[33] *Given an $n \times n$ selfadjoint matrix $\boldsymbol{M}$ there exists an $n \times n$ unitary matrix $\boldsymbol{U}$ such that the unitary transform $\boldsymbol{M}' = \boldsymbol{U}\ \boldsymbol{M}\ \boldsymbol{U}^\dagger$ is a diagonal matrix whose diagonal elements λ_ℓ are the eigenvalues of $\boldsymbol{M}$.*

Such a transformation is known as the *diagonalisation* of the selfadjoint matrix.

The theorem implies that an $n \times n$ selfadjoint matrix would have n eigenvectors. To show this we first note that the $n \times n$ diagonal matrix $\boldsymbol{M}'$ in the theorem have n orthonormal eigenvectors $\boldsymbol{C}'_\ell$ with the diagonal elements λ_ℓ as eigenvalues.[34] These eigenvectors lead to n new vectors $\boldsymbol{C}_\ell = \boldsymbol{U}^\dagger\boldsymbol{C}'_\ell$. These new vectors are the eigenvectors of $\boldsymbol{M}$. To prove this we first observe that $\boldsymbol{M} = \boldsymbol{U}^\dagger\boldsymbol{M}'\boldsymbol{U}$. It follows that

$$\boldsymbol{M}\boldsymbol{C}_\ell = \left(\boldsymbol{U}^\dagger\boldsymbol{M}'\boldsymbol{U}\right)\boldsymbol{C}_\ell = \boldsymbol{U}^\dagger\boldsymbol{M}'\boldsymbol{C}'_\ell \tag{7.169}$$

$$= \boldsymbol{U}^\dagger\lambda_\ell\boldsymbol{C}'_\ell = \lambda_\ell\boldsymbol{C}_\ell. \tag{7.170}$$

We can conclude that the original matrix $\boldsymbol{M}$ also possesses n independent eigenvectors. This important result is stated below.

Theorem 7.7.4(5)[35] *The eigenvectors an $n \times n$ selfadjoint matrix can be chosen to form an orthonormal basis for the set of $n \times 1$ column vectors.*

The realness of the eigenvalues and the scalar product $\langle \boldsymbol{C} \mid \boldsymbol{M}\boldsymbol{C} \rangle$ and the complete orthonormal nature of the eigenvectors are crucial in the applications of selfadjoint matrices in quantum mechanics.

---

[33] Hohn p. 296.
[34] *See* the examples in Eqs. (7.106) and (7.107).
[35] Hohn p. 299. Kreyszig p. 350.

As examples consider the Pauli matrices. We can verified that:

(1) Pauli matrices are unitary.

(2) Pauli matrices are selfadjoint.

(3) Pauli matrices possess only real eigenvalues, i.e., ± 1.

(4) The eigenvectors of each Pauli matrix corresponding to the two different eigenvalues, given explicitly by Eqs. (7.113), (7.114) and (7.115), are orthonormal. The two eigenvectors for the matrix σ_z span the set of all 2×1 column vectors, and hence they form an orthonormal basis in the set of all 2×1 column vectors. The same applies to the eigenvectors of σ_x and σ_y.

(5) Pauli matrices satisfy Theorems 7.7.4(1), 7.7.4(2), 7.7.4(4) and 7.7.4(5).

7.7.5 Projection Matrices

Definition 7.7.5(1)[36] *A square matrix $\boldsymbol{M}$ is a projection matrix if it is selfadjoint and idempotent, i.e.,*

$$\boldsymbol{M} = \boldsymbol{M}^\dagger \quad \text{and} \quad \boldsymbol{M} = \boldsymbol{M}^2. \tag{7.171}$$

Geometrically these matrices correspond to projection operations of vectors. Consider vectors in $\vec{I\!E}^3$. In basis $\{\vec{i}, \vec{j}, \vec{k}\}$ a vector $\vec{v}$ has a matrix representation $\boldsymbol{C}_{\vec{v}}$ as shown in Eq. (7.84). The projection $\vec{v}_{\vec{i}}$ of $\vec{v}$ onto $\vec{i}$ is given by Eq. (6.42). Let $\boldsymbol{C}_{\vec{v}_{\vec{i}}}$ be the matrix representation of $\vec{v}_{\vec{i}}$ in basis $\{\vec{i}, \vec{j}, \vec{k}\}$. Then the column vectors $\boldsymbol{C}_{\vec{v}}$ and $\boldsymbol{C}_{\vec{v}_{\vec{i}}}$ are related by a matrix $\boldsymbol{P}_{\vec{i}}$, i.e.,

$$\boldsymbol{C}_{\vec{v}_{\vec{i}}} = \boldsymbol{P}_{\vec{i}}\boldsymbol{C}_{\vec{v}} \quad \Leftrightarrow \quad \begin{pmatrix} v_x \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}. \tag{7.172}$$

We call $\boldsymbol{P}_{\vec{i}}$ the *projection matrix onto $\vec{i}$ in basis* $\{\vec{i}, \vec{j}, \vec{k}\}$. It possesses the two defining properties of projection matrices, i.e., selfadjointness $\boldsymbol{P}_{\vec{i}} = \boldsymbol{P}_{\vec{i}}^\dagger$ and idempotence $\boldsymbol{P}_{\vec{i}} = \boldsymbol{P}_{\vec{i}}^2$.

[36] Projection matrices defined here are also referred to as *orthogonal projection matrices* in mathematics literature where non-orthogonal projection matrices are introduced by their idempotent property without the selfadjointness requirement.

Next consider the projection $\vec{v}_{\vec{e}}$ of $\vec{v}$ onto an arbitrary unit vector $\vec{e}$ given by Eq. (6.47). Let $\boldsymbol{C}_{\vec{v}_{\vec{e}}}$ and $\boldsymbol{C}_{\vec{e}}$ be the matrix representations of $\vec{v}_{\vec{e}}$ and $\vec{e}$ in basis $\{\vec{i}, \vec{j}, \vec{k}\}$. Then

$$\boldsymbol{C}_{\vec{v}_{\vec{e}}} = \langle \vec{e} \mid \vec{v} \rangle \boldsymbol{C}_{\vec{e}}. \tag{7.173}$$

The matrix representation of this projection operation is

$$\boldsymbol{P}_{\vec{e}} := \boldsymbol{C}_{\vec{e}} \boldsymbol{C}_{\vec{e}}^{\dagger}, \tag{7.174}$$

since, on account of Eq. (7.73), we have[37]

$$\begin{aligned} \boldsymbol{P}_{\vec{e}} \boldsymbol{C}_{\vec{v}} &= \left(\boldsymbol{C}_{\vec{e}} \boldsymbol{C}_{\vec{e}}^{\dagger}\right) \boldsymbol{C}_{\vec{v}} = \boldsymbol{C}_{\vec{e}} \left(\boldsymbol{C}_{\vec{e}}^{\dagger} \boldsymbol{C}_{\vec{v}}\right) \\ &= \langle \boldsymbol{C}_{\vec{e}} \mid \boldsymbol{C} \rangle \boldsymbol{C}_{\vec{e}} = \langle \vec{e} \mid \vec{v} \rangle \boldsymbol{C}_{\vec{e}} = \boldsymbol{C}_{\vec{v}_e}. \end{aligned} \tag{7.175}$$

This matrix is selfadjoint and idempotent, i.e., it is a projection matrix known as the *projection matrix generated by the column vector* $\boldsymbol{C}_{\vec{e}}$. We shall employ the *Dirac notation* for projection matrices, i.e., we shall express this projection matrix as

$$\boldsymbol{P}_{\vec{e}} = |\boldsymbol{C}_{\vec{e}}\rangle\langle \boldsymbol{C}_{\vec{e}}|, \tag{7.176}$$

with the understanding that its action on any column vector is given by Eq. (7.175), i.e.,[38]

$$\boldsymbol{P}_{\vec{e}} \boldsymbol{C}_{\vec{v}} = \left(|\boldsymbol{C}_{\vec{e}}\rangle\langle \boldsymbol{C}_{\vec{e}}|\right) \boldsymbol{C}_{\vec{v}} = \langle \boldsymbol{C}_{\vec{e}} \mid \boldsymbol{C}_{\vec{v}} \rangle \boldsymbol{C}_{\vec{e}}. \tag{7.177}$$

As illustrations we can calculate the projection matrices generated by the eigenvectors of the Pauli matrices in Eqs. (7.113) to (7.118)[39]:

(1) From σ_x with eigenvectors $\boldsymbol{C}_{\vec{\alpha}_x}$ and $\boldsymbol{C}_{\vec{\beta}_x}$ we get

$$\boldsymbol{P}_{\vec{\alpha}_x} = \boldsymbol{C}_{\vec{\alpha}_x} \boldsymbol{C}_{\vec{\alpha}_x}^{\dagger} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{7.178}$$

$$\boldsymbol{P}_{\vec{\beta}_x} = \boldsymbol{C}_{\vec{\beta}_x} \boldsymbol{C}_{\vec{\beta}_x}^{\dagger} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{7.179}$$

[37] Matrix representation preserves scalar product due to Eq. (7.87).
[38] Here we do not attach any separate meaning to symbols $|\boldsymbol{C}_{\vec{v}}\rangle$ and $\langle \boldsymbol{C}_{\vec{v}}|$.
[39] Isham p. 91.

(2) From σ_y with eigenvectors $\boldsymbol{C}_{\vec{\alpha}_y}$ and $\boldsymbol{C}_{\vec{\beta}_y}$ we get

$$\boldsymbol{P}_{\vec{\alpha}_y} = \boldsymbol{C}_{\vec{\alpha}_y}\boldsymbol{C}^{\dagger}_{\vec{\alpha}_y} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}, \tag{7.180}$$

$$\boldsymbol{P}_{\vec{\beta}_y} = \boldsymbol{C}_{\vec{\beta}_y}\boldsymbol{C}^{\dagger}_{\vec{\beta}_y} = \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}. \tag{7.181}$$

(3) From σ_z with eigenvectors $\boldsymbol{C}_{\vec{\alpha}_z}$ and $\boldsymbol{C}_{\vec{\beta}_z}$ we get

$$\boldsymbol{P}_{\vec{\alpha}_z} = \boldsymbol{C}_{\vec{\alpha}_z}\boldsymbol{C}^{\dagger}_{\vec{\alpha}_z} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.182}$$

$$\boldsymbol{P}_{\vec{\beta}_z} = \boldsymbol{C}_{\vec{\beta}_z}\boldsymbol{C}^{\dagger}_{\vec{\beta}_z} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7.183}$$

The projection properties can be illustrated with these matrices, e.g., an arbitrary 2×1 column vector $\boldsymbol{C}$ can be projected onto $\boldsymbol{C}_{\vec{\alpha}_z}$ and $\boldsymbol{C}_{\vec{\alpha}_x}$ by $\boldsymbol{P}_{\vec{\alpha}_z}$ and $\boldsymbol{P}_{\vec{\alpha}_x}$, i.e.,

$$\boldsymbol{P}_{\vec{\alpha}_z}\boldsymbol{C} = \langle \boldsymbol{C}_{\vec{\alpha}_z} \mid \boldsymbol{C} \rangle \boldsymbol{C}_{\vec{\alpha}_z}, \quad \boldsymbol{P}_{\vec{\alpha}_x}\boldsymbol{C} = \langle \boldsymbol{C}_{\vec{\alpha}_x} \mid \boldsymbol{C} \rangle \boldsymbol{C}_{\vec{\alpha}_x}. \tag{7.184}$$

Explicitly we have

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = C_1\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{7.185}$$

$$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \frac{1}{2}(C_1 + C_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{7.186}$$

Theorem 7.7.5(1) *A projection matrix possesses only two eigenvalues, i.e.,* 0 *and* 1.

Proof Let a be an eigenvalue of a projection matrix $\boldsymbol{P}$ corresponding to an eigenvector $\boldsymbol{C}_a$, i.e., we have

$$\boldsymbol{P}\boldsymbol{C} = a\boldsymbol{C}. \tag{7.187}$$

Multiplying both sides by $\boldsymbol{P}$ we get

$$\boldsymbol{P}^2\boldsymbol{C} = a\boldsymbol{P}\boldsymbol{C} = a^2\boldsymbol{C}. \tag{7.188}$$

The idempotent property of $\boldsymbol{P}$ implies that $a = a^2$ which implies $a = 0$ or 1, bearing in mind that being selfadjoint the eigenvalues of $\boldsymbol{P}$ are real. **QED**

The projection matices in Eqs. (7.178) to (7.183) can be used to demonstrate the above theorem.

7.7.6 Spectral Decomposition of Selfadjoint Matrices

Given a selfadjoint matrix we can solve for its eigenvalues and eigenvectors. Conversely a selfadjoint matrix can be characterised by its eigenvalues and eigenvectors. Let $\boldsymbol{M}$ be an $n \times n$ selfadjoint matrix. The matrix possesses n orthonormal eigenvectors $\boldsymbol{C}_\ell$ together with their corresponding eigenvalues λ_ℓ.[40] Let $\boldsymbol{P}_\ell = |\boldsymbol{C}_\ell\rangle\langle\boldsymbol{C}_\ell|$ be the projection matrix generated by the eigenvector $\boldsymbol{C}_\ell$ in accordance with Eqs. (7.174) and (7.176). Then we have the following theorem.

Theorem 7.7.6(1) *A selfadjoint matrix $\boldsymbol{M}$ can be expressed in terms of its eigenvalues λ_ℓ and their associated the projection matrices $\boldsymbol{P}_\ell$ as*

$$\boxed{\boldsymbol{M} = \sum_{\ell=1}^{n} \lambda_\ell \boldsymbol{P}_\ell.} \tag{7.189}$$

This is known as the **spectral theorem** for selfadjoint matrices. The expression in Eq. (7.189) is referred to as the **spectral decomposition** of $\boldsymbol{M}$ and the set of eigenvalues is called the **spectrum** of $\boldsymbol{M}$. As illustrations we can check that the Pauli matrices have the following spectral decompositions:

$$\sigma_x = \boldsymbol{P}_{\vec{\alpha}_x} - \boldsymbol{P}_{\vec{\beta}_x}, \tag{7.190}$$

$$\sigma_y = \boldsymbol{P}_{\vec{\alpha}_y} - \boldsymbol{P}_{\vec{\beta}_y}, \tag{7.191}$$

$$\sigma_z = \boldsymbol{P}_{\vec{\alpha}_z} - \boldsymbol{P}_{\vec{\beta}_z}. \tag{7.192}$$

We can see that the spectral theorem tells us that a selfadjoint matrix is decomposable as a sum of projection matrices with real coefficients.

This spectral theorem can be extended to apply to operators acting on finite and infinite dimensional vector spaces, as will be discussed in §9.4.5, §15.3, and §20.2.

[40] While the n eigenvectors are distinct the eigenvalues are not necessarily different.

Exercises and Problems

Q7(1) Verify Eq. (7.37) on the trace of square matrices.

Q7(2) Show that the inverse of a square invertible matrix is unique.

Q7(3) Prove Eq. (7.71).

Q7(4) What is the inverse of a diagonal matrix with non-zero diagonal elements M_{jj}?

Q7(5) Show that the orthogonal matrix $\boldsymbol{R}_z(\theta_z)$ in Eq. (7.149) is not selfadjoint.

Q7(6) Prove Eq. (7.160).

Q7(7) Verify Eq. (7.162) on simultaneous unitary transformations.

Q7(8) Pauli matrices are denoted by $\sigma_x,\ \sigma_y,\ \sigma_z$.

- **(a)** Verify the six properties of Pauli matrices shown in Eqs. (7.41) to (7.49).
- **(b)** Show that $\left(a_x\sigma_x + a_y\sigma_y + a_z\sigma_z\right)^2 = \left(a_x^2 + a_y^2 + a_z^2\right)\boldsymbol{I}_{2\times 2}$, where $\boldsymbol{I}_{2\times 2}$ is the 2×2 identity matrix and $a_x,\ a_y,\ a_z$ are real numbers.
- **(c)** What are the determinant and trace of each of the Pauli matrices?
- **(d)** What is the inverse of each of Pauli matrices?
- **(e)** Verify that Pauli matrix σ_x satisfies the selfadjointness condition in Eq. (7.163) for the vectors $\boldsymbol{C}_{\vec{\alpha}_z}$ and $\boldsymbol{C}_{\vec{\beta}_z}$ in Eq. (7.118).
- **(f)** Show that the eigenvectors of each Pauli matrix corresponding to the eigenvalues ± 1 given by Eqs. (7.116), (7.117) and (7.118) are orthonormal.

Q7(9) Show that

- **(a)** If $\boldsymbol{P}$ is an $n \times n$ projection matrix then $\boldsymbol{I}_{n\times n} - \boldsymbol{P}$ is also a projection matrix.
- **(b)** If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are $n \times n$ projection matrices then $\boldsymbol{P} \cdot \boldsymbol{Q}$ is also a projection matrix if $\boldsymbol{P}$ and $\boldsymbol{Q}$ commute.

(c) If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are $n \times n$ projection matrices then $\boldsymbol{P} + \boldsymbol{Q}$ is also a projection matrix if $\boldsymbol{P} \cdot \boldsymbol{Q} = \boldsymbol{Q} \cdot \boldsymbol{P} = \boldsymbol{0}$, where $\boldsymbol{0}$ is the $n \times n$ zero matrix.

Q7(10) Show that the 2×2 matrices in Eqs. (7.178) to (7.183) are projection matrices.[41] Find their eigenvectors.

[41] Isham p. 91.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 8

Operations on Vectors in $\vec{I\!E}^3$

8.1 Functionals on $\vec{I\!E}^3$ and the Riesz Theorem

Mappings and functions discussed in §3.2.2 are very general concepts used to relate two sets of quantities. We shall now introduce a special class of mappings known as *functionals*. These are mappings F of the vector space $\vec{I\!E}^3$ into the set $I\!R$ of real numbers, i.e.,

$$F : I\!E^3 \to I\!R \quad \text{by} \quad \vec{u} \to F(\vec{u}) \in I\!R. \tag{8.1}$$

A functional associates every vector $\vec{u}$ in $\vec{I\!E}^3$ with a real number $F(\vec{u}) \in I\!R$. A functional F is said to be *linear* if

$$F(a\vec{u} + b\vec{v}) = aF(\vec{u}) + bF(\vec{v}) \tag{8.2}$$

for all a, b in $I\!R$ and for all $\vec{u}$, $\vec{v}$ in $I\!E^3$. An immediate consequence is that $F(\vec{0}) = 0$ if F is linear. We already have a mechanism to associate vectors with numbers, i.e., the scalar product on $\vec{I\!E}^3$ which also possesses linear property. The scalar product enables every vector $\vec{v}$ to generate a linear functional $F_{\vec{v}}$ on $\vec{I\!E}^3$ by

$$F_{\vec{v}}(\vec{u}) := \langle \vec{v} \mid \vec{u} \rangle \quad \forall \vec{u} \in I\!E^3. \tag{8.3}$$

The converse is also true as confirmed by the Riesz theorem.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

Theorem 8.1(1) Riesz Theorem[1] *Every linear functional F on the vector space $\vec{I\!E}^3$ is generated by a unique vector $\vec{v}$ by*

$$F(\vec{u}) = F_{\vec{v}}(\vec{u}) := \langle \vec{v} \mid \vec{u} \rangle. \tag{8.4}$$

Proof Let $\{\vec{e}_\ell\}$ be an orthonormal basis for $\vec{I\!E}^3$. We can express $\vec{u}$ in terms of the basis vectors as in Eq. (6.19), i.e.,

$$\vec{u} = \sum_\ell u_\ell \vec{e}_\ell, \quad u_\ell = \langle \vec{e}_\ell \mid \vec{u} \rangle. \tag{8.5}$$

Let F be a linear functional. Then we have, by linearity,

$$F(\vec{u}) = \sum_\ell u_\ell F(\vec{e}_\ell). \tag{8.6}$$

Define a vector $\vec{v}$ in terms of $F(\vec{e}_\ell)$ by

$$\vec{v} = \sum_\ell F(\vec{e}_\ell)\vec{e}_\ell. \tag{8.7}$$

We get, on ccount of Eq. (8.6),

$$\langle \vec{v} \mid \vec{u} \rangle = \sum_\ell F(\vec{e}_\ell)u_\ell = F(\vec{u}). \tag{8.8}$$

QED

8.2 Linear Operators

8.2.1 The Concept

We have already introduced two types of operations in previous discussions: an operation which changes the length of a given vector by scalar multiplication and an operation which changes the length as well as the direction of a given vector by projecting it onto another direction. These operations can be considered as mappings of $\vec{I\!E}^3$ into itself. In many applications we need to consider many different mappings of $\vec{I\!E}^3$ into itself. These mappings are describable by an equation, i.e.,

$$\widehat{A}: \ \vec{I\!E}^3 \to \vec{I\!E}^3 \quad \text{by} \quad \widehat{A}\,\vec{u} = \vec{v} \quad \forall \vec{u} \in \vec{I\!E}^3. \tag{8.9}$$

[1] Riesz (1880–1956) is a Hungarian mathematician. Scalar product in $\vec{I\!E}^3$ is real-valued.

The mapping is denoted by a letter with a "hat", i.e., $\widehat{A}$. We call these mappings *operators* since they represent operations on vectors. In the above equation $\vec{u}$ is called an **input vector** and $\vec{v}$ in called an **output vector**. As with functionals we are interested in operators which are *linear*.

Definition 8.2.1 (1) *An operator $\widehat{A}$ is said to be linear if it satisfies the following property:*

$$\widehat{A}(a\vec{u} + b\vec{v}) = a\widehat{A}\vec{u} + b\widehat{A}\vec{v} \quad \forall\, \vec{u}, \vec{v} \in \vec{I\!E}^3 \text{ and } \forall a, b \in I\!R. \quad (8.10)$$

Linear operator do not act on numbers. A number can be taken out and moved to the left of the operator. An immediate result is that $\widehat{A}\,\vec{0} = \vec{0}$. From now on all operators are assumed to be linear.[2]

8.2.2 General Definitions

8.2.2.1 Domain and range

The vector space $\vec{I\!E}^3$ on which $\widehat{A}$ acts is the *domain* of $\widehat{A}$. The set of images

$$\{\vec{v} : \vec{v} = \widehat{A}\vec{u},\ \vec{u} \in \vec{I\!E}^3\} \quad (8.11)$$

is the *range* of $\widehat{A}$, which is generally a subset of $\vec{I\!E}^3$. We shall denote the domain and the range of $\widehat{A}$ by $\vec{\mathcal{D}}(\widehat{A})$ and $\vec{\mathcal{R}}(\widehat{A})$, respectively. We then have

$$\vec{\mathcal{D}}(\widehat{A}) = \vec{I\!E}^3 \quad \text{and} \quad \vec{\mathcal{R}}(\widehat{A}) \subset \vec{I\!E}^3. \quad (8.12)$$

We shall call $\widehat{A}$ an operator on $\vec{I\!E}^3$.

The operator $\widehat{I\!I}$ which maps $\vec{I\!E}^3$ onto itself by $\widehat{I\!I}\,\vec{u} = \vec{u}\ \ \forall \vec{u} \in \vec{I\!E}^3$ is called the *identity operator* on $\vec{I\!E}^3$. The operator $\widehat{0}$ which maps every vector to the zero vector, i.e., $\widehat{0}\,\vec{u} = \vec{0}\ \ \forall \vec{u} \in \vec{I\!E}^3$, is called the *zero operator*.

Two operators $\widehat{A}$ and $\widehat{B}$ are equal, i.e., $\widehat{A} = \widehat{B}$, if and only if $\widehat{A}\vec{u} = \widehat{B}\vec{u}\ \ \forall \vec{u} \in \vec{I\!E}^3$.

[2] A list of definitions similar to and motivated by the corresponding definitions on mappings and on matrices is given in the next subsection.

8.2.2.2 Norm of operators

Consider the set of values $||\widehat{A}\vec{u}\,||$ for all unit vectors $\vec{u} \in \vec{I\!E}^3$, where $||\widehat{A}\vec{u}\,||$ is the norm of the vector $\widehat{A}\vec{u}$. The supremum of this set of values is defined to be the norm of the operator which is denoted by $||\widehat{A}\,||$, i.e.,[3]

$$||\widehat{A}\,|| := \sup\left\{||\widehat{A}\vec{u}\,||, \quad ||\vec{u}\,|| = 1\right\}. \tag{8.13}$$

An alternative expression is[4]

$$||\widehat{A}\,|| = \sup\left\{||\widehat{A}\vec{v}^{\,(u)}\,|| = \frac{||\widehat{A}\vec{v}\,||}{||\vec{v}||}, \quad \vec{v} \in \vec{I\!E}^3\right\}. \tag{8.14}$$

It follows that[5]

$$||\widehat{A}\vec{v}\,|| \leq ||\widehat{A}\,||\,||\vec{v}\,||. \tag{8.15}$$

8.2.2.3 Algebraic operations

Scalar multiplication, addition and multiplication of operators are defined by:

(1) *Scalar multiplication* $\widehat{A}' = a\widehat{A}$, $a \in I\!R$, is defined to be

$$\widehat{A}'\vec{u} := a(\widehat{A}\vec{u}\,) \quad \forall\vec{u} \in \vec{I\!E}^3. \tag{8.16}$$

(2) *Addition* $\widehat{C} = a\widehat{A} + b\widehat{B}$, $a, b \in I\!R$, is defined to be

$$\widehat{C}\vec{u} := a(\widehat{A}\vec{u}\,) + b(\widehat{B}\vec{u}\,) \quad \forall\vec{u} \in \vec{I\!E}^3. \tag{8.17}$$

(3) *Linear combination* A sum of operators of the form $a\widehat{A} + b\widehat{B} + \cdots$ is called a *linear combination* of the operators with coefficients $a, b, \cdots$.

(4) *Product* $\widehat{D} = (\widehat{A}\widehat{B}\,)$ is defined to be

$$\widehat{D}\vec{u} := \widehat{A}\big(\widehat{B}\vec{u}\,\big) \quad \forall\vec{u} \in \vec{I\!E}^3. \tag{8.18}$$

Note that $\widehat{A}$ does not act on $\widehat{B}$. It acts on the vector $\widehat{B}\vec{u}$. A simple example is the product of an operator with itself, i.e., $\widehat{A}^2 := \widehat{A}\widehat{A}$.

---

[3] The *supremum* of a set of real numbers is the least number that is greater than or equal to all numbers in the set. It is also called the *least upper bound*.

[4] Here $\vec{v}^{\,(u)}$ is the unit directional vector introduced by Eq. (6.17).

[5] Equation (8.15) applies to bounded operators in a complex scalar product space (*see* Eq. (17.7)).

(5) An operator $\widehat{A}$ is said to be *idempotent* if it is equal to its square, i.e., $\widehat{A}^2 = \widehat{A}$. A trivial example is the identity and the zero operators. As will be seen later, some very important non-trivial operators are also idempotent.

8.2.2.4 Commutators and anticommutators

The *commutator* of $\widehat{A}$ and $\widehat{B}$ is defined to be

$$[\,\widehat{A},\,\widehat{B}\,] := \widehat{A}\widehat{B} - \widehat{B}\widehat{A}. \tag{8.19}$$

An equation $[\,\widehat{A},\,\widehat{B}\,] = \widehat{C}$ is called a *commutation relation*. Two operators $\widehat{A}$, $\widehat{B}$ are said to *commute* if their commutator vanishes, i.e., if

$$[\,\widehat{A},\,\widehat{B}\,] = \widehat{0} \quad \text{or} \quad \widehat{A}\widehat{B} = \widehat{B}\widehat{A}. \tag{8.20}$$

In other words, two operators $\widehat{A}$, $\widehat{B}$ are said to commute if their multiplication is commutative, i.e., $\widehat{A}\widehat{B} = \widehat{B}\widehat{A}$.

Two operators are said to *anticommute* if their *anticommutator* defined to be

$$\{\widehat{A},\,\widehat{B}\,\} := \widehat{A}\widehat{B} + \widehat{B}\widehat{A} \tag{8.21}$$

vanishes.

8.2.2.5 Inverse operators

An operator $\widehat{A}$ is said to be *invertible* if it generates a one-to-one mapping of $\vec{I\!E}^3$ onto $\vec{I\!E}^3$, i.e., the range $\vec{\mathcal{R}}(\widehat{A})$ of $\widehat{A}$ coincides with $\vec{I\!E}^3$ and different input vectors are mapped into distinct output vectors. We can then define the inverse $\widehat{A}^{-1}$ of $\widehat{A}$ by Eq. (3.21), Explicitly, if $\widehat{A}\vec{u} = \vec{v}$, then the inverse $\widehat{A}^{-1}$ is defined by[6]

$$\widehat{A}^{-1}: \ \widehat{A}^{-1}\vec{v} := \vec{u} \quad \forall \vec{v} \in \vec{\mathcal{R}}(\widehat{A}). \tag{8.22}$$

It follows that

$$\widehat{A}^{-1}\widehat{A} = \widehat{I\!I} \quad \text{and} \quad \widehat{A}\widehat{A}^{-1} = \widehat{I\!I}. \tag{8.23}$$

[6] Halmos p. 62.

The invertibility of an operator $\widehat{A}$ is closely related to the requirement that the zero vector $\vec{0}$ is the only vector mapped into the zero vector by $\widehat{A}$, i.e.,

$$\widehat{A}\vec{u} = \vec{0} \quad \Rightarrow \quad \vec{u} = \vec{0}. \tag{8.24}$$

Firstly this requirement implies that the mapping generated by $\widehat{A}$ is one-to-one. To prove this we first observe that

$$\vec{u} \neq \vec{0} \;\Rightarrow\; \widehat{A}\vec{u} \neq \vec{0}. \tag{8.25}$$

Then:

$$\begin{aligned} \vec{u}_1 \neq \vec{u}_2 \quad &\Rightarrow \quad \vec{u}_1 - \vec{u}_2 \neq \vec{0} \\ &\Rightarrow \quad \widehat{A}\Big(\vec{u}_1 - \vec{u}_2\Big) \neq \vec{0} \\ &\Rightarrow \quad \widehat{A}\vec{u}_1 \neq \widehat{A}\vec{u}_2. \end{aligned} \tag{8.26}$$

Secondly Eq. (8.24) also implies that $\widehat{A}$ generates an onto mapping, i.e., the range of the operator is the entire $\vec{I\!E}^3$. To prove this let $\{\vec{e}_\ell\}$ be an orthonormal basis for $\vec{I\!E}^3$ and let $\vec{e}_\ell{}' = \widehat{A}\vec{e}_\ell$, $\ell = 1, 2, 3$. These new vectors are linearly independent. To show this consider an arbitrary linear combination of $\vec{e}_\ell{}'$, i.e.,

$$\sum_{\ell=1}^{3} c_\ell\, \vec{e}_\ell{}' = \sum_{\ell-1}^{3} c_\ell\, \widehat{A}\vec{e}_\ell = \widehat{A}\left(\sum_{\ell=1}^{3} c_\ell\, \vec{e}_\ell\right). \tag{8.27}$$

It follows from Eq. (8.24) that

$$\sum_{\ell=1}^{3} c_\ell\, \vec{e}_\ell{}' = \vec{0} \;\Rightarrow\; \sum_{\ell=1}^{3} c_\ell\, \vec{e}_\ell = \vec{0} \;\Rightarrow\; c_\ell = 0 \;\; \forall \ell. \tag{8.28}$$

The last result is due to the linear independence of $\vec{e}_\ell$. This result in turn implies the linear independence of $\vec{e}_\ell{}'$. It then follows that $\vec{e}_\ell{}'$ form a complete set of $\vec{I\!E}^3$. Consequently every vector $\vec{v}$ in $\vec{I\!E}^3$ is expressible as a linear combination of $\vec{e}_\ell{}'$, i.e.,

$$\vec{v} = \sum_{\ell=1}^{3} a_\ell\, \vec{e}_\ell{}' = \sum_{\ell=1}^{3} a_\ell\, \widehat{A}\vec{e}_\ell = \widehat{A}\left(\sum_{\ell=1}^{3} a_\ell\, \vec{e}_\ell\right). \tag{8.29}$$

In other words, every vector $\vec{v}$ in $\vec{I\!E}^3$ is expressible in the form

$$\vec{v} = \widehat{A}\,\vec{u}, \quad \vec{u} = \sum_{\ell=1}^{3} u_\ell\, \vec{e}_\ell. \tag{8.30}$$

This means that every vector $\vec{v}$ in $\vec{I\!E}^3$ is the image of an input vector $\vec{u}$ under the operator $\widehat{A}$, i.e., the range of the operator is the entire space $\vec{I\!E}^3$.

Theorem 8.2.2(1) *An operator $\widehat{A}$ is invertible if and only if*[7]

$$\widehat{A}\vec{u} = \vec{0} \quad \Rightarrow \quad \vec{u} = \vec{0}. \tag{8.31}$$

This corresponds to Eq. (7.70) on the condition for inverse for matrices.[8] Definition 7.4(1) for matrices also applies to operators here as seen in the following theorem.

Theorem 8.2.2(2)[9] *An operator $\widehat{A}$ is invertible if there is another operator $\widehat{B}$ such that*

$$\widehat{B}\widehat{A} = \widehat{I\!I} = \widehat{A}\widehat{B}. \tag{8.32}$$

The inverse $\widehat{A}^{-1}$ is equal to the operator $\widehat{B}$.

For operators in $\vec{I\!E}^3$ either of the conditions $\widehat{B}\widehat{A} = \widehat{I\!I}$ and $\widehat{A}\widehat{B} = \widehat{I\!I}$ in Eq. (8.32) implies the other. In other words, $\widehat{B}\widehat{A} = \widehat{I\!I}$ is sufficient to imply the invertibility of $\widehat{A}$ with $\widehat{B}$ as its inverse.[10]

As with matrices an invertible operator has a unique inverse. We also have $\left(\widehat{A}^{-1}\right)^{-1} = \widehat{A}$ and $(\widehat{A}\widehat{B})^{-1} = \widehat{B}^{-1}\widehat{A}^{-1}$.[11]

8.2.2.6 Adjoint operators

The scalar product on $\vec{I\!E}^3$ enables us to define the notion of the *adjoint* of a given operator $\widehat{A}$ in the same way the adjoint of a matrix is definable by Eq. (7.81). An operator, denoted by $\widehat{A}^\dagger$, is called the *adjoint* of $\widehat{A}$, if the following condition is satisfied:

$$\langle \widehat{A}^\dagger \vec{v} \mid \vec{u} \rangle = \langle \vec{v} \mid \widehat{A}\vec{u} \rangle \ \ \forall\, \vec{u},\ \vec{v} \in \vec{I\!E}^3. \tag{8.33}$$

[7] Halmos pp. 62–63.

[8] This theorem also applies to operators in a finite-dimensional complex vector space discussed in Chapter 13. The situation is more complicated in an infinite-dimensional vector space discussed in §17.5.

[9] Halmos p. 62.

[10] Halmos pp. 64, 142. This theorem remains true in finite-dimensional complex vector spaces (*see* also Fano p. 74). As will be discussed in §17.5 this theorem is not true in an infinite-dimensional vector space.

[11] Halmos p. 63.

Using the commutative property SP6.3.1(1) of scalar product we can see that the adjoint operator also satisfies

$$\langle \vec{u} \mid \widehat{A}^\dagger \vec{v} \rangle = \langle \widehat{A}\vec{u} \mid \vec{v} \rangle \ \ \forall\, \vec{u},\ \vec{v} \ \in \vec{I\!E}^3. \tag{8.34}$$

The existence of the adjoint operator follows from Riesz theorem. Given a vector $\vec{w}$ we can define a linear functional $F^{\hat{A}}_{\vec{w}}$ on $\vec{I\!E}^3$ by

$$F^{\hat{A}}_{\vec{w}}(\vec{u}\,) := \langle \vec{w} \mid \widehat{A}\vec{u} \rangle \ \ \forall\, \vec{u} \ \in \vec{I\!E}^3. \tag{8.35}$$

Riesz theorem tells us that the functional $F^{\hat{A}}_{\vec{w}}$ must be generated by a vector in the form of Eq. (8.4), i.e., there exists a vector $\vec{w}'$ such that

$$F^{\hat{A}}_{\vec{w}}(\vec{u}\,) = \langle \vec{w}' \mid \vec{u}\,\rangle. \tag{8.36}$$

It follows that

$$\langle \vec{w}' \mid \vec{u}\,\rangle = \langle \vec{w} \mid \widehat{A}\vec{u}\,\rangle \ \ \forall \vec{u} \in \vec{I\!E}^3. \tag{8.37}$$

The vector $\vec{w}'$ is uniquely related to $\vec{w}$ because of Eq. (6.31). We can then define an operator $\widehat{A}^\dagger$ by $\widehat{A}^\dagger\vec{w} = \vec{w}'$. Then Eq. (8.37) reduces to Eq. (8.33). A trivial example is that the adjoint of the identity operator is equal to itself, i.e., $\widehat{I\!I}^\dagger = \widehat{I\!I}$.

The notation and properties of adjoint operators resemble that of matrices in §7.3. In particular Eqs. (7.58) to (7.61) for the adjoint operation for matrices apply to operators here, i.e., we have

(1) The adjoint operation is linear, i.e.,

$$(a\widehat{A} + b\widehat{B}\,)^\dagger = a\widehat{A}^\dagger + b\widehat{B}^\dagger, \quad a,\ b \in I\!R. \tag{8.38}$$

In particular we have, for any $a \in I\!R$,

$$(a\widehat{A}\,)^\dagger = a\widehat{A}^\dagger \quad \Rightarrow \quad \langle \vec{u} \mid a\widehat{A}\vec{v}\,\rangle = a\,\langle \widehat{A}^\dagger\vec{u} \mid \vec{v}\,\rangle. \tag{8.39}$$

(2) The adjoint of $\widehat{A}^\dagger$, denoted by $\widehat{A}^{\dagger\dagger}$, is equal to $\widehat{A}$, i.e.,

$$\widehat{A}^{\dagger\dagger} = (\widehat{A}^\dagger)^\dagger = \widehat{A}. \tag{8.40}$$

(3) The adjoint operation on the product of two operators is given by the product of the adjoints of the operators in the reverse order as in Eq. (7.61) for matrices, i.e.,

$$(\widehat{A}\widehat{B}\,)^\dagger = \widehat{B}^\dagger\widehat{A}^\dagger. \tag{8.41}$$

(4) The adjoint of the inverse is the inverse of the adjoint, i.e.,

$$(\widehat{A}^{-1}\,)^\dagger = (\widehat{A}^\dagger\,)^{-1}. \tag{8.42}$$

8.2.2.7 Quadratic form

An operator on its own does not generally have a numerical value. However, an operator can generate numerical values with the help of vectors. For example, we have a value for every vector $\vec{u}$ in $\vec{I\!E}^3$ defined by

$$\boxed{\mathcal{Q}(\widehat{A}, \vec{u}) := \langle \vec{u} \mid \widehat{A}\vec{u} \rangle.} \tag{8.43}$$

We call $\mathcal{Q}(\widehat{A}, \vec{u})$ the *quadratic form generated by the operator* $\widehat{A}$ *on the vector space* $\vec{I\!E}^3$. In §10 we shall discuss how a link between this quadratic form and the expectation values in a probability theory can emerge for certain types of operators.

8.2.3 Matrix Representation of Operators

An operator is defined by its effect on vectors through Eq. (8.9). It would be useful to have explicit expressions for operators in terms of familiar quantities. This can be achieved in terms of matrices. Since we can represent vectors by column vectors we should be able to represent operators by square matrices to acting on column vectors. A matrix expression of operators can be established as follows:

(1) Choose an orthonormal basis $\{\vec{e}_\ell\}$ to express the vectors $\vec{u}$ and $\vec{v}$ as done in Eqs. (7.82) and (7.83).

(2) Rewrite Eq. (8.9) explicitly as

$$\vec{v} = \sum_{\ell=1}^{3} v_\ell \vec{e}_\ell = \widehat{A}\vec{u} = \widehat{A} \sum_{\ell=1}^{3} u_\ell \, \vec{e}_\ell, = \sum_{\ell=1}^{3} u_\ell \, \widehat{A}\vec{e}_\ell. \tag{8.44}$$

(3) Taking the scalar product with $\vec{e}_k$ we get

$$v_k = \sum_{\ell=1}^{3} u_\ell \, \langle \vec{e}_k \mid \widehat{A}\vec{e}_\ell \rangle \quad \text{or} \quad \sum_{\ell=1}^{3} \langle \vec{e}_k \mid \widehat{A}\vec{e}_\ell \rangle u_\ell = v_k. \tag{8.45}$$

(4) Construct the matrix representation $\boldsymbol{C}_{\vec{u}}$ and $\boldsymbol{C}_{\vec{v}}$ of $\vec{u}$ and $\vec{v}$ in basis $\{\vec{e}_\ell\}$ in accordance with Eq. (7.84), i.e.,

$$\boldsymbol{C}_{\vec{u}} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \qquad \boldsymbol{C}_{\vec{v}} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \tag{8.46}$$

(5) Define a 3×3 matrix $\boldsymbol{M}_{\hat{A}}$ with elements

$$M_{\hat{A}k\ell} = \langle \vec{e}_k \mid \widehat{A}\vec{e}_\ell \rangle. \tag{8.47}$$

(6) We can now rewrite Eq. (8.45) as a matrix equation

$$\sum_{\ell=1}^{3} M_{\hat{A}k\ell}\, u_\ell = v_k \quad \text{or} \quad \boldsymbol{M}_{\hat{A}} \boldsymbol{C}_{\vec{u}} = \boldsymbol{C}_{\vec{v}}, \tag{8.48}$$

We can then regard Eq. (8.48) as the matrix representation of Eq. (8.9) in basis $\{\vec{e}_\ell\}$, and matrix $\boldsymbol{M}_{\hat{A}}$ as the matrix representation of $\widehat{A}$ in basis $\{\vec{e}_\ell\}$.

(7) With a different choice of basis we would obtain a different matrix to $\widehat{A}$ and different column vectors to represent $\vec{u}$ and $\vec{v}$.

(8) One can shown that

(a) The matrix representation of the inverse $\widehat{A}^{-1}$ of $\widehat{A}$ is the inverse matrix to the matrix $\boldsymbol{M}_{\hat{A}}$.

(b) The matrix representation of the adjoint of $\widehat{A}$ is the adjoint matrix to the matrix $\boldsymbol{M}_{\hat{A}}$.

Operators have many properties similar to that of matrices, e.g., we have operator eigenvalue problem and we also have special operators to correspond to the special matrices introduced in §7.7. A whole chapter, Chapter 9, will be devoted to discuss those special operators. We shall look into the eigenvalue problem of operators here first.

8.2.4 Eigenvalue Problem for Operators

Operators are abstract quantities which do not manifestly possess any numerical values. For physical applications we need numerical values. As with matrices operators can generate numerical values in the form of eigenvalues. Any operator, apart from the identity operator, will affect a vector. For an arbitrary input vector $\vec{u}$ in Eq. (8.9) the output vector $\vec{v}$ may be very different from $\vec{u}$, e.g., $\vec{v}$ is not linearly dependent on $\vec{u}$. As in the eigenvalue problem for matrices we look for input vectors which are least affected by the operator. An input vector $\vec{u}$ is deemed to be *least affected* by $\widehat{A}$ if the operator

changes $\vec{u}$ only to the extend of a multiplicative constant. For such input vectors we have

$$\widehat{A}\vec{u} = a\vec{u}, \quad a \in I\!R. \tag{8.49}$$

The operator only changes the length of $\vec{u}$ but not its *orientation* in the sense that the output vector has a length $|a|\,||\vec{u}\,||$ and it is either parallel or antiparallel to $\vec{u}$.[12] The following are standard definitions:

(1) Equation (8.49) is known as the **eigenvalue equation** for the operator $\widehat{A}$.

(2) We call $\vec{u}$ in Eq. (8.49) an **eigenvector** of $\widehat{A}$ and a the **eigenvalue** corresponding to the eigenvector $\vec{u}$. We also call a an eigenvalue of $\widehat{A}$ and $\vec{u}$ an eigenvector of $\widehat{A}$ corresponding to the eigenvalue a. An operator may have different eigenvalues and an eigenvalue may correspond to different eigenvectors.[13]

(3) As for matrices an eigenvalue may be *degenerate* or *nondegenerate*.

   (a) If $\vec{u}$ is an eigenvector corresponding to an eigenvalue a then $b\vec{u}$, $b \in I\!R$, is another eigenvector corresponding to the eigenvalue a. However, we do not regard $b\vec{u}$ to be a different eigenvector. We consider two eigenvectors to be different only if they are linearly independent.

   (b) It is possible for an eigenvalue to correspond to two or more linearly independent eigenvectors. The eigenvalue is then said to be *degenerate* and the number of linearly independent eigenvectors corresponding to the eigenvalue is called the *degeneracy* of the eigenvalue. This is equivalent to saying that not all the eigenvalues have to be different. An eigenvalue of degeneracy 1 is said to be *nondegenerate*.

   (c) An operator is said to be *nondegenerate* if all its eigenvalues are nondegenerate. For a nondegenerate operator all its eigenvalues are different.

[12] We use the term *orientation* to mean both the parallel and antiparallel directions of a vector.

[13] As will be commented later the existence of eigenvectors for operators is related to the existence of eigenvectors for matrices.

(4) The set of eigenvalues of $\widehat{A}$ is called the **spectrum** of $\widehat{A}$.

Since operators can be represented by matrices we can study the eigenvalue problem of operators through their corresponding matrix eigenvalue equations. This includes the existence and solutions of the eigenvalues. On account of Eq. (8.48) we can write down the following matrix representation of the operator eigenvalue equation (8.49) in a chosen orthonormal basis $\{\vec{e}_\ell\}$, i.e.,

$$\widehat{A}\vec{u} = a\vec{u} \quad \rightarrow \quad \boldsymbol{M}_{\widehat{A}}\boldsymbol{C}_{\vec{u}} = a\,\boldsymbol{C}_{\vec{u}}. \tag{8.50}$$

A different choice of basis will change the matrix for $\widehat{A}$ and the column vector for the eigenvector $\vec{u}$, but not the eigenvalue a. Hence the eigenvalues of an operator can be solved in terms of a corresponding matrix eigenvalue equation.

Exercises and Problems

Q8(1) Show that an invertible operator has a unique inverse.

Q8(2) Show that in $\vec{I\!E}^3$ the condition $\widehat{B}\widehat{A} = \widehat{I\!I}$ is sufficient to imply the invertibility of $\widehat{A}$ with $\widehat{B}$ as its inverse.

Q8(3) Let $\boldsymbol{A}$ be a matrix representation of an invertible operator $\widehat{A}$. Show that the matrix representation of the inverse $\widehat{A}^{-1}$ of $\widehat{A}$ is the inverse matrix $\boldsymbol{A}^{-1}$ to the matrix $\boldsymbol{A}$.

Q8(4) Let $\{\vec{e}_\ell, \ell = 1, 2, 3\}$ be a basis for $\vec{I\!E}^3$ and let $\widehat{A}$ be an invertible operator. Show that $\{\vec{e}\,'_\ell = \widehat{A}\vec{e}_\ell, \ell = 1, 2, 3\}$ is also a basis for $\vec{I\!E}^3$.[14]

Q8(5) Show that an operator $\widehat{A}$ is invertible if and only if every vector $\vec{v} \in \vec{I\!E}^3$ can be expressed as $\vec{v} = \widehat{A}\vec{u}$ for some $\vec{u} \in \vec{I\!E}^3$.[15]

Q8(6) Show that the matrix representation of the adjoint of $\widehat{A}$ is the adjoint matrix $\boldsymbol{M}^\dagger_{\widehat{A}}$ to the matrix $\boldsymbol{M}_{\widehat{A}}$.

[14] Halmos p. 63.
[15] Halmos pp. 62–63.

Q8(7) Let F be a linear functional on $\vec{I\!E}^3$. Show that $F(\vec{0}) = 0$. Prove a similar result for linear operators.[16]

Q8(8) Prove Eqs. (8.38) to (8.42) on adjoint operations.

Q8(9) Associated with the projection operation given by Eq. (6.41) we can define an operator $\widehat{P}_{\vec{i}}$ by

$$\widehat{P}_{\vec{i}}\,\vec{v} = v_x\,\vec{i} \quad \forall\,\vec{v} \in \vec{I\!E}^3. \tag{8.51}$$

Show that

(a) The operator $\widehat{P}_{\vec{i}}$ is idempotent.

(b) The operator $\widehat{P}_{\vec{i}}$ is selfadjoint.

(c) The operator possesses only two eigenvalues 0 and 1. Find the corresponding eigenvectors.

(d) The matrix representation $\boldsymbol{M}_{\hat{P}_{\vec{i}}}$ of $\widehat{P}_{\vec{i}}$ in basis $\{\vec{i}, \vec{j}, \vec{k}\}$ is

$$\boldsymbol{M}_{\hat{P}_{\vec{i}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{8.52}$$

Show also that $\boldsymbol{M}_{\hat{P}_{\vec{i}}}$ is a projection matrix. Find the eigenvalues and their corresponding eigenvectors of $\boldsymbol{M}_{\hat{P}_{\vec{i}}}$.

[16] Halmos pp. 20, 55.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 9

Special Operators on $\vec{I\!E}^3$

Here we shall examine in detail some special operators defined in close analogy to the special matrices introduced in §7.7.

9.1 Scalar Multiplication Operators

A simple operation is to lengthen or to shorten a vector by a factor a without changing its orientation. The simplest operator of this kind is the *identity operator* $\widehat{I\!I}$. It is often convenient to write $\widehat{I\!I} = 1$. A more general operation can be carried out by scalar multiplication representable by an operator $\widehat{M}_a$, i.e.,

$$\widehat{M}_a \, \vec{u} := a\vec{u} \quad \text{or} \quad \widehat{M}_a = a \, \widehat{I\!I}, \;\; a \in I\!R. \tag{9.1}$$

This is called a *scalar multiplication operator*. Here a can be negative. If we multiply a vector by a negative number we would reverse the direction of the vector, as well as changing its norm.

9.2 Rotations and Orthogonal Operators

9.2.1 Rotations of Vectors

Another operation is to change the direction of a vector but not its norm. This is the operation of rotation of the vector. For a rotation

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

about the z-axis by an angle θ_z the rotated vector $\vec{v}\,'$ and the original vector $\vec{v}$ are related by Eqs. (7.146) to (7.148). Such a rotation can be represented by an orthogonal matrix in Eq. (7.149). We can also define an operator $\widehat{R}_z(\theta_z)$ to relate $\vec{v}\,'$ and $\vec{v}$ by

$$\vec{v}\,' = \widehat{R}_z(\theta_z)\vec{v} \quad \forall \vec{v} \in \vec{I\!E}^3. \tag{9.2}$$

We can similarly introduce rotation operators $\widehat{R}_x(\theta_x)$ and $\widehat{R}_y(\theta_y)$ for rotations about the x and the y axes. In addition to proper rotations we can also define an operator to correspond to every improper rotation in the same way. As shown in Eq. (7.126) orthogonal matrices preserve the norm of column vectors. This property can be used to define orthogonal matrices. We can define *orthogonal operators* to correspond to orthogonal matrices in the same way.

9.2.2 Orthogonal Operators

Definition 9.2.2(1) *An operator $\widehat{R}$ on $\vec{I\!E}^3$ which preserves the norm of all the vectors in $\vec{I\!E}^3$, i.e.,*

$$\langle \widehat{R}\vec{v} \mid \widehat{R}\vec{v} \rangle = \langle \vec{v} \mid \vec{v} \rangle \quad \text{or} \quad ||\widehat{R}\vec{v}|| = ||\vec{v}|| \qquad \forall\, \vec{v} \in \vec{I\!E}^3 \tag{9.3}$$

is called an orthogonal operator on $\vec{I\!E}^3$.

Orthogonal operators possess the following properties:

P9.2.2(1) Orthogonal operators correspond to orthogonal matrices. Every 3×3 orthogonal matrix defines an orthogonal operator on $\vec{I\!E}^3$. For example, we can define:

(1) The operators for all the proper rotations discussed in §7.7.2, including rotations about any chosen axis through the coordinate origin.

(2) The operators for reflections about coordinate axes. For example, we have

$$\begin{aligned} \widehat{R}_{rx}\vec{v} &= \widehat{R}_{rx}\left(v_x\vec{i} + v_y\vec{j} + v_z\vec{k}\right) \\ &:= \left(-v_x\vec{i} + v_y\vec{j} + v_z\vec{k}\right). \end{aligned} \tag{9.4}$$

which causes a reflection of the x component of the vector, i.e., from v_x to $-v_x$, and

$$\begin{aligned}\widehat{R}_{rxyz}\vec{v} &= \widehat{R}_{rxyz}\left(v_x\vec{i} + v_y\vec{j} + v_z\vec{k}\right)\\ &:= \left(-v_x\vec{i} - v_y\vec{j} - v_z\vec{k}\right) = -\vec{v}.\end{aligned} \tag{9.5}$$

which cause a refection of all components of the vector.

Conversely orthogonal operators are represented by orthogonal matrices.

P9.2.2(2) Orthogonal operators are invertible.

P9.2.2(3) An operator on $\vec{I\!E}^3$ is orthogonal if and only if it is invertible and its inverse is equal to it adjoint, i.e., $\widehat{R}^\dagger = \widehat{R}^{-1}$.[1]

We can define orthogonal transformations by orthogonal operators in the same way orthogonal matrices defines orthogonal transforms of column vectors by Eq. (7.127).

Definition 9.2.2(2) *Let $\widehat{R}$ be an orthogonal operator on $\vec{I\!E}^3$.*

(1) *The vector $\vec{u}\,' := \widehat{R}\,\vec{u}$ is called the* **orthogonal transform** *of $\vec{u}$ generated by the orthogonal operator $\widehat{R}$.*

(2) *The operator $\widehat{A}' := \widehat{R}\widehat{A}\widehat{R}^\dagger$ is called the* **orthogonal transform** *of $\widehat{A}$ generated by the orthogonal operator $\widehat{R}$.*

Orthogonal transformations possess the following properties:

P9.2.2(4) An orthogonal transformation of vectors preserves scalar products, i.e.,

$$\vec{u}\,' := \widehat{R}\vec{u},\ \ \vec{v}\,' := \widehat{R}\vec{v}, \quad \Rightarrow \quad \langle\vec{u}\,' \mid \vec{v}\,'\rangle = \langle\vec{u} \mid \vec{v}\,\rangle. \tag{9.6}$$

P9.2.2(5) An orthogonal transformation preserves orthonormal bases, i.e., given an orthonormal basis $\{\vec{e}_\ell\}$ in $\vec{I\!E}^3$ their orthogonal transforms $\{\vec{e}_\ell{}'\}$ generated by an orthogonal operator $\widehat{R}$ is again an orthonormal basis in $\vec{I\!E}^3$.

[1] This corresponds to Definition 7.7.2(1) for matrices.

P9.2.2(6) For a *simultaneous orthogonal transformation* of vectors and operators we have:

(1) Preservation of the scalar product $\langle \vec{u} \mid \widehat{A}\vec{v} \rangle$, i.e.,

$$\vec{u}\,' := \widehat{R}\vec{u}, \quad \vec{v}\,' := \widehat{R}\vec{v}, \quad \widehat{A}' := \widehat{R}\widehat{A}\widehat{R}^{\dagger}$$
$$\Rightarrow \quad \langle \vec{u}\,' \mid \widehat{A}'\vec{v}\,' \rangle = \langle \vec{u} \mid \widehat{A}\vec{v} \rangle. \tag{9.7}$$

(2) Preservation of the quadratic form $\mathcal{Q}(\widehat{A}, \vec{u}\,)$ generated by $\widehat{A}$, i.e.,

$$\mathcal{Q}(\widehat{A}', \vec{u}\,') = \langle \vec{u}\,' \mid \widehat{A}'\vec{u}\,' \rangle = \langle \vec{u} \mid \widehat{A}\vec{u} \rangle = \mathcal{Q}(\widehat{A}, \vec{u}\,). \tag{9.8}$$

P9.2.2(7) Orthogonal operators on $\vec{I\!E}^3$ have a maximum of two eigenvalues, i.e., ± 1.[2] For example, the operator $\widehat{R}_z(\theta_z)$ defined by Eq. (9.2) admits only one real eigenvalue, i.e., the value 1 corresponding to eigenvectors of the form $\vec{v} = v_z\vec{k}$, while the operator $\widehat{R}_{rxyx}$ given by Eq. (9.5) admits only the eigenvalue -1. The operator $\widehat{R}_{rx}$ given by Eq. (9.4) admits eigenvalues ± 1. These eigenvalues are the same as their representative matrices.

9.3 Projections and Projection Operators

We can change both the norm and the direction of vectors. An example is a projection operation discussed in §6.4. Here we shall discuss the mathematical description of projection operation in terms of *projection operators*.

9.3.1 Projectors onto Unit Vectors

To be specific, consider the projection of $\vec{v}$ onto the x-axis, i.e., projection onto the unit vector $\vec{i}$ given by Definition 6.4.1(1). We can describe the projection operation from $\vec{v}$ to $\vec{v}_{\vec{i}}$ mathematically by an operator $\widehat{P}_{\vec{i}}$, i.e.,

$$\widehat{P}_{\vec{i}}\,\vec{v} := \vec{v}_{\vec{i}} \qquad \forall \vec{v} \in \vec{I\!E}^3, \tag{9.9}$$

[2] We confine ourselves to real eigenvalues and real eigenvectors. These quantities exist for $\widehat{R}$ on $\vec{I\!E}^3$. The situation is different in a two-dimensional space $\vec{I\!E}^2$ like the x-y plane. Orthogonal operators acting on $\vec{I\!E}^2$ corresponding to rotations given the matrix in Eq. (7.110) have no real eigenvalues.

or explicitly[3]

$$\widehat{P}_{\vec{i}}\,\vec{v} := v_x\,\vec{i} = \langle\,\vec{i}\mid\vec{v}\,\rangle\,\vec{i} \qquad \forall\vec{v}\in I\!\vec{E}^3. \tag{9.10}$$

We call $\widehat{P}_{\vec{i}}$ the **projection operator onto** $\vec{i}$, or the **projector** onto $\vec{i}$. The projectors on $\vec{j}$ and $\vec{k}$ are similarly defined by

$$\widehat{P}_{\vec{j}}\,\vec{v} := \vec{v}_{\vec{j}} = v_y\,\vec{j} = \langle\,\vec{j}\mid\vec{v}\,\rangle\,\vec{j}, \tag{9.11}$$

$$\widehat{P}_{\vec{k}}\,\vec{v} := \vec{v}_{\vec{k}} = v_z\,\vec{k} = \langle\,\vec{k}\mid\vec{v}\,\rangle\,\vec{k}. \tag{9.12}$$

We can similarly introduce projectors onto arbitrary unit vectors which is depicted by Figure 6.4.2 and given by Definition 6.4.2 (1).

Definition 9.3.1(1)

(1) *The projector* $\widehat{P}_{\vec{e}}$ *onto a unit vector* $\vec{e}$ *is defined by*

$$\widehat{P}_{\vec{e}}\,\vec{v} := \vec{v}_{\vec{e}} = \langle\vec{e}\mid\vec{v}\,\rangle\,\vec{e} \qquad \forall\vec{v}\in I\!\vec{E}^3. \tag{9.13}$$

(2) *We call* $\widehat{P}_{\vec{e}}$ *the one-dimensional projector onto* $\vec{e}$*. We also call* $\widehat{P}_{\vec{e}}$ *the projector generated by* $\vec{e}$*. The vector* $\vec{v}_{\vec{e}} = \langle\vec{e}\mid\vec{v}\,\rangle\,\vec{e}$ *is called the projection of* $\vec{v}$ *onto* $\vec{e}$*.*

We shall present a list of properties under 8 headings:

P9.3.1(1) *Dirac notation* As with projection matrices shown in Eq. (7.176) projectors are often denoted in *Dirac notation*, e.g., we write

$$\boxed{\widehat{P}_{\vec{i}} = |\,\vec{i}\,\rangle\langle\,\vec{i}\,|, \quad \widehat{P}_{\vec{e}} = |\,\vec{e}\,\rangle\langle\,\vec{e}\,|.} \tag{9.14}$$

Equations (9.10) to (9.13) become

$$\widehat{P}_{\vec{i}}\,\vec{v} = |\,\vec{i}\,\rangle\langle\,\vec{i}\,|\,\vec{v} = \langle\,\vec{i}\mid\vec{v}\,\rangle\,\vec{i}, \tag{9.15}$$

$$\widehat{P}_{\vec{j}}\,\vec{v} = |\,\vec{j}\,\rangle\langle\,\vec{j}\,|\,\vec{v} = \langle\,\vec{j}\mid\vec{v}\,\rangle\,\vec{j}, \tag{9.16}$$

$$\widehat{P}_{\vec{k}}\,\vec{v} = |\,\vec{k}\,\rangle\langle\,\vec{k}\,|\,\vec{v} = \langle\,\vec{k}\mid\vec{v}\,\rangle\,\vec{k}; \tag{9.17}$$

$$\widehat{P}_{\vec{e}}\,\vec{v} = |\,\vec{e}\,\rangle\langle\,\vec{e}\,|\,\vec{v} = \langle\vec{e}\mid\vec{v}\,\rangle\,\vec{e}. \tag{9.18}$$

P9.3.1(2) *Norms of projections and projectors*

(1) The norm of the projection is generally smaller than that of the original vector, since by Schwarz inequality, we have

$$||\vec{v}_{\vec{e}}\,|| = |\langle\vec{e}\mid\vec{v}\,\rangle| \le ||\vec{v}\,||. \tag{9.19}$$

It follows that for all unit vectors $\vec{u}$ we have

$$||\,\widehat{P}_{\vec{e}}\,\vec{u}\,|| \le 1 \;\text{ and }\; ||\,\widehat{P}_{\vec{e}}\,\vec{e}\,|| = 1. \tag{9.20}$$

This means that the norm $||\,\widehat{P}_{\vec{e}}\,||$ of the projector is 1.

[3] *See* Q8(9).

(2) The norm $||\vec{v}_{\vec{e}}||$ of the projection $\vec{v}_{\vec{e}}$ is equal to the square root of the value of the quadratic form generated by the projector for the vector $\vec{v}$, i.e.,

$$||\vec{v}_{\vec{e}}||^2 = \langle \vec{v} \mid \widehat{P}_{\vec{e}}\, \vec{v} \rangle, \tag{9.21}$$

since

$$\langle \vec{v} \mid \widehat{P}_{\vec{e}}\, \vec{v} \rangle = \langle \vec{e} \mid \vec{v} \rangle \langle \vec{v} \mid \vec{e} \rangle = |\langle \vec{e} \mid \vec{v} \rangle|^2. \tag{9.22}$$

For example, we have

$$\langle \vec{v} \mid \widehat{P}_{\vec{i}}\, \vec{v} \rangle = v_x^2, \quad \langle \vec{v} \mid \widehat{P}_{\vec{j}}\, \vec{v} \rangle = v_y^2, \quad \langle \vec{v} \mid \widehat{P}_{\vec{k}}\, \vec{v} \rangle = v_z^2. \tag{9.23}$$

These values are obviously less than or at most equal to the norm of $\vec{v}$.

P9.3.1(3) *Mutually orthogonal projectors*[4] Two orthogonal vectors $\vec{e}_1$ and $\vec{e}_2$ have zero projection onto each other. Consequently the product of their corresponding projectors $\widehat{P}_{\vec{e}_1}$ and $\widehat{P}_{\vec{e}_2}$ is zero. Naturally we call these two projectors orthogonal. For example, $\widehat{P}_{\vec{i}}$, $\widehat{P}_{\vec{j}}$ and $\widehat{P}_{\vec{k}}$ are mutually orthogonal projectors and we have

$$\widehat{P}_{\vec{i}}\widehat{P}_{\vec{j}} = \widehat{P}_{\vec{k}}\widehat{P}_{\vec{i}} = \widehat{P}_{\vec{j}}\widehat{P}_{\vec{k}} = \widehat{0}. \tag{9.24}$$

P9.3.1(4) *Summing up to identity* $\widehat{I\!I}$ By Eq. (9.23) we get

$$\begin{aligned} ||\vec{v}||^2 &= v_x^2 + v_y^2 + v_x^2 = \langle \vec{v} \mid \widehat{P}_{\vec{i}}\, \vec{v} \rangle + \langle \vec{v} \mid \widehat{P}_{\vec{j}}\, \vec{v} \rangle + \langle \vec{v} \mid \widehat{P}_{\vec{k}}\, \vec{v} \rangle \\ &= \langle \vec{v} \mid \left(\widehat{P}_{\vec{i}} + \widehat{P}_{\vec{j}} + \widehat{P}_{\vec{k}} \right) \vec{v} \rangle \quad \forall \vec{v} \in \vec{I\!E}^3. \end{aligned} \tag{9.25}$$

It follows that the sum of projectors $\widehat{P}_{\vec{i}}$, $\widehat{P}_{\vec{j}}$ and $\widehat{P}_{\vec{k}}$ is equal to the identity operator, i.e.,

$$\widehat{P}_{\vec{i}} + \widehat{P}_{\vec{j}} + \widehat{P}_{\vec{k}} = \widehat{I\!I}. \tag{9.26}$$

This can also be proved directly, i.e., we have, $\forall\, \vec{v} \in \vec{I\!E}^3$,

$$\begin{aligned} \left(\widehat{P}_{\vec{i}} + \widehat{P}_{\vec{j}} + \widehat{P}_{\vec{k}} \right) \vec{v} \;&=\; \widehat{P}_{\vec{i}}\, \vec{v} + \widehat{P}_{\vec{j}}\, \vec{v} + \widehat{P}_{\vec{k}}\, \vec{v} \\ &= v_x\, \vec{i} + v_y\, \vec{j} + v_z\, \vec{k} = \vec{v} \\ &\Rightarrow \widehat{P}_{\vec{i}} + \widehat{P}_{\vec{j}} + \widehat{P}_{\vec{k}} = \widehat{I\!I}. \end{aligned} \tag{9.27}$$

[4] Not to be confused with orthogonal operators.

This result applies to projectors $\widehat{P}_{\vec{e}_\ell}$ generated by any complete orthonormal set of vectors $\vec{e}_\ell$, i.e., we have[5]

$$\widehat{P}_{\vec{e}_1} + \widehat{P}_{\vec{e}_2} + \widehat{P}_{\vec{e}_3} = \widehat{I\!I}. \tag{9.28}$$

P9.3.1(5) *Quadratic form* For a unit vector $\vec{u}$ we have

$$\|\vec{u}\,\|^2 = \langle \vec{u} \mid \vec{u}\,\rangle = u_x^2 + u_y^2 + u_z^2 = 1. \tag{9.29}$$

It follows from Eq. (9.23) that the quadratic forms generated by the projectors $\widehat{P}_{\vec{i}}$, $\widehat{P}_{\vec{j}}$ and $\widehat{P}_{\vec{k}}$ on $\vec{I\!E}^3$ satisfy

$$\langle \vec{u} \mid \widehat{P}_{\vec{i}}\,\vec{u}\,\rangle + \langle \vec{u} \mid \widehat{P}_{\vec{j}}\,\vec{u}\,\rangle + \langle \vec{u} \mid \widehat{P}_{\vec{k}}\,\vec{u}\,\rangle = 1. \tag{9.30}$$

Consequently we have

$$0 \leq \langle \vec{u} \mid \widehat{P}_{\vec{i}}\,\vec{u}\,\rangle = u_x^2 \leq 1, \tag{9.31}$$

$$0 \leq \langle \vec{u} \mid \widehat{P}_{\vec{j}}\,\vec{u}\,\rangle = u_y^2 \leq 1, \tag{9.32}$$

$$0 \leq \langle \vec{u} \mid \widehat{P}_{\vec{k}}\,\vec{u}\,\rangle = u_z^2 \leq 1. \tag{9.33}$$

The projectors $\widehat{P}_{\vec{e}_\ell}$ generated by any three orthonormal vectors $\vec{e}_\ell$ also satisfy these inequalities. It should be pointed out that the quadratic form generated by a projector for any vector has a non-negative value, i.e., $\langle \vec{v} \mid \widehat{P}_{\vec{e}}\,\vec{v}\,\rangle \geq 0$, on account of Eq. (9.22).

P9.3.1(6) *Selfadjointness condition and selfadjointness* When calculating scalar product a scalar multiplication operator $\widehat{M}_a = a\widehat{I\!I}$, $a \in I\!R$ can be placed to act on the vector on the right or on the left of the scalar product, i.e.,

$$\langle \vec{u} \mid \widehat{M}_a\,\vec{v}\,\rangle = \langle \widehat{M}_a\,\vec{u} \mid \vec{v}\,\rangle \quad \forall\, \vec{u}, \vec{v} \in \vec{I\!E}^3. \tag{9.34}$$

This is similar to the selfadjointness condition for matrices in Eq. (7.163). Projectors also satisfy this property, i.e., we have

$$\langle \vec{u} \mid \widehat{P}_{\vec{e}}\,\vec{v}\,\rangle = \langle \widehat{P}_{\vec{e}}\,\vec{u} \mid \vec{v}\,\rangle \quad \forall\, \vec{u},\ \vec{v} \in \vec{I\!E}^3. \tag{9.35}$$

We can conclude from Eq. (8.33) that $\widehat{P}_{\vec{e}}$ is equal to its adjoint $\widehat{P}^\dagger_{\vec{e}}$. Following Definition 7.7.4(1) for selfadjoint matrices we call projector $\widehat{P}_{\vec{e}}$ *selfadjoint* and Eq. (9.35) the *selfadjointness condition* for projectors.

[5] Projectors $\widehat{P}_{\vec{e}_\ell}$ are defined by $\widehat{P}_{\vec{e}_\ell}\,\vec{v} := v_\ell\,\vec{e}_\ell,\ v_\ell = \langle \vec{e}_\ell \mid \vec{v}\,\rangle$.

P9.3.1(7) *Idempotence* Like projection matrices projectors are idempotent, i.e.,

$$\widehat{P}_{\vec{e}} = \widehat{P}_{\vec{e}}^{\,2}. \tag{9.36}$$

As with projection matrices the selfadjointness and idempotent properties are so characteristic of projectors that we can even determine whether a given operator is a projector by its selfadjointness and idempotent properties. A formal statement of this will be given in Theorem 9.4.2(1).

P9.3.1(8) *Matrix representations* Projectors are represented by projection matrices. In the basis $\{\vec{i}, \vec{j}, \vec{k}\}$ the matrix representations of $\widehat{P}_{\vec{i}}$, $\widehat{P}_{\vec{j}}$ and $\widehat{P}_{\vec{k}}$ are respectively,

$$\boldsymbol{P}_{\vec{i}} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{P}_{\vec{j}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{9.37}$$

$$\boldsymbol{P}_{\vec{k}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9.38}$$

9.3.2 Projectors onto Subspaces

As shown in Eq. (6.59) an arbitrary vector $\vec{v}$ can be expressed as a unique sum of a vector $\vec{v}_{\vec{\mathcal{S}}}$ in a given subspace $\vec{\mathcal{S}}$ and a vector $\vec{v}_{\vec{\mathcal{S}}^\perp}$ in the orthogonal complement $\vec{\mathcal{S}}^\perp$ of $\vec{\mathcal{S}}$. This enables us to introduce projectors onto subspaces, based on the fact that the vectors $\vec{v}_{\vec{\mathcal{S}}}$ and $\vec{v}_{\vec{\mathcal{S}}^\perp}$ in Eq. (6.59) are the projections of the vector $\vec{v}$ onto the subspaces $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^\perp$, respectively.

Definition 9.3.2(1) *The projector $\widehat{P}_{\vec{\mathcal{S}}}$ onto the subspace $\vec{\mathcal{S}}$ is the operator defined by*

$$\widehat{P}_{\vec{\mathcal{S}}}\,\vec{v} := \vec{v}_{\vec{\mathcal{S}}} \quad \forall\, \vec{v} \in \vec{I\!E}^3. \tag{9.39}$$

The projector $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ onto the subspace $\vec{\mathcal{S}}^\perp$ is the operator defined by

$$\widehat{P}_{\vec{\mathcal{S}}^\perp}\vec{v} := \vec{v}_{\vec{\mathcal{S}}^\perp} \quad \forall\, \vec{v} \in \vec{I\!E}^3. \tag{9.40}$$

This is a generalisation of Definition 9.3.1(1) which applies to one-dimensional subspaces. For example, when the subspace is one-dimensional spanned by the unit vector $\vec{e}$ the projector $\widehat{P}_{\vec{\mathcal{S}}_{\vec{e}}}$ defined

above is the same as the projector $\widehat{P}_{\vec{e}}$ in Eq. (9.13),[6] e.g., we have $\widehat{P}_{\vec{\mathcal{S}}_z} = \widehat{P}_{\vec{k}}$. The properties listed after Definition 9.3.1(1) can be generalised to apply to $\widehat{P}_{\vec{\mathcal{S}}}$.

Theorem 9.3.2(1) *The sum of the projectors $\widehat{P}_{\vec{\mathcal{S}}}$ and $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ onto a subspace $\vec{\mathcal{S}}$ and its orthogonal complement $\vec{\mathcal{S}}^\perp$ is equal to the identity operator, i.e.,*

$$\widehat{P}_{\vec{\mathcal{S}}} + \widehat{P}_{\vec{\mathcal{S}}^\perp} = \widehat{I\!I}. \tag{9.41}$$

A plane in $\vec{I\!E}^3$ is a two-dimensional subspace spanned by two orthonormal vectors $\vec{e}_1$ and $\vec{e}_2$, e.g., the set $\vec{\mathcal{S}}_{12}$ of vectors $\vec{v}_{12}$ in Eq. (6.55) constitutes a plane. The projector $\widehat{P}_{\vec{\mathcal{S}}_{12}}$ onto the plane $\vec{\mathcal{S}}_{12}$ spanned by two orthonormal vectors $\vec{e}_1$ and $\vec{e}_2$ is defined by

$$\widehat{P}_{\vec{\mathcal{S}}_{12}}\,\vec{v} := \langle \vec{e}_1 \mid \vec{v}\,\rangle\,\vec{e}_1 + \langle \vec{e}_2 \mid \vec{v}\,\rangle \vec{e}_2 \quad \forall \vec{v} \in \vec{I\!E}^3. \tag{9.42}$$

This is called a *two-dimensional projector*. Such a projector is a sum of two one-dimensional projectors, i.e.,

$$\widehat{P}_{\vec{\mathcal{S}}_{12}} = \widehat{P}_{\vec{e}_1} + \widehat{P}_{\vec{e}_2} = \widehat{P}_{\vec{\mathcal{S}}_1} + \widehat{P}_{\vec{\mathcal{S}}_2}. \tag{9.43}$$

An example is the projector $\widehat{P}_{\vec{\mathcal{S}}_{xy}}$ onto the x-y plane. We have $\widehat{P}_{\vec{\mathcal{S}}_{xy}} = \widehat{P}_{\vec{\mathcal{S}}_x} + \widehat{P}_{\vec{\mathcal{S}}_y}$. If we extend Eq. (9.42) to three orthonormal vectors we would obtain a *three-dimensional projector*. However, this is trivial as the resulting projector is the identity operator which projects every vector onto itself. The identity operator is the only three-dimensional projector on $\vec{I\!E}^3$.

Definition 9.3.2(2)[7] *Two projectors are orthogonal if the subspaces onto which they project are orthogonal.*

This is a generalisation of the concept first introduced in property P9.3.1(3). Obvious examples are $\widehat{P}_{\vec{e}_1}$ and $\widehat{P}_{\vec{e}_2}$ if the unit vectors $\vec{e}_1$ and $\vec{e}_2$ are orthogonal, e.g., $\widehat{P}_{\vec{i}}$ and $\widehat{P}_{\vec{j}}$. The projectors $\widehat{P}_{\vec{\mathcal{S}}}$ and $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ are also orthogonal, e.g., $\widehat{P}_{\vec{\mathcal{S}}_z}$ and $\widehat{P}_{\vec{\mathcal{S}}_{xy}}$.

Theorem 9.3.2(2) *Two projectors are orthogonal if and only if their product is zero.*[8]

---

[6] The notation $\vec{\mathcal{S}}_{\vec{e}}$ introduced in Eq. (6.52) denotes the subspace spanned by $\vec{e}$.
[7] A general definition of orthogonal subspaces is given by Definition 6.4.4(3).
[8] *See* Theorem 13.2.2(1) and its proof in the solution for Q13(7).

For two one-dimensional projectors $\widehat{P}_{\vec{e}_1}$ and $\widehat{P}_{\vec{e}_2}$ we can see that

$$\langle \vec{e}_1 \mid \vec{e}_2 \rangle = 0 \quad \Leftrightarrow \quad \widehat{P}_{\vec{e}_1} \widehat{P}_{\vec{e}_2} = \widehat{0}. \tag{9.44}$$

Definition 9.3.2(3) *A discrete set of mutually orthogonal projectors $\widehat{P}_{\vec{\mathcal{S}}_\ell}$ is said to be a* **complete orthogonal family of projectors** *on $\vec{I\!E}^3$ if the sum of the projectors is equal to the identity operator, i.e., if*

$$\sum_\ell \widehat{P}_{\vec{\mathcal{S}}_\ell} = \widehat{I\!I}. \tag{9.45}$$

The following examples and remarks illustrate the above definition:

(1) There is a maximum of three mutually orthogonal projectors in a complete orthogonal family of projectors on $\vec{I\!E}^3$. An example is the set $\{\widehat{P}_{\vec{i}}, \widehat{P}_{\vec{j}}, \widehat{P}_{\vec{k}}\}$. Generally the set of projectors $\widehat{P}_{\vec{e}_\ell}$ generated by a complete orthonormal set of vectors $\vec{e}_\ell$ in $\vec{I\!E}^3$ is a complete orthogonal family of projectors on $\vec{I\!E}^3$.

(2) The projectors need not all be one-dimensional, e.g., the two projectors $\widehat{P}_{\vec{\mathcal{S}}_z}$ and $\widehat{P}_{\vec{\mathcal{S}}_{xy}}$ also form a complete orthogonal family of projectors on $\vec{I\!E}^3$.

(3) Equations (9.30) to (9.33) apply to any complete orthogonal family of projectors. This important result is stated in Theorem 9.3.2(1) below.

(4) Projectors $\widehat{P}_{\vec{\mathcal{S}}}$ are selfadjoint and idempotent. This statement follows P9.3.1(6) and (7).

Theorem 9.3.2(3) *Given a unit vector $\vec{u}$ the quadratic forms generated by a complete orthogonal family of projectors $\widehat{P}_{\vec{\mathcal{S}}_\ell}$ on $\vec{I\!E}^3$ satisfy the following properties*[9]*:*

$$0 \leq \langle \vec{u} \mid \widehat{P}_{\vec{\mathcal{S}}_\ell} \vec{u} \rangle \leq 1, \quad \text{and} \quad \sum_\ell \langle \vec{u} \mid \widehat{P}_{\vec{\mathcal{S}}_\ell} \vec{u} \rangle = 1. \tag{9.46}$$

In Chapter 10 we shall discuss how these results can lead to the establishment of a probability theory in terms of unit vectors and a complete orthogonal family of projectors.

[9] *See* Theorem 13.2.2(1) and its proof in the solution for Q13(8).

9.3.3 Eigenvalue Problem

A projector $\widehat{P}$ admits only two eigenvalues 0 and 1 since it is selfadjoint and idempotent. To prove these results consider the eigenvalue equation $\widehat{P}\,\vec{u} = \lambda\,\vec{u}$, where $\vec{u}$ is normalised. First the eigenvalue is real since

$$\lambda = \langle \vec{u} \mid \widehat{P}\vec{u} \rangle = \langle \vec{u} \mid \widehat{P}\,\widehat{P}\vec{u} \rangle = \langle \widehat{P}\vec{u} \mid \widehat{P}\vec{u} \rangle = ||\widehat{P}\vec{u}\,||^2. \tag{9.47}$$

Next we have

$$\widehat{P}\,\vec{u} = \widehat{P}^2\,\vec{u} = \widehat{P}(\widehat{P}\,\vec{u}) = \lambda\,\widehat{P}\vec{u} = \lambda^2\vec{u} \tag{9.48}$$

$$\Rightarrow \quad \lambda\,\vec{u} = \lambda^2\,\vec{u} \Rightarrow \quad \lambda = 0,\, 1. \tag{9.49}$$

In the case of $\widehat{P}_{\vec{e}}$ the eigenvalue 1 is nondegenerate corresponding to the eigenvector $\vec{e}$, i.e., $\widehat{P}_{\vec{e}}\,\vec{e} = \vec{e}$. The eigenvalue 0 is degenerate. The corresponding eigenvectors are all the vectors $\vec{v}$ orthogonal to $\vec{e}$ since $\widehat{P}_{\vec{e}}\,\vec{v} = 0$ if $\langle \vec{e} \mid \vec{v} \rangle = 0$. We can also appreciate this in terms of matrices. For example, the projector $\widehat{P}_{\vec{i}}$ has a matrix representation $\boldsymbol{P}_{\vec{i}}$ given in Eq. (9.38). One can see that the eigenvectors for the matrix are the column vectors in Eq. (7.93) corresponding to the basis vectors $\vec{i}$, $\vec{j}$, $\vec{k}$.

9.4 Selfadjoint Operators

9.4.1 Concept and Definition

The operators introduced so far have a clear geometric meaning. We can use these operators as building blocks to establish other operators which may not have distinctive geometric properties. The building blocks we have in mind here are projectors. Let $\{\widehat{P}_{\vec{e}_\ell}\}$ be a complete orthogonal family of projectors associated with an orthonormal basis $\{\vec{e}_\ell\}$ for $\vec{I\!E}^3$. In the same way we can generate new vectors in terms of linear combinations of the basis vectors $\vec{e}_\ell$, we can generate new operators by *linear combinations* of a complete orthogonal family of projectors $\widehat{P}_{\vec{e}_\ell}$, i.e., we can construct

$$\widehat{A} := a_1\,\widehat{P}_{\vec{e}_1} + a_2\,\widehat{P}_{\vec{e}_2} + a_3\,\widehat{P}_{\vec{e}_3}, \quad a_1,\ a_2,\ a_3 \in I\!R. \tag{9.50}$$

We can also have linear combinations of non-orthogonal projectors. For these operators we have $\widehat{A} = \widehat{A}^\dagger$ on account of Eq. (8.38), i.e., these operators satisfy the selfadjointness condition discussed in P9.3.1(6). But these operators are not idempotent. In other words, they are not projectors. We have therefore generated a new class of operators characterised by their selfadjointness property. Following Definition 7.7.4(1) for selfadjoint matrices we can use this selfadjointness property to define this new class of operators, rather than employing the construction in Eq. (9.50).[10]

Definition 9.4.1(1) *An operator $\widehat{A}$ on $\vec{I\!E}^3$ is said to be selfadjoint if it is equal to its adjoint, i.e., $\widehat{A} = \widehat{A}^\dagger$.*

This is equivalent to the requirement that $\widehat{A}$ satisfies the following **selfadjointness condition**:

$$\langle \vec{u} \mid \widehat{A}\vec{v} \rangle = \langle \widehat{A}\vec{u} \mid \vec{v} \rangle \quad \forall\ \vec{u},\ \vec{v} \in \vec{I\!E}^3. \tag{9.51}$$

The following comments serve to clarify the definition:

C9.4.1(1) The identity operator $\widehat{I\!I}$ and scalar multiplication operators $\widehat{M}_a$ in Eq. (9.1) are selfadjoint.[11]

C9.4.1(2) Projectors are selfadjoint.

C9.4.1(3) A linear combination of projectors with real coefficients is selfadjoint, e.g., $\widehat{A}$ in Eq. (9.50).

C9.4.1(4) While a linear combination of projectors generates a selfadjoint operator it is not obvious that the converse is true. The spectral theorem, i.e., Theorem 7.7.6(1), tells us that a selfadjoint matrix can be decomposed as a linear combination of projection matrices. We would expect a similar statement is also true for selfadjoint operators. In §9.4.5 we shall present a theorem to this effect, i.e.,

> *all selfadjoint operators on $\vec{I\!E}^3$ can be constructed as a linear combination of a complete orthogonal family of projectors.*

[10] The significance of linear combinations of orthogonal projectors will become clear in §9.4.5 on the spectral theorem.

[11] For real vector spaces like $\vec{I\!E}^3$ scalar multiplications are by real numbers. For complex vector spaces scalar multiplications can be by complex numbers and complex functions which will not give rise to selfadjoint operators.

C9.4.1(5) Selfadjoint operators are represented by selfadjoint matrices.

C9.4.1(6) Selfadjointness is a very distinctive property, not satisfied by many other operators. Take the orthogonal operator $\widehat{R}_z(\theta_z)$ in Eq. (9.2) for example. For vectors $\vec{u}$ and $\vec{v}$ lying in the x-y plane, i.e.,

$$\vec{u} = u_x\vec{i} + u_y\vec{j}, \quad \vec{v} = v_x\vec{i} + v_y\vec{j},$$

we have, in accordance with Eqs. (7.146) and (7.148),

$$\widehat{R}_z(\theta_z)\vec{v} = (v_x\cos\theta_z - v_y\sin\theta_z)\vec{i} + (v_x\sin\theta_z + v_y\cos\theta_z)\vec{j}. \quad (9.52)$$

It is easily verified that generally

$$\langle\vec{u} \mid \widehat{R}_z(\theta_z)\vec{v}\,\rangle \neq \langle\widehat{R}_z(\theta_z)\vec{u} \mid \vec{v}\,\rangle. \quad (9.53)$$

It follows that $\widehat{R}_z(\theta_z)$ is not selfadjoint. The corresponding matrix $\boldsymbol{R}_z(\theta_z)$ in Eq. (7.149) is also not selfadjoint. Generally if we want to have an operator to act on $\vec{u}$ on the left in the scalar product $\langle\vec{u} \mid \widehat{A}\vec{v}\,\rangle$ we have to employ the adjoint $\widehat{A}^\dagger$ to do it. When acting on $\vec{u} = u_x\vec{i} + u_y\vec{j}$ the required adjoint operator $\widehat{R}_z^\dagger(\theta_z)$ is given by[12]

$$\widehat{R}_z^\dagger(\theta_z)\vec{u} = u_x'\vec{i} + u_y'\vec{j}, \quad (9.54)$$

where

$$u_x' = u_x\cos\theta_z + u_y\sin\theta_z, \quad u_y' = -u_x\sin\theta_z + u_y\cos\theta_z. \quad (9.55)$$

We have

$$\langle\vec{u} \mid \widehat{R}_z(\theta_z)\vec{v}\,\rangle = \langle\widehat{R}_z^\dagger(\theta_z)\vec{u} \mid \vec{v}\,\rangle. \quad (9.56)$$

The operator $\widehat{R}_z(\theta_z)$ is not equal to its adjoint $\widehat{R}_z^\dagger(\theta_z)$.

The following theorems are intuitive and useful.[13]

[12] This corresponds to the adjoint of the rotation matrix in Eq. (7.149). For a real matrix its adjoint is equal to its transpose.

[13] These are special cases of Theorem 13.1(1) and Corollary 13.1(1) which apply to complex vector spaces.

Theorem 9.4.1(1)[14] *If $\widehat{A}$ is selfadjoint then*

$$\langle \vec{u} \mid \widehat{A}\vec{u} \rangle = 0 \ \ \forall \vec{u} \quad \Rightarrow \quad \widehat{A} = \widehat{0}. \tag{9.57}$$

Theorem 9.4.1(2)[15] *If $\widehat{A}$ and $\widehat{B}$ are selfadjoint then*

$$\langle \vec{u} \mid \widehat{A}\vec{u} \rangle = \langle \vec{u} \mid \widehat{B}\vec{u} \rangle \ \ \forall \vec{u} \quad \Rightarrow \quad \widehat{A} = \widehat{B}. \tag{9.58}$$

Corollary 9.4.1(1) For a selfadjoint operator $\widehat{A}$ we have

$$\langle \vec{u} \mid \widehat{A}\vec{u} \rangle = \langle \vec{u} \mid \vec{u} \rangle \ \ \forall \vec{u} \quad \Rightarrow \quad \widehat{A} = \widehat{I\!I}. \tag{9.59}$$

Selfadjoint matrices and selfadjoint operators have many distinctive properties which are crucial to the probabilistic formulation of quantum mechanics. These properties are studied in detail in the remaining subsections of this chapter.

9.4.2 Properties and Relations with Projectors

The following properties are easily verified:

P9.4.2(1) Scalar multiplication does not affect the selfadjointness of an operator, i.e., if $\widehat{A}$ is selfadjoint, then $a\widehat{A}$ is selfadjoint, i.e.,

$$\widehat{A} = \widehat{A}^\dagger \quad \Rightarrow \quad (a\widehat{A})^\dagger = a\widehat{A}, \quad a \in I\!R. \tag{9.60}$$

P9.4.2(2) Linear combinations of selfadjoint operators are again selfadjoint because of Eq. (8.38), i.e., $\forall\ a_\ell \in I\!R$ we have

$$\widehat{A}_\ell = \widehat{A}_\ell^\dagger \quad \Rightarrow \quad \Big(\sum_\ell a_\ell \widehat{A}_\ell\Big) = \Big(\sum_\ell a_\ell \widehat{A}_\ell\Big)^\dagger. \tag{9.61}$$

It follows that the sum of two selfadjoint operators is selfadjoint.

P9.4.2(3) The product of two selfadjoint operators is not necessarily selfadjoint because of Eq. (8.41). However, the product of two commuting selfadjoint operators is selfadjoint, i.e.,

$$(\widehat{A}\widehat{B})^\dagger = \widehat{B}^\dagger \widehat{A}^\dagger = \widehat{B}\widehat{A} = \widehat{A}\widehat{B}. \tag{9.62}$$

[14] Halmos p. 138. *See* Theorem 13.1(1) and Corollary 13.1(1) for complex vector spaces.
[15] Halmos p. 138.

In particular the square of a selfadjoint operator is selfadjoint:

$$\widehat{A} = \widehat{A}^\dagger \;\Rightarrow\; (\widehat{A})^2 = (\widehat{A}^2)^\dagger \tag{9.63}$$

$$\Rightarrow\; \langle \vec{u} \mid \widehat{A}^2 \vec{v} \rangle = \langle \widehat{A}^2 \vec{u} \mid \vec{v} \rangle. \tag{9.64}$$

P9.4.2(4) The products of an operator with its adjoint are selfadjoint, i.e., $\widehat{A}^\dagger \widehat{A}$ and $\widehat{A}\widehat{A}^\dagger$ are selfadjoint.[16]

P9.4.2(5) A selfadjoint operator is a projector if it is idempotent. This is a very useful result which we shall highlight in the form of a theorem for easy reference later.

Theorem 9.4.2(1)[17] *An operator on $\vec{I\!E}^3$ is a projector if and only if it is selfadjoint and idempotent.*

Corollary 9.4.2(1)

(1) *The sum of two projectors is a projector if and only if the two projectors are orthogonal.*

(2) *The product of two projectors is a projector if and only if the two projectors commute.*

9.4.3 Eigenvalue Problem and Quadratic Form

Orthogonal operators and projectors on $\vec{I\!E}^3$ have only two eigenvalues, i.e., ± 1 for orthogonal operators and 0, 1 for projectors. Selfadjoint operators can have arbitrary eigenvalues. Having a wide range of eigenvalues enables us to use selfadjoint operators to describe a wide range of physical quantities. The selfadjoint operator $\widehat{A}$ in Eq. (9.50) possesses values a_1, a_2, a_3. These are the eigenvalues of $\widehat{A}$ corresponding to eigenvectors $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ since they satisfy the following eigenvalue equations:

$$\widehat{A}\,\vec{e}_1 = a_1\vec{e}_1, \quad \widehat{A}\,\vec{e}_2 = a_2\vec{e}_2, \quad \widehat{A}\,\vec{e}_3 = a_3\vec{e}_3. \tag{9.65}$$

The operator operates like a multiplication operator when acting on its eigenvectors. Depending on the choice of a_1, a_2 and a_3 the operator has at least one eigenvalue and at most three distinct eigenvalues. When an eigenvalue is degenerate, say $a_1 = a_2$, then

[16] This can be proved using Eqs. (8.40) and (8.41).

[17] Fano p. 77 for a proof. This theorem can be used to prove an operator to be a projector, e.g., in establishing Corollary 9.4.2(1).

any linear combination $c_1\vec{e}_1 + c_2\vec{e}_2$ of the eigenvectors $\vec{e}_1$, $\vec{e}_2$ is again an eigenvector corresponding to the eigenvalue a_1.

Since selfadjoint operators are represented by selfadjoint matrices we can solve their eigenvalue problem in terms of their representative selfadjoint matrices to obtain similar results, e.g., the contents of Theorems 7.7.4(1) and 7.7.4(2) for selfadjoint matrices also carry over to selfadjoint operators.

There is another way to generate values for a selfadjoint operator $\widehat{A}$. Let $\vec{u}$ be a unit vector. Then the quadratic form $\mathcal{Q}(\widehat{A}, \vec{u}) = \langle \vec{u} \mid \widehat{A}\vec{u} \rangle$ introduced in §8.2.2 produces a value for each $\vec{u}$. Eigenvalues are special cases of the quadratic form, i.e.,

$$\mathcal{Q}(\widehat{A}, \vec{e}_1) = a_1, \quad \mathcal{Q}(\widehat{A}, \vec{e}_2) = a_2, \quad \mathcal{Q}(\widehat{A}, \vec{e}_3) = a_3. \tag{9.66}$$

9.4.4 Eigensubspaces and Eigenprojectors

Definition 9.4.4(1)

(1) *A subspace, denoted by $\vec{\mathcal{S}}^{\hat{A}}(a)$, is called the eigensubspace of a selfadjoint operator $\widehat{A}$ corresponding to the eigenvalue a if it is spanned by the eigenvectors of $\widehat{A}$ corresponding to the eigenvalue a.*

(2) *The projector $\widehat{P}^{\hat{A}}(a)$ onto the eigensubspace $\vec{\mathcal{S}}^{\hat{A}}(a)$ is called the eigenprojector of the operator $\widehat{A}$ corresponding to or associated with the eigenvalue a.*

Let a_ℓ be the eigenvalues of a selfadjoint operator $\widehat{A}$. Then:

(1) The eigensubspace $\vec{\mathcal{S}}^{\hat{A}}(a_\ell)$ and its associated eigenprojector $\widehat{P}^{\hat{A}}(a_\ell)$ are one-dimensional if the corresponding eigenvalue a_ℓ is nondegenerate. The eigenprojector is equal to the projector generated by the corresponding normalised eigenvector $\vec{e}_\ell$, i.e.,

$$\widehat{P}^{\hat{A}}(a_\ell) = \widehat{P}_{\vec{e}_\ell} = |\vec{e}_\ell\rangle\langle\vec{e}_\ell|. \tag{9.67}$$

(2) If a_2 is degenerate with degeneracy 2 corresponding to two orthonormal eigenvectors $\vec{e}_{21}$, $\vec{e}_{22}$, then corresponding eigensubspace is two-dimensional and the eigenprojector $\widehat{P}^{\hat{A}}(a_2)$ is equal to the sum of two projectors in accordance with Eq. (9.43),

i.e.,[18]

$$\widehat{P}^{\hat{A}}(a_2) = \widehat{P}_{\vec{e}_{21}} + \widehat{P}_{\vec{e}_{22}} = |\vec{e}_{21}\rangle\langle\vec{e}_{21}| + |\vec{e}_{22}\rangle\langle\vec{e}_{22}|. \tag{9.68}$$

We also call $|\vec{e}_{21}\rangle\langle\vec{e}_{21}|$ and $|\vec{e}_{22}\rangle\langle\vec{e}_{22}|$ eigenprojectors. Let $\vec{e}'_{21}$ and $\vec{e}'_{22}$ be two orthonormal vectors formed by linear combinations of $\vec{e}_{21}$ and $\vec{e}_{22}$. Then these new vectors are also eigenvectors of $\widehat{A}$ corresponding to the eigenvalue a_2. They will generate two new eigenprojectors, i.e.,

$$\widehat{P}_{\vec{e}'_{21}} = |\vec{e}'_{21}\rangle\langle\vec{e}'_{21}| \quad \text{and} \quad \widehat{P}_{\vec{e}'_{22}} = |\vec{e}'_{22}\rangle\langle\vec{e}'_{22}|. \tag{9.69}$$

Their sum is the same as that of the original projectors, i.e.,

$$\widehat{P}^{\hat{A}}(a_2) = \widehat{P}_{\vec{e}_{21}} + \widehat{P}_{\vec{e}_{22}} = \widehat{P}_{\vec{e}'_{21}} + \widehat{P}_{\vec{e}'_{22}}. \tag{9.70}$$

The eigenvalues and eigenvectors of selfadjoint operators possess the same properties as selfadjoint matrices in Theorem 7.7.4(1).

Theorem 9.4.4(1)

(1) *Eigenvalues of a selfadjoint operator are real.*

(2) *Eigenvectors of a selfadjoint operator corresponding to different eigenvalues are orthogonal.*

(3) *The eigenvectors of a selfadjoint operator can be chosen to form an orthonormal basis in* $\vec{I\!E}^3$.

(4) *A selfadjoint operator has associated with it a complete orthogonal family of eigenprojectors generated by a complete orthonormal set of eigenvectors of the operator.*

To prove this theorem we first note that a selfadjoint operator is representable by a selfadjoint matrix. The operator and its matrix representation possess the same eigenvalues in accordance with Eq. (8.50). The proof of this theorem then follows from Theorems 7.7.4(2) and 7.7.4(4).

Selfadjoint operators with non-negative eigenvalues can be singled out to form a distinctive group with many useful properties.

[18] There is no need to have the superscript $\hat{A}$ on the projector $\widehat{P}_{\vec{e}_{21}}$ which is determined by the vector $\vec{e}_{21}$.

Definition 9.4.4(2) *A selfadjoint operator $\widehat{A}$ is said to be positive if its eigenvalues are not negative.*[19]

A positive selfadjoint operator can have zero eigenvalue. Projectors are examples of positive operators.

9.4.5 Spectral Theorem of Selfadjoint Operators on $\vec{I\!E}^3$

Projectors are closely related to selfadjoint operators as seen in Eq. (9.50) and Theorem 9.4.2(1). The relation goes deeper. For matrices we have Theorem 7.7.6(1) which tells us that a sefladjoint matrix can be written in terms of its eigenvalues and their associated projection matrices. Equation (9.50) suggests that this statement should also apply to selfadjoint operators on $\vec{I\!E}^3$. We know that every selfadjoint operator $\widehat{A}$ on $\vec{I\!E}^3$ possesses a complete orthonormal set of eigenvectors $\vec{e}_\ell$ which would generate a complete orthogonal family of eigenprojectors $\widehat{P}_{\vec{e}_\ell}$. We can construct an operator by forming a linear combination of the eigenprojectors $\widehat{P}_{\vec{e}_\ell}$ with the eigenvalues a_ℓ as coefficients as in Eq. (9.50). Let us summarise this result into a theorem.

Theorem 9.4.5(1) *A selfadjoint operator $\widehat{A}$ on $\vec{I\!E}^3$ is expressible as a linear combination of a complete orthogonal family of eigenprojectors $\widehat{P}_{\vec{e}_\ell}$ generated by a complete orthonormal set of eigenvectors $\vec{e}_\ell$ with the corresponding eigenvalues a_ℓ as coefficients, i.e.,*

$$\boxed{\widehat{A} = \sum_{\ell=1}^{3} a_\ell \, \widehat{P}_{\vec{e}_\ell} = \sum_{\ell=1}^{3} a_\ell \, |\vec{e}_\ell\rangle\langle\vec{e}_\ell|.} \tag{9.71}$$

Theorem 9.4.5(1) is known as the **spectral theorem** for selfadjoint operators on $\vec{I\!E}^3$.[20] The above expression for $\widehat{A}$ is referred to as a **spectral decomposition** of $\widehat{A}$.

Proof We know that $\widehat{A}$ possesses a complete orthonormal set of eigenvectors $\vec{e}_\ell$ and a corresponding complete orthogonal family of

[19] *See* Definition 13.3.1(1) on positive operators on a complex vector space.

[20] If any of the eigenvalues is degenerate then the choice of $\vec{e}_\ell$, and hence $\widehat{P}_{\vec{e}_\ell}$, is not unique. *See* the discussion after the proof of this theorem for a detailed discussion. *See* also Theorem 9.4.5(2).

eigenprojectors $\widehat{P}_{\vec{e}_\ell}$ corresponding to a set of eigenvalues a_ℓ. For any vector $\vec{v} \in \vec{I\!E}^3$ we have $\vec{v} = \sum_{\ell=1}^3 v_\ell \vec{e}_\ell$, $v_\ell = \langle \vec{e}_\ell \mid \vec{v} \rangle$, and

$$\widehat{A}\vec{v} = \sum_{\ell=1}^{3} v_\ell \widehat{A}\vec{e}_\ell = \sum_{\ell=1}^{3} a_\ell v_\ell \vec{e}_\ell. \tag{9.72}$$

Construct the operator

$$\widehat{A}' := \sum_{\ell=1}^{3} a_\ell \, \widehat{P}_{\vec{e}_\ell}. \tag{9.73}$$

Then for any vector $\vec{v} \in \vec{I\!E}^3$ we have

$$\widehat{A}'\vec{v} = \sum_{\ell=1}^{3} a_\ell \, \widehat{P}_{\vec{e}_\ell}\vec{v} = \sum_{\ell=1}^{3} a_\ell \, \langle \vec{e}_\ell \mid \vec{v} \, \rangle \vec{e}_\ell = \sum_{\ell=1}^{3} a_\ell v_\ell \vec{e}_\ell. \tag{9.74}$$

It follows that $\widehat{A}' = \widehat{A}$. **QED**

The following comments serve to clarify the theorem:

C9.4.5(1) The fact that the sum of the projectors in a complete orthogonal family is the identity operator can be rephrased as a *spectral decomposition of the identity operator* as a selfadjoint operator, i.e., we have

$$\boxed{\widehat{I\!I} = \sum_{\ell=1}^{3} \widehat{P}_{\vec{e}_\ell}.} \tag{9.75}$$

This expression is called a *spectral decomposition* of the identity.[21]

C9.4.5(2) The eigenvalues a_ℓ in Eq. (9.71) may not be all different. If an eigenvalue is degenerate, say $a_1 \neq a_2 = a_3$, then the spectral decomposition in Eq. (9.71) becomes

$$\widehat{A} = a_1 \widehat{P}_{\vec{e}_1} + a_2 \widehat{P}_{\vec{e}_2} + a_2 \widehat{P}_{\vec{e}_3} = a_1 \widehat{P}^{\hat{A}}(a_1) + a_2 \widehat{P}^{\hat{A}}(a_2), \tag{9.76}$$

where

$$\widehat{P}^{\hat{A}}(a_1) = \widehat{P}_{\vec{e}_1}, \quad \widehat{P}^{\hat{A}}(a_2) = \widehat{P}_{\vec{e}_2} + \widehat{P}_{\vec{e}_3}. \tag{9.77}$$

[21] The decomposition of the identity is not unique. This is because $\widehat{I\!I}$ has only one eigenvalue, i.e., 1, which is degenerate with degeneracy 3. The sum of any complete orthogonal family of projectors will add up to the identity.

Here $\widehat{P}^{\hat{A}}(a_2)$ is a two-dimensional projector. As discussed in relation to Eq. (9.70) it is possible to choose two different orthonormal eigenvectors for the eigenvalue a_2, but Eqs. (9.76) and (9.77) remain the same.[22] The spectral decomposition of $\widehat{A}$ can then be rewritten as

$$\widehat{A} = a_1 \widehat{P}^{\hat{A}}(a_1) + a_2 \widehat{P}^{\hat{A}}(a_2), \tag{9.78}$$

where $\widehat{P}^{\hat{A}}(a_1)$ and $\widehat{P}^{\hat{A}}(a_2)$ correspond to distinct eigenvalues of $\widehat{A}$.

For clarity we shall restate the spectral theorem in terms of eigenprojectors associated with distinct eigenvalues.[23]

Theorem 9.4.5(2) Spectral theorem *A selfadjoint operator $\widehat{A}$ is expressible as a linear combination of its complete orthogonal family of eigenprojectors $\widehat{P}^{\hat{A}}(a_m)$ corresponding distinct eigenvalues a_m, i.e.,*[24]

$$\boxed{\widehat{A} = \sum_m a_m \, \widehat{P}^{\hat{A}}(a_m).} \tag{9.79}$$

The above expression for $\widehat{A}$ is referred to as the *spectral decomposition* of $\widehat{A}$. The corresponding spectral decomposition of the identity takes the form

$$\boxed{\widehat{I\!I} = \sum_m \widehat{P}^{\hat{A}}(a_m).} \tag{9.80}$$

The spectral theorem tells that while a given selfadjoint operator determines its eigenvalues and eigenvectors through its eigenvalue equation the converse is also true, i.e., knowledge of the eigenvalues and eigenvectors determine a unique selfadjoint operator by

[22] It is convenient to re-label things, e.g., re-label both a_2 and a_3 as a_2, and $\vec{e}_2$, $\vec{e}_3$ as $\vec{e}_{21}$ and $\vec{e}_{22}$, and projectors $\widehat{P}_{\vec{e}_2}$ and $\widehat{P}_{\vec{e}_3}$ as $\widehat{P}_{\vec{e}_{21}}$ and $\widehat{P}_{\vec{e}_{22}}$.

[23] The eigenvalues a_ℓ in Eq. (9.71) may not be all different, while the eigenvalues a_m in Eq. (9.79) are all different. The analysis of eigenvalues and eigenvectors of selfadjoint operators is a part of a general study, known as **spectral theory**, of eigenvalues, eigenvectors and their generalisations and applications to a wide range of operators in a variety of spaces.

[24] The index m goes from 1 to 3 if all the eigenvalues are nondegenerate. Theorem 9.4.5(2) takes the form of Theorem 9.4.5(1). The index m runs from 1 to 2 if only one of the eigenvalues is degenerate. In the case where all the eigenvalues are the same, e.g., for $\widehat{M}_a$ in Eq. (9.1), the summation reduces to a single term. *See* also Isham p. 50 for the theorem in Dirac notation.

Eq. (9.71) or Eq. (9.79). As we shall see in §13.3 and §20.3 this theorem remains valid in higher dimensional complex vector spaces. The spectral theorem is the foundation for a mathematical formulation of quantum mechanics.

9.4.6 Functions of Selfadjoint Operators

A real-valued function $f : \mathbb{R} \to \mathbb{R}$ of a real variable $\tau \in \mathbb{R}$ is commonly denoted by $f(\tau)$. Let $\widehat{A}$ be a selfadjoint operator with nondegenerate eigenvalues a_ℓ corresponding to normalised eigenvectors $\vec{e}_\ell$ and eigenprojectors $\widehat{P}_{\vec{e}_\ell}$. The question posed here is how one would define a corresponding function of a selfadjoint operator $\widehat{A}$. For a polynomial function, e.g.,

$$f(\tau) = c_0 + c_1\tau + c_2\tau^2, \quad c_0, c_1, c_2 \in \mathbb{R} \tag{9.81}$$

we can define an operator $f(\widehat{A})$ by

$$f(\widehat{A}) := c_0 + c_1\widehat{A} + c_2\widehat{A}^2, \tag{9.82}$$

and call this operator a function of $\widehat{A}$ corresponding to the polynomial $f(x)$. The eigenvalues and eigenvectors of $f(\widehat{A})$ are simply $f(a_\ell)$ and $\vec{e}_\ell$, i.e.,

$$f(\widehat{A})\vec{e}_\ell = f(a_\ell)\vec{e}_\ell, \tag{9.83}$$

This is such an intuitive and desirable property that we want to have a general definition of an arbitrary function of $\widehat{A}$ to possess this property.

Definition 9.4.6(1) *Let $\widehat{A}$ be a selfadjoint operator with a spectral decomposition shown in* Eq. (9.79)*. For an arbitrary real-valued function $f(\tau)$ the operator $f(\widehat{A})$ given by*

$$f(\widehat{A}) := \sum_m f(a_m)\,\widehat{P}^{\widehat{A}}(a_m) \tag{9.84}$$

is defined to be the function $f(\widehat{A})$ of $\widehat{A}$.

This definition is consistent with Eq. (9.82) and satisfies Eq. (9.83). For real-valued functions $f(\tau)$ the resulting operators $f(\widehat{A})$ are all

selfadjoint. Three examples are

E9.4.6(1) *Square* The square of a selfadjoint operator $\widehat{A}$ is defined by

$$\widehat{A}^2 := \sum_m a_m^2 \, \widehat{P}^{\hat{A}}(a_m). \tag{9.85}$$

This agrees with the expression $\widehat{A}^2 = \widehat{A}\widehat{A}$. The resulting operator is positive.

E9.4.6(2) *The inverse* The inverse of an invertible selfadjoint operator $\widehat{A}$ is defined by

$$\widehat{A}^{-1} := \sum_m a_m^{-1} \, \widehat{P}^{\hat{A}}(a_m). \tag{9.86}$$

This agrees with the expression $\widehat{A}^{-1}\widehat{A} = \widehat{I\!I}$. The definition is possible only if $a_m \neq 0$ for all m, i.e., if $\widehat{A}$ is invertible.

E9.4.6(3) *Square root* For a positive selfadjoint operator its square root is defined by

$$\sqrt{\widehat{A}} := \sum_m \sqrt{a_m} \, \widehat{P}^{\hat{A}}(a_m). \tag{9.87}$$

This agrees with the intuition that the square of $\sqrt{\widehat{A}}$ is $\widehat{A}$.

Exercises and Problems

Q9(1) Show that an operator on $\vec{I\!E}^3$ is orthogonal if and only if it is invertible and its inverse is equal to its adjoint.

Q9(2) Show that orthogonal operators on $\vec{I\!E}^3$ have at most two eigenvalues, i.e., ± 1. Find the eigenvalues and eigenvectors of $\widehat{R}_{rx}$ in Eq. (9.4) and $\widehat{R}_{rxyz}$ in Eq. (9.5).

Q9(3) Let $\{\vec{e}_\ell\}$ be an orthonormal basis for $\vec{I\!E}^3$ show that their orthogonal transforms $\{\vec{e}\,'_\ell\}$ generated by an orthogonal operator $\widehat{R}$ is also an orthonormal basis for $\vec{I\!E}^3$.

Q9(4) Prove Eqs. (9.6) to (9.8) on orthogonal transformations.

Q9(5) Prove Theorem 9.3.2(1).

Q9(6) Find the eigenvectors of the two-dimensional projector in Eq. (9.43) corresponding to the eigenvalues 1 and 0.

Q9(7) Verify Eq. (9.44).

Q9(8) Prove Theorems 9.4.1(1), 9.4.1(2) and Corollary 9.4.1(1).

Q9(9) Show that all projectors satisfy the sefladjointness condition in Eq. (9.35) and that they are also idempotent.

Q9(10) Show that selfadjoint operators are represented by selfadjoint matrices.

Q9(11) Demonstrate the inequality in Eq. (9.53) with $\vec{u} = u_y \vec{j}$ and $\vec{v} = v_x \vec{i}$. Verify Eq. (9.56).

Q9(12) Prove Eq. (9.70).

Q9(13) Show that the selfadjoint operator $\widehat{A}$ in Eq. (9.71) admits a_ℓ as its eigenvalues with the unit vectors $\vec{e}_\ell$ as the corresponding eigenvectors.

Q9(14) Show that the products of an operator with its adjoint, i.e., $\widehat{A}^\dagger \widehat{A}$ and $\widehat{A}\widehat{A}^\dagger$, are selfadjoint and positive.

Q9(15) Show that the expression of the inverse in Eq. (9.86) satisfy the condition $\widehat{A}\widehat{A}^{-1} = \widehat{I\!I}$.

Q9(16) The square root $\sqrt{\widehat{A}}$ of a positive operator $\widehat{A}$ is defined by Eq. (9.87). Show that $\left(\sqrt{\widehat{A}}\right)^2 = \widehat{A}$.

Q9(17) Show that the quadratic form $\mathcal{Q}(\widehat{A}, \vec{u})$ generated by a positive selfadjoint operator in any $\vec{u}$ is non-negative.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 10

Probability, Selfadjoint Operators, Unit Vectors and the Need for Complexness

In our earlier discussion of probability theory there is no mention of any mechanism to generate probabilities. There are of course many different ways of generating probabilities. In many classical statistical experiments we can assume all outcomes are equally likely to occur. This would enable us to equate the probability $\wp(E_1)$ of an event E_1 to the ratio of the number n_1 of outcomes in the event to the total number N of outcomes in the sample space, i.e., $\wp(E_1) = n_1/N$. The counting of various combinations and permutations of outcomes becomes important. Consider the tossing of a fair dice mentioned in §3.3. The sample space given by Eq. (3.23) has six outcomes. The probability of the event E_e consisting of even numbers has three outcomes. Since all the outcomes have an equal probability of occurrence the probability of the event E_e is simply equal to $3/6 = 1/2$. However, this method cannot be universally applied since not all outcomes are equally likely to occur in a general statistical experiment. Some well-known examples are cited in §4.3. We then have to introduce different ways to produce the required probability distribution. In the next section we shall show

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

how probability distributions can be generated using unit vectors and selfadjoint operators on $\vec{I\!E}^3$.

10.1 Generating Probability Distributions on $\vec{I\!E}^3$

Consider an experiment to measure an observable A of a physical system which can yield one of three different values a_1, a_2 and a_3 with different probabilities. The measurement of A then amounts to a statistical experiment with $S_{am} = \{a_1, a_2, a_3\}$ as its sample space. Our aim is to obtain a probability mass function $\wp$, a function from the sample space S_{am} into the interval [0, 1], to give a quantitative characterisation of the experiment. Generally the probability of each of measured outcome would depend on the state of the system as well as that particular observable in question. A simple counting of number of outcomes without reference to the state cannot possibly lead to the correct probability distribution.

Suppose the physical system possesses property QMP5.3(3) given in §5.3, i.e., the state ϕ^s of the system can determine the probability distribution of the values of A in the form of a probability mass function $\wp^A(\phi^s, a_\ell)$ on the sample space $\mathcal{S}_{am}$.[1] A change of state will give rise to a different probability mass function. Our task is to find a mathematical description of the state ϕ^s and the observable A which will lead to a probability mass function on the sample space S_{am}. Vectors and operators discussed in the preceding sections can provide such a description in the following way[2]:

(1) Starting with the given values a_1, a_2 and a_3 of the observable we can construct a selfadjoint operator $\widehat{A}$ on $\vec{I\!E}^3$ using a_1, a_2 and a_3 as its eigenvalues and an appropriate complete orthonormal set of vectors $\vec{e}_\ell$ as the corresponding eigenvectors. To proceed we first define the complete orthogonal family of projectors $\widehat{P}_{\vec{e}_\ell}$

[1] For a given observable the values of the probability density function will depend on both the state and the values of A, hence the notation $\wp^A(\phi^s, a_\ell)$.

[2] There are many descriptions which can lead to a probability mass function. Here we are just trying a description in terms of vectors and operators. There is no guarantee that the resulting probability mass function will agree with the actual probability mass function $\wp^A(\phi^s, a_\ell)$ for the statistical experiment.

generated by $\vec{e}_\ell$. Then the operator can be explicitly defined in terms of a_1, a_2, a_3 and $\widehat{P}_{\vec{e}_\ell}$ by

$$\widehat{A} := \sum_{\ell=1}^{3} a_\ell \, \widehat{P}_{\vec{e}_\ell}, \qquad \widehat{P}_{\vec{e}_\ell} = |\vec{e}_\ell\rangle\langle\vec{e}_\ell|. \tag{10.1}$$

This construction amounts to defining the selfadjoint operator by its spectral decomposition as in Theorem 9.4.5(1).

(2) Next we observe that the complete orthogonal family of projectors $\widehat{P}_{\vec{e}_\ell}$ can generate a function on the sample space $S_{am} = \{a_1, a_2, a_3\}$ for each given unit vector $\vec{u}$ by the quadratic form $\mathcal{Q}(\widehat{P}_{\vec{e}_\ell}, \vec{u}\,)$, i.e.,[3]

$$\wp^{\hat{A}}(\vec{u}, a_\ell) := \mathcal{Q}(\widehat{P}_{\vec{e}_\ell}, \vec{u}\,) = \langle \vec{u} \mid \widehat{P}_{\vec{e}_\ell}\vec{u}\,\rangle. \tag{10.2}$$

More explicitly we have

$$\wp^{\hat{A}}(\vec{u}, a_\ell) = |\langle \vec{e}_\ell \mid \vec{u}\,\rangle|^2. \tag{10.3}$$

This function satisfies the defining properties of probability mass functions specified in Eq. (3.24), i.e.,[4]

$$\boxed{0 \leq \wp^{\hat{A}}(\vec{u}, a_\ell) \leq 1, \quad \textstyle\sum_{\ell=1}^{3} \wp^{\hat{A}}(\vec{u}, a_\ell) = 1,} \tag{10.4}$$

on account of Eq. (9.46).

(3) The above mathematical results strongly suggest the following description of the system:

(a) Describe observable A by the operator $\widehat{A}$ in Eq. (10.1).

(b) Describe the state ϕ^s by a unit vector $\vec{u}$ in $\vec{I\!E}^3$.

(c) We can then identify the function $\wp^{\hat{A}}(\vec{u}, a_\ell)$ in Eq. (10.2) as the probability mass function for the distribution of the values of A in state described by unit vector $\vec{u}$, i.e., we have

$$A \quad \rightarrow \quad \widehat{A}, \tag{10.5}$$

$$\phi^s \quad \rightarrow \quad \vec{u}, \tag{10.6}$$

$$\wp^{A}(\phi^s, a_\ell) \quad = \quad \wp^{\hat{A}}(\vec{u}, a_\ell). \tag{10.7}$$

---

[3] Note that $\wp^{A}$ and $\wp^{\hat{A}}$ are meant to be different functions with $\wp^{\hat{A}}$ having $\vec{u}$ and a_ℓ as its arguments and $\wp^{A}$ having ϕ^s and a_ℓ as its arguments.

[4] *See* Eq. (13.29) and its proof on the solution to Q13(7).

(d) The expectation value $\mathcal{E}(\wp^A)$ defined by Eq. (3.32) can be calculated directly in terms of the vector $\vec{u}$ and the operator $\widehat{A}$, i.e., using Eq. (10.2) and the spectral theorem 9.4.5 (1) we get

$$\begin{aligned}\mathcal{E}(\wp^A) &:= \sum_{\ell=1}^{3} a_\ell \, \wp^A(\phi^s, a_\ell) = \sum_{\ell=1}^{3} a_\ell \, \langle \vec{u} \mid \widehat{P}_{\vec{e}_\ell} \vec{u} \rangle \\ &= \langle \vec{u} \mid \Big(\sum_{\ell=1}^{3} a_\ell \widehat{P}_{\vec{e}_\ell} \Big) \vec{u} \rangle \\ &= \langle \vec{u} \mid \widehat{A}\vec{u} \rangle. \end{aligned} \tag{10.8}$$

(e) It is more transparent to denote the expectation value by $\mathcal{E}(A, \phi^s)$ or more directly by $\mathcal{E}(\widehat{A}, \vec{u})$.

(f) Equation (10.8) shows that the expectation value $\mathcal{E}(A, \phi^s)$ is equal to the quadratic form $\mathcal{Q}(\widehat{A}, \vec{u})$ generated by $\widehat{A}$ for the vector $\vec{u}$, i.e., we have

$$\mathcal{E}(A, \phi^s) = \mathcal{E}(\widehat{A}, \vec{u}) = \mathcal{Q}(\widehat{A}, \vec{u}). \tag{10.9}$$

This enables us to express the uncertainty in accordance with Theorem 3.5(1) as

$$\Delta(\widehat{A}, \vec{u}) = \sqrt{\mathcal{Q}(\widehat{A}^2, \vec{u}) - \mathcal{Q}(\widehat{A}, \vec{u})^2}, \tag{10.10}$$

where $\mathcal{Q}(\widehat{A}^2, \vec{u}) := \langle \vec{u} \mid \widehat{A}^2 \vec{u} \rangle$.

(4) The description presented above is applicable to cases when not all the values a_1, a_2 and a_3 are different. Suppose $a_1 \neq a_2 = a_3$. We re-label the two distinct eigenvalues as $a_{m=1} = a_1$ and $a_{m=2} = a_2 = a_3$. In accordance with the spectral theorem 9.4.5 (2) we have

$$\widehat{A} = \sum_{m=1}^{2} a_m \, \widehat{P}^{\hat{A}}(a_m), \tag{10.11}$$

where $\widehat{P}^{\hat{A}}(a_1) = \widehat{P}_{\vec{e}_1}$ and $\widehat{P}^{\hat{A}}(a_2) = \widehat{P}_{\vec{e}_2} + \widehat{P}_{\vec{e}_3}$. We get

$$\wp^{\hat{A}}(\vec{u}, a_1) := \langle \vec{u} \mid \widehat{P}^{\hat{A}}(a_1)\vec{u} \rangle = \langle \vec{u} \mid \widehat{P}_{\vec{e}_1} \vec{u} \rangle, \tag{10.12}$$

$$\wp^{\hat{A}}(\vec{u}, a_2) := \langle \vec{u} \mid \widehat{P}^{\hat{A}}(a_2)\vec{u} \rangle = \langle \vec{u} \mid \widehat{P}_{\vec{e}_2} \vec{u} \rangle + \langle \vec{u} \mid \widehat{P}_{\vec{e}_3} \vec{u} \rangle. \tag{10.13}$$

(5) Finally we would compare $\wp^{\hat{A}}(\vec{u}, a_\ell)$ with the experimental probability mass function $\wp^{A}(\phi^s, a_\ell)$.

A system described above will also satisfy Properties QMP5.3(1) and QMP5.3(2) in §5.3.

(1) The probabilistic behaviour of the system produced by the above description automatically implies that not all observables are compatible. Let $\widehat{A}'$ be another selfadjoint operator with eigenvalues a'_ℓ corresponding eigenvectors $\vec{e}\,'_\ell$. Generally there may not exist a unit vector $\vec{u}$ such that

$$\wp^{\hat{A}}(\vec{u}, a_\ell) = \langle \vec{u} \mid \widehat{P}_{\vec{e}_\ell} \vec{u} \rangle = |\langle \vec{e}_\ell \mid \vec{u} \rangle|^2 = 1. \tag{10.14}$$

$$\wp^{\hat{A}'}(\vec{u}, a'_\ell) = \langle \vec{u} \mid \widehat{P}_{\vec{e}\,'_\ell} \vec{u} \rangle = |\langle \vec{e}\,'_\ell \mid \vec{u} \rangle|^2 = 1. \tag{10.15}$$

If $\widehat{A}$ and $\widehat{A}'$ have the same eigenvectors, e.g., $\vec{e}_\ell = \vec{e}\,'_\ell$, then

(a) We have

$$\widehat{A} = \sum_{\ell=1}^{3} a_\ell \, \widehat{P}_{\vec{e}_\ell} \quad \text{and} \quad \widehat{A}' = \sum_{\ell=1}^{3} a'_\ell \, \widehat{P}_{\vec{e}_\ell}. \tag{10.16}$$

(b) The two operators $\widehat{A}$ and $\widehat{A}'$ commute.[5]

(c) For probability distributions we have, setting $\vec{u} = \vec{e}_\ell$,

$$\wp^{\hat{A}}(\vec{e}_\ell, a_\ell) = \wp^{\hat{A}'}(\vec{e}_\ell, a'_\ell) = \langle \vec{e}_\ell \mid \widehat{P}_{\vec{e}_\ell} \vec{e}_\ell \rangle = 1. \tag{10.17}$$

This means that in the state described by vector $\vec{e}_\ell$ observable A possesses the value a_ℓ and observable A' possesses the value a'_ℓ. The two observables are compatible.

(2) A state φ^s_ℓ exists in which A possesses the value a_ℓ. Such a state can be described by the unit vector $\vec{e}_\ell$ since we have

$$\wp^{\hat{A}}(\vec{e}_\ell, a_\ell) = \langle \vec{e}_\ell \mid \widehat{P}_{\vec{e}_\ell} \vec{e}_\ell \rangle = 1. \tag{10.18}$$

[5] The link between compatibility and commutativity has important consequences in quantum mechanics.

10.2 A Model Probability Theory Based on $\vec{I\!E}^3$

10.2.1 The Model

Consider a system possessing quantum characteristics QMP5.3(1) to QMP5.3(4), i.e., the system possesses the following properties:

P10.2.1(1) Not all observables are simultaneously measurable.

P10.2.1(2) There are states compatible with any given observable, e.g., there is a state $\varphi^s_{a_1}$ in which a given observable A possesses the value a_1.

P10.2.1(3) In an arbitrary state ϕ^s a measurement of an observable A would result in the value a_ℓ with probability $\wp^A(\phi^s, a_\ell)$ which is determined by the state ϕ^s and observable A.

P10.2.1(4) Superposition principle is satisfied.

Moreover, suppose observables of the system can yield a maximum of three different values on measurement, i.e., observable A can yield a maximum of three different values a_1, a_2, a_3. Our task is to establish a possible description of such a system.

Following the discussion in the preceding section we can formulate a model theory for the system based on $\vec{I\!E}^3$ as follows:

(1) Choose the state space to be $\vec{I\!E}^3$ with states described by unit vectors in $\vec{I\!E}^3$. These unit vectors are referred to as **state vectors**.

(2) Describe observable A by a selfadjoint operator $\widehat{A}$ in Eq. (10.1) with a suitable choice of a complete orthonormal set $\{\vec{e}_\ell\}$ and the values a_ℓ of A as the eigenvalues of $\widehat{A}$.

(3) In an arbitrary state described by a unit vector $\vec{u}$ the probability mass function for the values of A is given by

$$\wp^{\hat{A}}(\vec{u}, a_\ell) := \mathcal{Q}(\widehat{P}_{\vec{e}_\ell}, \vec{u}) = \langle \vec{u} \mid \widehat{P}_{\vec{e}_\ell} \vec{u} \rangle = |\langle \vec{e}_\ell \mid \vec{u} \rangle|^2. \quad (10.19)$$

The expectation value and uncertainty are given by Eqs. (10.9) and (10.10).

(4) As discussed in the preceding subsection the probabilistic behaviour of the system produced by the above description automatically implies QMP 5.3(1) to QMP 5.3(2), i.e.,

   (a) not all observables are compatible, and

   (b) in the state described by vector $\vec{e}_\ell$ observable A possesses the value a_ℓ.

(5) The superposition principle can be formulated in terms of normalised linear combinations of state vectors. Let $\vec{u}_1$ and $\vec{u}_2$ be two unit vectors we can generate new unit vectors $\vec{u}$ by

$$\vec{u} = c_1\vec{u}_1 + c_2\vec{u}_2 \quad \text{where} \quad |c_1|^2 + |c_2|^2 = 1. \tag{10.20}$$

   This new vector can serve as a new state vector which contains information not available in $\vec{u}_1$ and $\vec{u}_2$ taken separately, e.g.,

$$\begin{aligned}\mathcal{E}(\widehat{A}, \vec{u}\,) = \langle \vec{u} \mid \widehat{A}\vec{u}\,\rangle = \mathcal{E}(\widehat{A}, \vec{u}_1\,) + \mathcal{E}(\widehat{A}, \vec{u}_2\,) \\ + c_1c_2\,\langle \vec{u}_1 \mid \widehat{A}\vec{u}_2\rangle + c_2c_1\,\langle \vec{u}_2 \mid \widehat{A}\vec{u}_1\rangle.\end{aligned} \tag{10.21}$$

   The last two terms which depend on how the two states are combined together, i.e., on c_1 and c_2, are known as the **correlation** or **interference terms**.[6]

The theory presented can be extended to lower as well as higher dimensional real vector spaces. For example, if a system with observables capable of only having two different values we would take the state space to be the two-dimensional space $\vec{I\!E}^2$.

10.2.2 A Need for Generalisation

The fact that vectors and selfadjoint operators on a real vector space like $\vec{I\!E}^3$ can provide a mechanism for generating probability distributions satisfying properties QMP5.3(1) to QMP5.3(4) does not guarantee that they can give a correct description of quantum systems. For a model theory to have any chance of success it

[6] *See* §31.5 for more discussion.

is necessary for the model to satisfy the following additional requirements:

(1) There must be enough unit vectors and selfadjoint operators to correspond to all the states and all the observables of the system.

(2) The probability distribution given in Eq. (10.19) for every observable in every state must agree with experimental results.

The choice of real vector spaces turns out to be too restrictive for the description of quantum systems. Let us illustrate the difficulty with the example of electron spin. An electron spin can align itself along the x, y and z directions and the spin along any direction is capable of taking only two values, e.g., S_z possesses only two values $\pm\hbar/2$. In accordance with QMP5.3(2) there is a state α_z^s in which S_z possesses the value $\hbar/2$ and a state β_z^s in which S_z possesses the value $-\hbar/2$. Having only two values an electron spin should be described in a two-dimensional vector space, not a three-dimensional vector space like $I\!\vec{E}^3$. It is tempting to employ a two-dimensional real vector space $I\!\vec{E}^2$ like the y-z plane as the state space to describe an electron spin. For example, we may attempt to describe the *spin-up* state α_z^s by the unit vector $\vec{k}$ in the positive z direction and the *spin-down* state β_z^s by the unit vector $-\vec{k}$ in the negative z-direction. But this is incompatible with the model theory in the preceding subsection. In accordance with the model theory the two states must be described two eigenvectors of the operator representing S_z corresponding to eigenvalues $\hbar/2$ and $-\hbar/2$. This means that the two vectors must be orthogonal to each other. But $\vec{k}$ and $-\vec{k}$ are not orthogonal to each other. It follows that we cannot described α_z^s by $\vec{k}$ and β_z^s by $-\vec{k}$. Intuitively we can also see that vectors in the y-z plane cannot describe the spin along the x direction.[7]

A way out would be to set up a theory on two-dimensional *complex vector spaces*.

[7] A more detailed analysis on why a two-dimensional real vector space cannot accommodate the three observables of spin is presented in Chapter 14.

10.3 Need for Complexness

Many quadratic equations permit no solutions in terms of real numbers, the equation $x^2 + 1 = 0$ being an example. Complex numbers were introduced in order to provide some formal solutions. A symbolic quantity $\sqrt{-1}$, known as the *imaginary unit* and denoted by i, was introduced. It was defined by the property that its square was equal to 1, i.e., $i^2 = -1$. Then the equation $x^2 + 1 = 0$ admits two formal solutions, i.e., $x_\pm = \pm i$. Complex numbers were defined to be quantities of the form $a + bi$, where a, b were real numbers. These quantities were regarded as fictitious and unrelated to our physical world. Over the years complex numbers and complex-valued functions do eventually find many applications in engineering and classical physics. In classical physics a wave is a propagation of disturbances in a medium. The simplest kind of waves are harmonic waves travelling along the x-axis described by sinusoidal or cosinusoidal functions of x and time t, e.g., a sinusoidal wave in a vibrating string discussed in §2.2. Harmonic waves propagating along the positive and negative x directions are describable by the following functions:

$$Y_+(x, t) = A\cos(kx - \omega t), \quad Y_-(x, t) = A\cos(kx + \omega t). \quad (10.22)$$

The functions Y_+ and Y_- may be visualised as the displacement of the medium along the y axis. They are solutions of Eq. (2.10), the classical wave equation in one-dimension. We can represent such wave motion in terms of complex-valued functions of x and t, i.e.,

$$F_+(x, t) = Ae^{i(kx-\omega t)}, \quad F_-(x, t) = Ae^{-i(kx+\omega t)}. \quad (10.23)$$

These complex functions are often referred to as **plane waves**, to contrast the real harmonic waves. They are used to simplify certain mathematical manipulation. At the end of the calculations we have to take the real or imaginary parts of the complex functions to recover all the physical quantities of interest. For example, if we wish to superpose the two opposing waves $Y_+(x, t)$ and $Y_-(x, t)$ we can proceed in two different ways:

(1) We add the two real functions in Eq. (10.22):

$$Y(x, t) = Y_+(x, t) + Y_-(x, t) = 2A\cos kx\cos\omega t. \quad (10.24)$$

(2) We add the two complex functions in Eq. (10.23):

$$F(x,t) = F_+(x,t) + F_-(x,t) = 2A\cos kx\, e^{-i\omega t}. \tag{10.25}$$

We have to take the real part of $F(x,t)$ in order to recover $Y(x,t)$ in Eq. (10.24), using *Euler's formula*[8]

$$e^{i\theta} = \cos\theta + i\sin\theta, \quad \theta \in I\!R. \tag{10.26}$$

While complex numbers and functions are widely used they are not necessary for classical physics where all the fundamental equations such as wave equations in Eq. (2.10) and the equation for Newton's second law are real.

The situation is entirely different in quantum physics. The time-dependent Schrödinger equation which is the *quantum wave equation* is complex. It explicitly contains the imaginary unit i.[9] For the free motion of a quantum particle of mass m in one-dimension along the x axis the time-dependent Schrödinger equation is

$$i\hbar\frac{\partial\phi(x,t)}{\partial t} = \widehat{H}\,\phi(x,t), \quad \widehat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}. \tag{10.27}$$

Solutions of the time-dependent Schrödinger equation are generally known as *quantum wave functions* or simply *wave functions*. Intuitively we may imagine these quantum wave functions describing a kind of *quantum waves*. In contrast to solutions of the classical wave equation Eq. (2.10) the solutions to the quantum wave equation are necessarily complex, i.e., a quantum wave has to be described by a complex wave function.[10]

An important common property of waves, satisfied by both classical and quantum waves, is that they can be superposed to produce a new wave.[11] By examining the superposition of waves we can gain an intuitive understanding as to why quantum waves have to correspond to complex functions:

[8] Euler (1707–1783) is a Swiss mathematician and physicist.

[9] Schrödinger (1887–1961) is an Austrian physicist famous for his equation and a quantum paradox by his name. He shared the 1933 Nobel Prize in Physics with Dirac.

[10] The term *wave functions* is also used to refer to square-integrable functions as stated in §18.4.2.1.

[11] Mathematically this is due to the fact that both the classical and quantum wave equations are linear. For a linear equation a linear combination of two solutions is also a solution.

(1) The superposition of two opposing classical waves in Eq. (10.22) produces what is known as a *standing wave* described by $Y(x, t)$ in Eq. (10.24). A standard wave corresponds to the medium vibrating up and down without propagation. The vibration ceases periodically over the entire medium, e.g., at $t_1 = \pi/2\omega$, $t_2 = 3\pi/2\omega$, we have $Y(x, t_1) = 0$ and $Y(x, t_2) = 0$ for all x.

(2) The above result implies that the real function $Y(x, t)$ cannot describe a quantum wave. A quantum wave is known to represent the position probability amplitude of the particle, i.e., the absolute value square is the position probability density function of the particle. Such a density function cannot be zero everywhere at any time since that would mean a zero probability of finding the particle anywhere. A description of position probability amplitude by $Y(x, t)$ would mean the periodic disappearance of the particle. This is not acceptable. On the other hand the complex function $F(x, t)$ in Eq. (10.25) can never vanish for all x at any given time, a property necessary for the description of position probability amplitude. This would also imply that the corresponding quantum wave equation, i.e., the time-dependent Schrödinger equation, must be complex as well.[12]

We have now seen two examples, the electron spin in the preceding section and the interpretation of the wave function discussed above, to show that complex numbers and complex functions are intrinsically related to the quantum world.[13] Quantum systems have to be described by complex-valued functions, *complex vectors* and operators on complex vectors. Before looking into complex vectors we shall familiarise ourselves with complex numbers first.

Complex numbers introduced earlier in an intuitive manner in terms of the symbolic imaginary unit $i = \sqrt{-1}$ possesses the usual algebraic properties of real numbers in the following sense:

P10.3(1) The imaginary unit can be multiplied by itself, i.e., we can square the imaginary unit, i.e., $i^2 = i \times i$ and this square is defined to be the real number -1.

[12] O'hanian p. 157.
[13] Hence complex numbers are not fictitious and devoid of reality.

P10.3(2) The imaginary unit can be multiplied by a real numbers b to produce a formal quantity of the form bi. This quantity is known as an *imaginary number*. The imaginary unit multiplied by the real number zero is assumed to be zero.

P10.3(3) An imaginary number bi can be added to a real number a to produce a formal quantity z of the form $z = a+bi$ which is called a *complex number*. Real and imaginary numbers are then special cases of complex numbers. We also have the following terminology:

(1) The real number a and b are referred to respectively, as the *real* and *imaginary parts* of z.

(2) The *complex conjuage* of z, denoted by z^*, is defined to be the complex number $z^* := a - b\,i$. It follows that $z^{**} = z$.

(3) The *absolute value* of z, denoted by $|z|$, is defined to be the real number $|z| := \sqrt{a^2 + b^2}$.

P10.3(4) Two complex numbers are deemed equal if and only if their real and imaginary parts are equal.

P10.3(5) Complex numbers can be manipulated according to the following rules[14]:

$$(a_1 + b_1\,i) + (a_2 + b_2\,i) := (a_1 + a_2) + (b_1 + b_2)\,i, \tag{10.28}$$

$$(a_1 + b_1\,i)(a_2 + b_2\,i) := (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)\,i. \tag{10.29}$$

P10.3(6) With above rules for addition and multiplication the absolute value of z can be calculated in terms of the product of z and its complex conjugate, i.e.,

$$|z| = \sqrt{z^*z}. \tag{10.30}$$

We have, for any two complex numbers z and w

$$|z^*| = |z| \quad \text{and} \quad |zw| = |z|\,|w|. \tag{10.31}$$

There is a need to put things on a more rigorous mathematical basis in oder to make the concept of complex vectors easier to understand. In the next chapter we shall present a definition of complex numbers which can be generalised to define complex vectors.

[14] It is possible to define division of complex numbers as well.

Exercises and Problems

Q10(1) In the model theory in §10.2.1 the observable A has three possible values, i.e., a_1, a_2, and a_3. Find the probabilities for the value a_1 when the state vector $\vec{u}$ is one of the following unit vectors:

$$\vec{u}_1 = \vec{e}_1, \quad \vec{u}_2 = \vec{e}_2, \quad \text{and} \quad \vec{u}_3 = \frac{1}{\sqrt{3}}\,\vec{e}_1 + \sqrt{\frac{2}{3}}\,\vec{e}_3. \tag{10.32}$$

What are the expectation values $\mathcal{E}(\widehat{A}, \vec{u}\,)$ and the uncertainties $\Delta(\widehat{A}, \vec{u}\,)$ in these cases?

Q10(2) Show that $\widehat{A}$ and $\widehat{A}'$ in Eq. (10.16) commute.

Q10(3) Prove Eq. (10.31).

Q10(4) Let θ, a and b be real numbers. For the complex numbers $z = e^{i\theta}$ and $z = e^{i\theta}\big(a + ib\big)$ show that

$$|e^{i\theta}| = 1 \quad \text{and} \quad |z|^2 = a^2 + b^2. \tag{10.33}$$

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 11

Complex Vectors

11.1 Complex Numbers

It is not obvious how one would define complex vectors. It is not meaningful to introduce *imaginary vectors* by just taking the "square root" of a "negative vector", similar to the way we take the square root of the negative number -1, to arrive at an "imaginary number". Also the multiplication of a real vector $\vec{u} \in \vec{I\!E}^3$ by a complex number, e.g., $(a + ib)\vec{u}$, does not have an intuitive meaning. We need an approach which is not based on "imaginary vectors".

Let us illustrate such an approach with complex numbers. The idea is to define complex numbers as *order pairs* (a, b) of real numbers a and b subject to the following rules of algebraic operations:

Equality $\quad (a_1, b_1) = (a_2, b_2)$ if and only if $a_1 = a_2$, $b_1 = b_2$.

Addition $\quad (a_1, b_1) + (a_2, b_2) := (a_1 + a_2, b_1 + b_2)$.

Multiplication $\quad (a_1, b_1)(a_2, b_2) := (a_1a_2 - b_1b_2, a_1b_2 + b_1a_2)$.

This is equivalent to defining complex numbers as members of the Cartesian product $I\!R \times I\!R$ formed by the set of all ordered pairs (a, b) of real numbers a and b subject to above rules of equality, addition and multiplication.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

To compare with the intuitive definition in Eqs. (10.28) and (10.29) given earlier we can introduce the following terms:[1]

(1) We call a and b, respectively, the *real* and the *imaginary* parts of the complex number $z := (a, b)$.

(2) We call the pair $(0, 1)$ the *imaginary unit*. The square of the imaginary unit can be calculated according the multiplication rule, i.e., we have

$$(0, 1)(0, 1) = (-1, 0). \tag{11.1}$$

If we denote the imaginary unit $(0, 1)$ by i then Eq. (11.1) becomes the usual expression $i^2 = -1$.

(3) We identify a pair of the form $(a, 0)$ with the real number a and we call a pair of the form $(0, b)$ an imaginary number. The following are illustrative examples:

(a) We can identify $(-1, 0)$ with the real number -1 and $(0, 1)$ with the imaginary number i.

(b) We can denote the multiplication of $(a', 0)$ and (a, b) simply by $a'(a, b)$. By the multiplication rule we have

$$a'(a, b) = (a', 0)(a, b) = (a'a, a'b). \tag{11.2}$$

(c) The order in the pair is important since the definition of equality stated above implies $(a, b) \neq (b, a)$ if $a \neq b$.

(4) Real and imaginary numbers are a subset of the complex numbers.

(5) The complex conjugate z^* of z is defined to be $z^* := (a, -b)$.

(6) The absolute value or *norm* of $|z|$ of z is defined to be $|z| := \sqrt{a^2 + b^2}$.

We can summarise the results by making the following identification with our previous definition of complex numbers:

$$(0, 1) = i, \quad (a, 0) = a, \quad (a, b) = a + ib. \tag{11.3}$$

In practical calculations it is easier to employ our previous notation $z = a + ib$, rather than (a, b) for complex numbers.

[1] Here a, b, a' are real numbers.

11.2 Complexification of the Vector Space $\vec{I\!E}^3$

11.2.1 Complex Vectors

Consider the Cartesian product $\vec{I\!E}^3 \times \vec{I\!E}^3$ formed by pairs $(\vec{u}, \vec{v})$ of real vectors $\vec{u}, \vec{v}$ in $\vec{I\!E}^3$, i.e.,

$$\vec{I\!E}^3 \times \vec{I\!E}^3 := \Big\{ (\vec{u}, \vec{v}) : \ \vec{u}, \vec{v} \in \vec{I\!E}^3 \Big\}. \tag{11.4}$$

Define algebraic operations of these pairs as follows:

Equality $(\vec{u}_1, \vec{v}_1) = (\vec{u}_2, \vec{v}_2)$ if and only if $\vec{u}_1 = \vec{u}_2, \ \ \vec{v}_1 = \vec{v}_2$.

Addition $(\vec{u}_1, \vec{v}_1) + (\vec{u}_2, \vec{v}_2) := (\vec{u}_1 + \vec{u}_2, \vec{v}_1 + \vec{v}_2)$.

Multiplication[2] $(a, b)(\vec{u}, \vec{v}) := (a\vec{u} - b\vec{v}, a\vec{v} + b\vec{u})$.

Let $\vec{I\!E}^{3c}$ denote the Cartesian product $\vec{I\!E}^3 \times \vec{I\!E}^3$ endowed with the above algebraic operations. We call $\vec{I\!E}^{3c}$ a **complex vector space** with its elements referred to as **complex vectors**. We also call $\vec{I\!E}^{3c}$ the *complexification* of $\vec{I\!E}^3$. Generally a complex vector cannot be visualised as "an arrow" with a direction in the physical space. In practical calculations it is easier to see what is going on if we express complex vectors formally in terms of the imaginary number i, i.e., we can make the following formal identification:

$$(\vec{u}, \vec{v}) = \vec{u} + i\vec{v}, \tag{11.5}$$

$$\begin{aligned} (\vec{u}_1, \vec{v}_1) + (\vec{u}_2, \vec{v}_2) &= (\vec{u}_1 + i\vec{v}_1) + (\vec{u}_2 + i\vec{v}_2) \\ &= (\vec{u}_1 + \vec{u}_2) + i(\vec{v}_1 + \vec{v}_2), \end{aligned} \tag{11.6}$$

$$\begin{aligned} (a, b)(\vec{u}, \vec{v}) &= (a + ib)(\vec{u} + i\vec{v}) \\ &= (a\vec{u} - b\vec{v}) + i(a\vec{v} + b\vec{u}). \end{aligned} \tag{11.7}$$

We often denote complex vectors by Greek letters to distinguish them from real vectors, e.g., we write

$$\vec{\zeta} := (\vec{u}, \vec{v}) = \vec{u} + i\vec{v} \quad \text{and} \quad \vec{\zeta}_\ell := (\vec{u}_\ell, \vec{v}_\ell) = \vec{u}_\ell + i\vec{v}_\ell. \tag{11.8}$$

Each vector $\vec{u}$ in $\vec{I\!E}^3$ has a counterpart $\vec{u}^{\,c}$ in $\vec{I\!E}^{3c}$, i.e.,

$$\vec{u} \in \vec{I\!E}^3 \quad \leftrightarrow \quad \vec{u}^{\,c} := (\vec{u}, \vec{0}) = \vec{u} + i\,\vec{0} \ \in \vec{I\!E}^{3c}. \tag{11.9}$$

[2] Here multiplication is of a vector by scalars, i.e., complex numbers (a, b), not the multiplication of two vectors. Such multiplication follows from the rule of multiplication of complex numbers in §11.1.

We call $\vec{u}^{\,c}$ a **real vector** in $\vec{I\!E}^{3c}$. Real vectors form a subset of $\vec{I\!E}^{3c}$. Equation (11.7) tells us that a real vector becomes a complex vector when multiplied by a complex number, i.e.,

$$(a+ib)\vec{u}^{\,c} = (a+ib)(\vec{u},\,0) = a\vec{u} + ib\vec{u} \in \vec{I\!E}^{3c}. \tag{11.10}$$

The algebraic operations of complex vectors defined earlier possess the following properties:

Addition[3]

CA11.2.1(1)	*Commutative*	$\vec{\zeta}_1 + \vec{\zeta}_2 = \vec{\zeta}_2 + \vec{\zeta}_1$.
CA11.2.1(2)	*Associative*	$(\vec{\zeta}_1 + \vec{\zeta}_2) + \vec{\zeta}_3 = \vec{\zeta}_1 + (\vec{\zeta}_2 + \vec{\zeta}_3)$.
CA11.2.1(3)	*Zero vector*	There is a unique vector in $\vec{I\!E}^{3c}$, the zero vector $\vec{0}^c$, such that $\vec{\zeta} + \vec{0}^c = \vec{\zeta} \quad \forall \vec{\zeta} \in \vec{I\!E}^{3c}$.
CA11.2.1(4)	*Inverse*	For each $\vec{\zeta} \in \vec{I\!E}^{3c}$ there exists a unique inverse denoted by $\vec{\zeta}^{\,-1}$ such that $\vec{\zeta} + \vec{\zeta}^{\,-1} = \vec{0}^c$.

The zero vector is given in terms of the zero real vector $\vec{0}$ by $\vec{0}^c = (\vec{0},\,\vec{0}\,)$.[4]

Scalar multiplication by complex numbers[5]

CSM11.2.1(1)	*Distributive*	$c(\vec{\zeta}_1 + \vec{\zeta}_2) = c\,\vec{\zeta}_1 + c\,\vec{\zeta}_2$ and $(c_1 + c_2)\vec{\zeta} = c_1\,\vec{\zeta} + c_2\,\vec{\zeta}$.
CSM11.2.1(2)	*Associative*	$(c_1c_2)\vec{\zeta} = c_1(c_2\,\vec{\zeta}\,)$.
CSM11.2.1(3)	*Multiplication by* 1	$1\,\vec{\zeta} = \vec{\zeta}$.

With scalar multiplication extended to complex numbers the definitions of linear dependence and independence, dimensions and bases discussed in §6.2.1 and §6.2.2 carry over to $\vec{I\!E}^{3c}$.

Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be a basis for $\vec{I\!E}^3$, then its counterpart in $\vec{I\!E}^{3c}$, denoted by $\{\vec{e}_1^{\,c}, \vec{e}_2^{\,c}, \vec{e}_3^{\,c}\}$, will form a basis in $\vec{I\!E}^{3c}$, since all vectors in $\vec{I\!E}^{3c}$ are expressible as a linear combination of $\vec{e}_\ell^{\,c} := (\vec{e}_\ell,\,\vec{0}\,)$. To

[3] These properties are comparable with the corresponding properties A6.1.1(1) to A6.1.1(4) for vectors in $\vec{I\!E}^3$. Equations (6.1) to (6.3) apply equally here.

[4] In contrast the zero complex number is defined in terms of the number zero by (0, 0).

[5] These properties are comparable with the corresponding properties SM6.1.2(1) to SM6.1.2(3) for vectors in $\vec{I\!E}^3$. From now on a scalar means a complex number in general, e.g., c, c_1 and c_2 are complex numbers.

appreciate this we shall start with a pair of vectors in $\vec{u}, \vec{v} \in \vec{I\!E}^3$ and effect their expansion in terms of $\vec{e}_\ell$, i.e.,

$$\vec{u} = \sum_{\ell=1}^{3} u_\ell \vec{e}_\ell \quad \text{and} \quad \vec{v} = \sum_{\ell=1}^{3} v_\ell \vec{e}_\ell. \tag{11.11}$$

where $u_\ell, v_\ell \in I\!R$. Then the complex vector $\vec{\zeta} = \vec{u} + i\vec{v}$ will have the following expansion[6]:

$$\begin{aligned} \vec{\zeta} = \vec{u} + i\vec{v} &= \left(\sum_{\ell=1}^{3} u_\ell \vec{e}_\ell \right) + i \left(\sum_{\ell=1}^{3} v_\ell \vec{e}_\ell \right) \\ &= \sum_{\ell=1}^{3} (u_\ell + i v_\ell)\, \vec{e}_\ell = \sum_{\ell=1}^{3} (u_\ell + i v_\ell)\, (\vec{e}_\ell, \vec{0}\,) \\ &= \sum_{\ell=1}^{3} (u_\ell + i v_\ell)\, \vec{e}_\ell^{\,c}. \end{aligned} \tag{11.12}$$

We can conclude that any complex vector $\vec{\zeta} \in \vec{I\!E}^{3c}$ is expressible as a linear combination of $\vec{e}_\ell^{\,c}$. The complex vector space $\vec{I\!E}^{3c}$ is therefore three-dimensional. When there is no risk of confusion we can drop the superscript to denote $(\vec{u}, \vec{0}\,)$ by $\vec{u}$, e.g., we can rewrite

$$(a + ib)\vec{u}^{\,c} \text{ as } (a + ib)\vec{u} \text{ and } \vec{0}^{\,c} = (\vec{0}, \vec{0}\,) \text{ as } \vec{0}. \tag{11.13}$$

11.2.2 Scalar Product and Orthonormal Bases

We can define the scalar product $\langle \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c$ and norm $||\vec{\zeta}\,||^c$ in $\vec{I\!E}^{3c}$ in terms of the scalar product and norm in $\vec{I\!E}^3$ by[7]

$$\begin{aligned} \langle \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c &= \langle (\vec{u}_1 + i\vec{v}_1) \mid (\vec{u}_2 + i\vec{v}_2) \rangle^c \\ &:= \langle \vec{u}_1 \mid \vec{u}_2 \rangle + \langle \vec{v}_1 \mid \vec{v}_2 \rangle + i \langle \vec{u}_1 \mid \vec{v}_2 \rangle - i \langle \vec{v}_1 \mid \vec{u}_2 \rangle, \end{aligned} \tag{11.14}$$

$$\begin{aligned} ||\vec{\zeta}\,||^c &= ||\vec{u} + i\vec{v}\,||^c \\ &:= \sqrt{\langle \vec{u} + i\vec{v} \mid \vec{u} + i\vec{v} \rangle^c} = \sqrt{\langle \vec{u} \mid \vec{u} \rangle + \langle \vec{v} \mid \vec{v} \rangle} \\ &= \sqrt{||\vec{u}\,||^2 + ||\vec{v}\,||}. \end{aligned} \tag{11.15}$$

[6] Making use of Eq. (11.7).

[7] For real vectors $\vec{u}^c, \vec{v}^c$ in $\vec{I\!E}^{3c}$ we have $\langle \vec{u}^c \mid \vec{v}^c \rangle^c = \langle \vec{u} \mid \vec{v} \rangle$ and $||\vec{u}^{\,c}\,|| = ||\vec{u}\,||$.

The present definition possesses the following properties:

CSP11.2.2(1)	*Non-commutative*	$\langle \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c = \langle \vec{\zeta}_2 \mid \vec{\zeta}_1 \rangle^{c*}$.
CSP11.2.2(2)	*Distributive*	$\langle \vec{\zeta} \mid (c_1 \vec{\zeta}_1 + c_2 \vec{\zeta}_2) \rangle^c$ $= c_1 \langle \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c + c_2 \langle \vec{\zeta} \mid \vec{\zeta}_2 \rangle^c$, where $c_1,\ c_2 \in \mathbb{C}$.
CSP11.2.2(3)	*Positive definite*	$\langle \vec{\zeta} \mid \vec{\zeta} \rangle^c$ is real, non-negative, and $\langle \vec{\zeta} \mid \vec{\zeta} \rangle = 0 \quad \Rightarrow \quad \vec{\zeta} = \vec{0}$.

Note and that CSP11.2.2(1) and CSP11.2.2(2) imply[8]

$$\langle \vec{\zeta}_1 \mid c\,\vec{\zeta}_2 \rangle^c = c\,\langle\, \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c \ \text{ and } \ \langle c\,\vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c = c^* \langle\, \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c. \quad (11.16)$$

The definitions of orthonormality and orthonormal basis given in §6.2 carry over to $\vec{I\!E}^{3c}$. An orthonormal basis $\{\vec{e}_\ell\}$ for $\vec{I\!E}^3$ correspond to an orthonormal basis $\{\vec{e}_\ell^{\,c}\}$ for $\vec{I\!E}^{3c}$, as shown in Eq. (11.12).[9] We can rewrite Eq. (11.12) as

$$\vec{\zeta} = \vec{u} + i\vec{v} = \sum_{\ell=1}^{3} \zeta_\ell \vec{e}_\ell^{\,c}, \quad \text{with} \ \ \zeta_\ell = \langle \vec{e}_\ell^{\,c} \mid \vec{\zeta} \rangle^c = u_\ell + i v_\ell. \quad (11.17)$$

For example, the basis $\{\vec{i},\ \vec{j},\ \vec{k}\}$ in $\vec{I\!E}^3$ correspond to an orthonormal basis $\{\vec{i}^{\,c},\ \vec{j}^{\,c},\ \vec{k}^{\,c}\}$ for $\vec{I\!E}^{3c}$. We can also have complex vectors forming an orthonormal basis in $\vec{I\!E}^{3c}$.

The Schwarz inequality and the triangle inequalities in §6.3.6 remain valid.[10] Following the expansion in Eq. (11.17) the Pythagoras theorem now takes the form

$$||\vec{\zeta}\,||^c = \Big(\sum_{\ell=1}^{3} \zeta_\ell^* \zeta_\ell \Big)^{1/2}. \quad (11.18)$$

$$\langle \vec{\zeta}_1 \mid \vec{\zeta}_2 \rangle^c = \sum_{\ell=1}^{3} \zeta_{1\ell}^* \zeta_{2\ell}. \quad (11.19)$$

The discussion on matrix representation of vectors in §7.5 can be extended to to complex vectors in $\vec{I\!E}^{3c}$. Generally complex vectors

[8] Apart from CSP11.2.2(1) these are the same as the properties of the scalar product for vectors in $\vec{I\!E}^3$ listed in §6.3.1.

[9] We also have $\langle \vec{e}_\ell^{\,c} \mid \vec{e}_{\ell'}^{\,c} \rangle^c = \delta_{\ell\ell'}$.

[10] *See* Q12(1) and Q12(3) in Exercise and Problems for Chapter 12.

will be represented by complex column vectors. For example, in the basis $\{\vec{i}^{\,c}, \vec{j}^{\,c}, \vec{k}^{\,c}\}$ any vector in $\vec{I\!E}^{3c}$ has the expansion

$$\vec{\zeta} = \zeta_x \vec{i}^{\,c} + \zeta_y \vec{j}^{\,c} + \zeta_z \vec{k}^{\,c} \tag{11.20}$$

with

$$\zeta_x = \langle\, \vec{i}^{\,c} \mid \vec{\zeta}\,\rangle^c, \;\; \zeta_y = \langle\, \vec{j}^{\,c} \mid \vec{\zeta}\,\rangle^c, \;\; \zeta_z = \langle\, \vec{k}^{\,c} \mid \vec{\zeta}\,\rangle^c. \tag{11.21}$$

In basis $\left\{\vec{i}^{\,c}, \vec{j}^{\,c}, \vec{k}^{\,c}\right\}$ the matrix representation of the basis vectors $\vec{i}^{\,c}, \vec{j}^{\,c}, \vec{k}^{\,c}$ and of $\vec{\zeta}$ are given respectively by[11]

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \zeta_x \\ \zeta_y \\ \zeta_z \end{pmatrix}. \tag{11.22}$$

While complex vectors in $\vec{I\!E}^{3c}$ does not have the usual geometric meaning of real vectors in physical space we can still imagine a complex vector defining a "direction" in an abstract sense. The discussion on operators in Chapters 8 and 9 as well as the discussion on probability in Chapter 10 can also be extended to $\vec{I\!E}^{3c}$. We shall go into these in more details in later chapters after we generalise the concept of complex vectors to arbitrary dimensions.

Exercises and Problems

Q11(1) Prove Eqs. (11.17), (11.18) and (11.19) in $\vec{I\!E}^{3c}$.

Q11(2) Let

$$\vec{\epsilon}_1 = \frac{1}{\sqrt{2}}\left(\vec{i}^{\,c} + i\,\vec{j}^{\,c}\right), \quad \vec{\epsilon}_2 = \frac{1}{\sqrt{2}}\left(\vec{i}^{\,c} - i\,\vec{j}^{\,c}\right), \quad \vec{\epsilon}_3 = \vec{k}^{\,c}.$$

(a) Write down the matrix representations of the above vectors in the orthonormal basis $\{\vec{i}^{\,c}, \vec{j}^{\,c}, \vec{k}^{\,c}\}$.

(b) Show that $\vec{\epsilon}_\ell$, $\ell = 1, 2, 3$ above form an orthonormal basis for $\vec{I\!E}^{3c}$.

[11] We can have complex basis vectors, e.g., the column vectors in Eq. (14.52).

(c) Show that an arbitrary vector $\vec{\zeta}$ can be expressed as

$$\vec{\zeta} = \sum_{\ell=1}^{3} \zeta_\ell \, \vec{\epsilon}_\ell, \quad \zeta_\ell = \langle \vec{\epsilon}_\ell \mid \vec{\zeta} \, \rangle^c. \tag{11.23}$$

Q11(3) Show that Eqs. (11.18) and (11.19) hold in a general complex orthonormal basis.

Chapter 12

N-Dimensional Complex Vector Spaces

12.1 Introductory Remarks

So far we have adopted a *constructive method* to define complex vectors. Using the familiar three-dimensional real vectors in $I\!\vec{E}^3$ we can construct three-dimensional complex vectors explicitly, e.g., we define their algebraic operations and scalar product explicitly. Important properties can then be derived from all these explicit definitions, e.g., CA11.2.1(1) to CA11.2.1(4) and CSP11.2.2(1) to CSP11.2.2(3). This constructive method provides an intuitive way to define things. However, such a method is not useful when we want to generalise our theory. Many of the explicit constructions cannot be easily generalised.

When extending to higher dimensions what we want is to preserve the intrinsic and desirable properties of our three-dimensional complex vector space $I\!\vec{E}^{3c}$, but not necessarily the constructive expressions. The idea is to define various quantities, including vector spaces themselves, by their properties from the start. We shall start by defining an N-dimensional complex vector space as a set of elements endowed with certain properties. There is no need to describe these elements explicitly as long they possess the desired properties.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

12.2 Definitions

Definition 12.2(1) *A set $\vec{\mathbb{V}}$ of elements $\vec{\zeta}, \vec{\eta}, \ldots$, is called a complex vector space and its elements are called complex vectors if*

(1) *An operation (addition) is defined which assigns to every pair of elements $\vec{\zeta}$ and $\vec{\eta}$ a new element in $\vec{\mathbb{V}}$, denoted by $\vec{\zeta} + \vec{\eta}$ and called the sum of $\vec{\zeta}$ and $\vec{\eta}$, in such a way that properties* CA11.2.1(1) *to* CA11.2.1(4) *are satisfied.*

(2) *An operation (scalar multiplication) is defined which assigns to every element $\vec{\zeta}$ and a scalar $c \in \mathbb{C}$ a new element in $\vec{\mathbb{V}}$, denoted by $c\,\vec{\zeta}$, in such a way that properties* CSM11.2.1(1) *to* CSM11.2.1(3) *are satisfied.*[1]

The definition of linear dependence and independence, dimensions and bases (complete sets) are the same as those given in §6.2.1 and §6.2.2. An N-dimensional space, denoted by $\vec{\mathbb{V}}^N$, is one which possesses a maximum of N linearly independent vectors. It is possible to have an infinite number of linearly independent vectors in a space. In such a case the space is *infinite-dimensional*. The notation $\vec{\mathbb{V}}^N$ is used when N is finite. When the dimension is infinite the notation $\vec{\mathbb{V}}^\infty$ is used.

Definition 12.2(2)

(1) *A* **scalar product** *on $\vec{\mathbb{V}}^N$ is an assignment of a scalar to each pair of vectors $\vec{\zeta}$ and $\vec{\eta}$ in $\vec{\mathbb{V}}^N$, denoted by $\langle \vec{\zeta} \mid \vec{\eta} \rangle$, satisfying properties* CSP11.2.2(1) , CSP11.2.2(2) *and* CSP11.2.2(3).[2]

(2) *A vector space $\vec{\mathbb{V}}^N$ with a scalar product defined on it is called a* **scalar product space**.[3]

[1] The set becomes a *real vector space* if the scalar multiplication is restricted to real numbers.

[2] The symbol $\langle \vec{\zeta} \mid \vec{\eta} \rangle$ instead of $\langle \vec{\zeta} \mid \vec{\eta} \rangle^c$ is used to simplify the notation.

[3] A scalar product is often referred to as an *inner product*, with a scalar product space referred to as an *inner product space*.

A scalar product $\langle \vec{\zeta} \mid \vec{\eta} \rangle$ is generally complex. It has the following properties:

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle = \langle \vec{\eta} \mid \vec{\zeta} \rangle^*, \tag{12.1}$$

$$\langle \vec{\zeta} \mid c\,\vec{\eta} \rangle = c\,\langle \vec{\zeta} \mid \vec{\eta} \rangle, \quad \langle c\,\vec{\zeta} \mid \vec{\eta} \rangle = c^*\langle \vec{\zeta} \mid \vec{\eta} \rangle. \tag{12.2}$$

Once a scalar product is defined on $\vec{\mathbb{V}}^N$ we can define the concepts of norm, orthonormality and orthonormal bases (complete orthonormal sets) as before.[4] Orthogonal vectors are linearly independent. The Gram-Schmidt orthogonalisation procedure in §6.3.5 applies to ensure the existence of an orthonormal basis consisting N orthonormal vectors $\vec{\varepsilon}_\ell$, $\ell = 1, 2, \cdots, N$, such that any vector $\vec{\zeta} \in \vec{\mathbb{V}}^N$ can be expressed as

$$\vec{\zeta} = \sum_{\ell=1}^{N} \zeta_\ell \vec{\varepsilon}_\ell, \quad \zeta_\ell = \langle \vec{\varepsilon}_\ell \mid \vec{\zeta} \rangle. \tag{12.3}$$

It immediately follows that Eqs. (6.30) and (6.31) for $\vec{E}^3$ apply here, i.e., we have

$$\langle \vec{\varepsilon}_\ell \mid \vec{\zeta} \rangle = 0 \ \forall \ell \quad \Rightarrow \quad \vec{\zeta} = \vec{0} \tag{12.4}$$

and

$$\langle \vec{\varepsilon}_\ell \mid \vec{\zeta} \rangle = \langle \vec{\varepsilon}_\ell \mid \vec{\eta} \rangle \ \forall \ell \quad \Rightarrow \quad \vec{\zeta} = \vec{\eta}. \tag{12.5}$$

The discussion on matrix representation of vectors in §7.5 can be extended to complex vectors in $\vec{\mathbb{V}}^N$. Generally complex vectors will be represented by complex column vectors.

In a complex vector space we can express a scalar product in terms of norms as follows[5]:

$$\begin{aligned}\langle \vec{\zeta} \mid \vec{\eta} \rangle = \frac{1}{2}\Big\{ &\langle \vec{\zeta} + \vec{\eta} \mid \vec{\zeta} + \vec{\eta} \rangle - i\langle \vec{\zeta} + i\vec{\eta} \mid \vec{\zeta} + i\vec{\eta} \rangle \\ &+(i-1)\big(\langle \vec{\zeta} \mid \vec{\zeta} \rangle + \langle \vec{\eta} \mid \vec{\eta} \rangle\big)\Big\}.\end{aligned} \tag{12.6}$$

[4] We define a norm in $\vec{\mathbb{V}}^N$ by Eq. (11.15). It is possible to define a norm independent of scalar product (*see* Weidmann p. 7).

[5] *See* Q12(1) for an alternative expression and also see Eq. (13.7) for a similar equality involving an operator. Note that

$$\langle \vec{\zeta} + \vec{\eta} \mid \vec{\zeta} + \vec{\eta} \rangle = ||\vec{\zeta} + \vec{\eta}||^2, \quad \langle \vec{\zeta} + i\vec{\eta} \mid \vec{\zeta} + i\vec{\eta} \rangle = ||\vec{\zeta} + i\vec{\eta}||^2.$$

12.3 Examples

12.3.1 Complex Column Matrices as Vectors

Consider the set $\mathbb{C}^N$ of $N \times 1$ column matrices of complex elements:

$$\mathbb{C}^N := \left\{ \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \cdot \\ \cdot \\ \cdot \\ \zeta_N \end{pmatrix} : \quad \zeta_\ell \in \mathbb{C}, \quad \ell = 1, 2, \cdots, N \right\}. \tag{12.7}$$

We can define addition and scalar multiplication by

$$\begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \cdot \\ \cdot \\ \cdot \\ \zeta_N \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdot \\ \cdot \\ \cdot \\ \eta_N \end{pmatrix} := \begin{pmatrix} \zeta_1 + \eta_1 \\ \zeta_2 + \eta_2 \\ \cdot \\ \cdot \\ \cdot \\ \zeta_N + \eta_N \end{pmatrix}, \tag{12.8}$$

$$c \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \cdot \\ \cdot \\ \cdot \\ \zeta_N \end{pmatrix} := \begin{pmatrix} c\,\zeta_1 \\ c\,\zeta_2 \\ \cdot \\ \cdot \\ \cdot \\ c\,\zeta_N \end{pmatrix}. \tag{12.9}$$

These operations satisfy properties CA11.2.1(1) to CA11.2.1(4) and CSM11.2.1(1) to CSM11.2.1(3). By Definition 12.2 (1) the set becomes an N-dimensional complex scalar product space, to be denoted by $\vec{\mathbb{C}}^N$, with column matrices as vectors. Hence we shall denote these column matrices in vector notation as $\vec{\zeta}, \vec{\eta}, \ldots$ and call them *column vectors*.

We can assign a complex number $\langle \vec{\zeta} \mid \vec{\eta} \rangle$ to every pair of vectors $\vec{\zeta}, \vec{\eta}$ in $\vec{\mathbb{C}}^N$ by[6]

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle := \zeta_1^* \eta_1 + \zeta_2^* \eta_2 + \cdots + \zeta_N^* \eta_N = \sum_{\ell=1}^{N} \zeta_\ell^* \eta_\ell. \tag{12.10}$$

[6] The notation $\langle \vec{\zeta} \mid \vec{\eta} \rangle$ anticipates the assignment to form a scalar product.

This assignment of a scalar to every pair of vectors in $\vec{\mathbb{C}}^N$ satisfy properties CSP11.2.2(1) to CSP11.2.2(3). Hence the assignment constitutes a scalar product. We shall denote resulting scalar product space also by $\vec{\mathbb{C}}^N$.

The complex conjugates in the expression for the scalar product ensure that

$$\langle \vec{\zeta} \mid \vec{\zeta} \rangle = \sum_{\ell=1}^{N} \zeta_\ell^* \zeta_\ell \tag{12.11}$$

is real and non-negative. We can then define norm of a vector to be

$$||\vec{\zeta}|| := \sqrt{\langle \vec{\zeta} \mid \vec{\zeta} \rangle} = \left(\sum_{\ell=1}^{N} \zeta_\ell^* \zeta_\ell \right)^{1/2}. \tag{12.12}$$

Two vectors $\vec{\zeta}$, $\vec{\eta}$ are orthonormal if

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle = 0 \quad \text{and} \quad \|\vec{z}\| = \|\vec{\eta}\| = 1. \tag{12.13}$$

Equations (12.11) and (12.12) are the Pythagoras theorem in $\vec{\mathbb{C}}^N$.

We can see that $\vec{\mathbb{C}}^N$ is N-dimensional with an obvious orthonormal basis[7]

$$\vec{e}_1^{\,c} := \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \quad \vec{e}_2^{\,c} := \begin{pmatrix} 0 \\ 1 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}, \quad \ldots, \quad \vec{e}_N^{\,c} := \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 1 \end{pmatrix}. \tag{12.14}$$

An arbitrary vector has the familiar expansion

$$\vec{\zeta} = \sum_{\ell=1}^{N} z_\ell \vec{e}_\ell^{\,c}, \quad \zeta_\ell = \langle \vec{e}_\ell^{\,c} \mid \vec{\zeta} \rangle \in \mathbb{C}. \tag{12.15}$$

When $N = 2$ we have a simple but important case of a two-dimensional space $\vec{\mathbb{C}}^2$ consisting of 2×1 column vectors, i.e.,

$$\vec{\mathbb{C}}^2 := \left\{ \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} : \quad \zeta_\ell \in \mathbb{C}, \quad \ell = 1, 2 \right\}, \tag{12.16}$$

[7] Using the notation in §11.2.2.

with scalar product of two such vectors $\vec{\zeta}$ and $\vec{\eta}$ given by

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle := \zeta_1^* \eta_1 + \zeta_2^* \eta_2. \tag{12.17}$$

If we restrict the coefficients in the linear combinations to real numbers in Eq. (12.15) we would obtain a set $I\!R^N$ of $N \times 1$ column matrices of real elements. This set forms an N-dimensional real vector space denoted by $\vec{I\!R}^N$. Explicitly we have

$$\vec{I\!R}^N := \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_N \end{pmatrix} : \quad a_\ell \in I\!R, \quad \ell = 1, 2, \cdots, N \right\}. \tag{12.18}$$

12.3.2 Complex $N \times N$ Matrices as Vectors

Consider the set of $N \times N$ complex matrices $\boldsymbol{M}$, $\boldsymbol{N}$, $\cdots$. With usual matrix addition and scalar multiplication the set becomes a vector space. To be specific let us consider the set of 2×2 complex matrices. There is a maximum of 4 linearly independent 2×2 matrices, e.g., in vector notation

$$\vec{\boldsymbol{m}}_{\boldsymbol{1}} := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \vec{\boldsymbol{m}}_{\boldsymbol{2}} := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{12.19}$$

$$\vec{\boldsymbol{m}}_{\boldsymbol{3}} := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \vec{\boldsymbol{m}}_{\boldsymbol{4}} := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{12.20}$$

such that any 2×2 complex matrix (in vector notation)

$$\vec{\boldsymbol{M}} := \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \tag{12.21}$$

can be written as a linear combination

$$\vec{\boldsymbol{M}} = M_{11}\vec{\boldsymbol{m}}_{\boldsymbol{1}} + M_{12}\vec{\boldsymbol{m}}_{\boldsymbol{2}} + M_{21}\vec{\boldsymbol{m}}_{\boldsymbol{3}} + M_{22}\vec{\boldsymbol{m}}_{\boldsymbol{4}}. \tag{12.22}$$

The set of matrices constitutes a four-dimensional vector space with $\{\vec{\boldsymbol{m}}_{\boldsymbol{\ell}}, \ell = 1, 2, 3, 4\}$ forming a basis. We can define a scalar product, known as the *Frobenius scalar product*,[8] by

$$\langle \vec{\boldsymbol{M}} \mid \vec{\boldsymbol{N}} \rangle := M_{11}^* N_{11} + M_{12}^* N_{12} + M_{21}^* N_{21} + M_{22}^* N_{22}. \tag{12.23}$$

Then the set $\{\vec{\boldsymbol{m}}_{\boldsymbol{\ell}}, \ \ell = 1, 2, 3, 4\}$ forms an orthonormal basis.

[8] Frobenius (1849–1917) was a German mathematician.

12.3.3 Complex-Valued Functions as Vectors

Consider the set $C(\Lambda)$ of continuous complex-valued functions $\phi(x)$, $\psi(x)$, $\cdots$ of a real variable x defined on a closed interval $\Lambda = [0, L]$.[9] The set is closed under addition and scalar multiplication, i.e., $\phi + \psi$ and $c\,\phi$, $c \in \mathbb{C}$ are again members of $C(\Lambda)$. It follows that the set $C(\Lambda)$ forms a complex vector space and we can consider functions in $C(\Lambda)$ as vectors. To emphasise the vector nature of these functions we shall denote them with vector notation,[10] e.g., $\vec{\phi}$ and $\vec{\psi}$ for $\phi(x)$ and $\psi(x)$. This relationship is denoted by

$$\boxed{\vec{\phi} := \phi(x) \quad \text{and} \quad \vec{\psi} := \psi(x)} \tag{12.24}$$

The symbol $:=$ means that the vectors on the left hand side are defined by the functions on the right hand side. Note that the variable x does not appear in the notation for the vector. Vectors $\vec{\phi}$, $\vec{\psi}$ are also said to correspond to functions $\phi(x)$ and $\psi(x)$.

It is easy to check that the assignment of a (complex) number to each pair of functions ϕ and ψ in $C(\Lambda)$ in terms of the following Riemann integral[11]

$$\int_0^L \phi^*(x)\psi(x)dx \tag{12.25}$$

satisfies CSP11.2.2(1) to CSP11.2.2(3) of scalar product. We can use Eq. (12.25) to define a scalar product in $C(\Lambda)$ with the notation

$$\langle \vec{\phi} \mid \vec{\psi} \rangle := \int_0^L \phi^*(x)\psi(x)dx. \tag{12.26}$$

The norm of vectors is given by

$$||\vec{\phi}\,||^2 := \langle \vec{\phi} \mid \vec{\phi} \rangle = \int_0^L \phi^*(x)\phi(x)dx. \tag{12.27}$$

Two vectors $\vec{\phi}$, $\vec{\psi}$ are orthonormal if $||\vec{\phi}\,||^2 = ||\vec{\psi}\,||^2 = 1$ and

$$\langle \vec{\phi} \mid \vec{\psi} \rangle = \int_0^L \phi^*(x)\psi(x)dx = 0. \tag{12.28}$$

[9] Weidmann p. 5. A complex-valued function of a real variable is decomposable into the form $\phi(x) = f(x) + i\,g(x)$ where f and g are real-valued functions of x. The function $\phi(x)$ is continuous in x if $f(x)$ and $g(x)$ are. The same applies to differentiability and integrability.

[10] Following Isham p. 28.

[11] The integral exists for continuous functions over a finite interval.

When using the notation $\vec{\phi}$ we are treating the function $\phi(x)$ as a whole as a vector. These functions define a set of vectors which would form a vector space which is denoted by $\vec{C}(\Lambda)$. We can express this relationship by $\vec{C}(\Lambda) := C(\Lambda)$.[12] Actual calculations are carried out in terms of $\phi(x)$, $\psi(x)$ as functions as in Eqs. (12.26) and (12.27).

It is not obvious how we can obtain a basis to span such a set of continuous functions and what the dimension of the space is. To make things easier here we shall do the opposite. Instead of looking for a basis for $\vec{C}(\Lambda)$ we shall select a set of n functions $\varphi_\ell(x)$ in $C(\Lambda)$ to define n orthonormal vectors $\vec{\varphi}_\ell$. These vectors can be used to construct an n-dimensional vector space. As an example consider the following functions defined on the interval $\Lambda = [0, L]$:

$$\varphi_\ell(x) = \frac{1}{\sqrt{L}}\, e^{i(2\pi\ell/L)x}, \quad \ell = 0, \pm 1, \pm 2, \cdots . \tag{12.29}$$

Being continuous these functions are members of $C(\Lambda)$. They define a corresponding set of vectors $\vec{\varphi}_\ell$ in $\vec{C}(\Lambda)$. As vectors they are also orthonormal, and hence they are linearly independent of each other. Let us consider the three vectors defined by $\varphi_{-1}(x)$, $\varphi_0(x)$ and $\varphi_1(x)$, i.e.,

$$\vec{\varphi}_{-1} := \frac{1}{\sqrt{L}}\, e^{-i(2\pi/L)x}, \quad \vec{\varphi}_0 := \frac{1}{\sqrt{L}}, \quad \vec{\varphi}_1 := \frac{1}{\sqrt{L}}\, e^{i(2\pi/L)x}.$$

Their linear combinations would produce a set of vectors, i.e.,

$$\left\{ \vec{\phi} := c_{-1}\vec{\varphi}_{-1} + c_0\vec{\varphi}_0 + c_1\vec{\varphi}_1, \quad c_{-1},\ c_0,\ c_1 \in \mathbb{C} \right\}.$$

This set of vectors forms a three-dimensional complex vector space in its own right with $\{\vec{\varphi}_\ell, \ell = -1, 0, 1\}$ as an orthonormal basis. Clearly the discussion can be extended to form vector spaces of higher dimensions by using a larger set of vectors $\vec{\varphi}_\ell$. For example, the set of functions $\varphi_\ell(x)$ in Eq. (12.29) for $\ell = 0, \pm 1, \pm 2, \cdots, \pm m$ would define a set of $n = 2m + 1$ orthonormal vectors $\vec{\varphi}_\ell$. These

[12] It is common to employ the same notation, e.g., $\phi(x)$ and $C(\Lambda)$ to denote the vector and the vector space.

vectors can be used to span an n-dimensional vector space

$$\left\{ \vec{\phi} = \sum_{\ell=-m}^{m} c_\ell \, \vec{\varphi}_\ell := \sum_{\ell=-m}^{m} c_\ell \, \frac{1}{\sqrt{L}} \, e^{i(2\ell\pi/L)\,x}, \quad c_\ell \in \mathbb{C} \right\}. \tag{12.30}$$

This is a subset of $\vec{C}(\Lambda)$. We can make the subset bigger and bigger by increasing n without limit. Intuitively we can see that $\vec{C}(\Lambda)$ must be an infinite-dimensional space.

12.4 Isomorphism between Spaces

Let f be a mapping of a given space $\vec{\mathbb{W}}_1^N$ onto another space $\vec{\mathbb{W}}_2^N$. The mapping is said to be linear if it has the property that when $\vec{\zeta}_1$ and $\vec{\eta}_1$ of $\vec{\mathbb{W}}_1^N$ are mapped onto $\vec{\zeta}_2 = f(\vec{\zeta}_1)$ and $\vec{\eta}_2 = f(\vec{\eta}_1)$ of $\vec{\mathbb{W}}_2^N$ then all their linear combinations $a\,\vec{\zeta}_1 + b\,\vec{\eta}_1$ in $\vec{\mathbb{W}}_1^N$ are mapped onto the corresponding linear combinations of $\vec{\zeta}_2$ and $\vec{\eta}_2$ in $\vec{\mathbb{W}}_2^N$, i.e., we have, $\forall a, b \in \mathbb{C}$,

$$f\left(a\,\vec{\zeta}_1 + b\,\vec{\eta}_1\right) = a\,f(\vec{\zeta}_1) + b\,f(\vec{\eta}_1) = a\,\vec{\zeta}_2 + b\,\vec{\eta}_2. \tag{12.31}$$

Let $\{\vec{\varphi}_{1\ell}\}$ be an orthonormal basis for $\vec{\mathbb{W}}_1^N$ and let $\{\vec{\varphi}_{2\ell}\}$ be an orthonormal basis for $\vec{\mathbb{W}}_2$. We can define a linear mapping of $\vec{\mathbb{W}}_1^N$ onto $\vec{\mathbb{W}}_2^N$ by first mapping the basis vectors $\vec{\varphi}_{1\ell}$ of $\vec{\mathbb{W}}_1^N$ to the basis vectors $\vec{\varphi}_{2\ell}$ of $\vec{\mathbb{W}}_2^N$, and then extend the mapping by linearity to the entire $\vec{\mathbb{W}}_1^N$. Complex vector spaces of the same dimension can be mapped onto each other in this way.

A one-to-one linear mapping of $\vec{\mathbb{W}}_1^N$ onto $\vec{\mathbb{W}}_2^N$ is called an *isomorphism* between $\vec{\mathbb{W}}_1^N$ and $\vec{\mathbb{W}}_2^N$. The two spaces are then said to be *isomorphic*.[13] An isomorphism can be demonstrated by relating an orthonormal basis in $\vec{\mathbb{W}}_1^N$ to that of $\vec{\mathbb{W}}_2^N$. For example, we can establish an isomorphism between the N-dimensional vector space in Eq. (12.30) spanned by $\vec{\varphi}_\ell$ and $\vec{\mathbb{C}}^N$ spanned by $\vec{e}_\ell^{\,c}$. We can achieve this by first mapping $\vec{\varphi}_\ell$ onto $\vec{e}_\ell^{\,c}$, i.e., we start with $f(\vec{\varphi}_\ell) = \vec{e}_\ell^{\,c}$. Then

[13] Halmos p. 14, Isham p. 22.

for any $\vec{\phi} = \sum_{\ell=1}^{N} c_\ell \, \vec{\varphi}_\ell$ we have

$$f(\vec{\phi}) = f\Big(\sum_{\ell=1}^{N} c_\ell \, \vec{\varphi}_\ell\Big) = \sum_{\ell=1}^{N} c_\ell \, f\big(\vec{\varphi}_\ell\big) = \sum_{\ell=1}^{N} c_\ell \, \vec{e}_\ell^{\,c}. \tag{12.32}$$

The mapping establishes an isomorphism between the two spaces. All N-dimensional vector spaces $\vec{\mathbb{V}}^N$ are isomorphic to $\vec{\mathbb{C}}^N$ in a similar way. This isomorphism enables us to establish a matrix representation of $\vec{\mathbb{V}}^N$, i.e., we can represent vectors in $\vec{\mathbb{V}}^N$ by column vectors in $\vec{\mathbb{C}}^N$.

An isomorphism, i.e., a one-to-one linear mapping of a vector space $\vec{\mathbb{V}}_1^N$ onto another vector space $\vec{\mathbb{V}}_2^N$, is said to be **unitary** if it preserves the scalar product, i.e., if

$$\langle \vec{\zeta}_1 \mid \vec{\eta}_1 \rangle_1 = \langle \vec{\zeta}_2 \mid \vec{\eta}_2 \rangle_2, \tag{12.33}$$

where $\langle \vec{\zeta}_1 \mid \vec{\eta}_1 \rangle_1$ denotes scalar product of two vectors $\vec{\zeta}_1$ and $\vec{\eta}_1$ in $\vec{\mathbb{V}}_1^N$ and $\langle \vec{\zeta}_2 \mid \vec{\eta}_2 \rangle_2$ denotes scalar product of two vectors $\vec{\zeta}_2$ and $\vec{\eta}_2$ in $\vec{\mathbb{V}}_2^N$. An example is the mapping in Eq. (12.32) which is clearly unitary.

12.5 Concluding Remarks

The examples in §12.3 suggest that one can go one step further to form infinite-dimensional vector spaces, e.g., $\vec{\mathbb{C}}^N$ becomes an infinite-dimensional vector space as N becomes arbitrarily large. Functions in $C(\Lambda)$ also form an infinite-dimensional vector space since not every continuous function on the interval Λ can be expressed as a linear combination of a finite set of functions like those in Eq. (12.29). Then things can become very complicated since many of the definitions introduced for finite-dimensional spaces cannot be applied to infinite-dimensional spaces. For example, the scalar product in Eq. (12.10) may produce an infinite value when summing over an infinite number of terms. We shall return to discuss infinite-dimensional spaces after reviewing operators in N-dimensional spaces in the next chapter.

Exercises and Problems

Q12(1) Prove Eq. (12.6).

Q12(2) Show that

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle = \frac{1}{4}\Big\{ ||\vec{\zeta} + \vec{\eta}||^2 - ||\vec{\zeta} - \vec{\eta}||^2 - i\left(||\vec{\zeta} + i\vec{\eta}||^2 - ||\vec{\zeta} - i\vec{\eta}||^2 \right) \Big\}. \tag{12.34}$$

Q12(3) Prove the Schwarz inequality in $\vec{\mathbb{V}}^N$[14]:

$$|\langle \vec{\zeta} \mid \vec{\eta} \rangle| \leq ||\vec{\zeta}||\,||\vec{\eta}||. \tag{12.35}$$

Q12(4) Prove the following triangle inequalities in $\vec{\mathbb{V}}^N$[15]:

$$||\vec{\zeta} + \vec{\eta}|| \leq ||\vec{\zeta}|| + ||\vec{\eta}||. \tag{12.36}$$

$$\Big|\, ||\vec{\zeta}|| - ||\vec{\eta}|| \,\Big| \leq ||\vec{\zeta} - \vec{\eta}||. \tag{12.37}$$

Q12(5) Show that the Frobenius expression in Eq. (12.23) satisfies the properties CSP11.2.2(1), CSP11.2.2(2) and CSP11.2.2(3) of a scalar product.

Q12(6) Show that the integral expression in Eq. (12.25) satisfies the properties CSP11.2.2(1), CSP11.2.2(2) and CSP11.2.2(3) of a scalar product.

Q12(7) Verify that the vectors $\vec{\varphi}_\ell$ corresponding to the functions $\varphi_\ell(x)$ in Eq. (12.29) are orthonormal.

Q12(8) Show that two orthogonal vectors are linearly independent.

Q12(9) Show that the three Pauli matrices σ_x, σ_y and σ_z in Eq. (7.9) together with the 2×2 identity matrix form a basis for the vector space of 2×2 complex matrices with Frobenius scalar product. Is this an orthonormal basis?

Q12(10) Show that the mapping in Eq. (12.32) is unitary.

---

[14] Roman p. 419. Prugovečki pp. 19–20.

[15] Work out $||\vec{\zeta} + \vec{\eta}||^2 = \langle \vec{\zeta} + \vec{\eta} \mid \vec{\zeta} + \vec{\eta} \rangle$ and $||\vec{\zeta} - \vec{\eta}||^2 = \langle \vec{\zeta} - \vec{\eta} \mid \vec{\zeta} - \vec{\eta} \rangle$. Note that $\langle \vec{\zeta} \mid \vec{\eta} \rangle + \langle \vec{\eta} \mid \vec{\zeta} \rangle = 2\times$ the real part of $\langle \vec{\zeta} \mid \vec{\eta} \rangle$ which is less than or equal to $2|\langle \vec{\zeta} \mid \vec{\eta} \rangle|$.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 13

Operators on *N*-Dimensional Complex Vector Spaces

13.1 Introduction

The contents of Chapters 8 and 9 can be generalised to N-dimensional scalar product spaces $\vec{\mathbb{V}}^N$. For example, with real numbers extended to complex numbers we can define a linear functional on $\vec{\mathbb{V}}^N$ to be a linear mapping of $\vec{\mathbb{V}}^N$ into $\mathbb{C}$ and Riesz theorem remains valid.

As in $\vec{I\!E}^3$ we shall assume from the outset that operators defined on a finite-dimensional space $\vec{\mathbb{V}}^N$ act on every vector in the space, i.e., we shall assume that all operators have $\vec{\mathbb{V}}^N$ as their domain. But not all operators would have $\vec{\mathbb{V}}^N$ as their range. With $\vec{I\!E}^3$ and $I\!R$ replaced respectively by $\vec{\mathbb{V}}^N$ and $\mathbb{C}$ the definition of linear operators in Eq. (8.10) and the properties these operators given in §8.2.2 apply here:

P13.1(1) The definition of the norm of an operator in Eq. (8.13) applies. The definitions of algebraic operations of operators, i.e., Eqs. (8.15) to (8.18), the definition of linear combinations and the condition for the equality of two operators, remain valid here.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

P13.1(2) The definitions of the commutator and anticommutator of two operators given respectively by Eqs. (8.19) and (8.21) apply. The definitions of commuting and anticommuting operators also apply here.

P13.1(3) The definition of inverse operators and the property of inverse operators in Eq. (8.23), Theorems 8.2.2(1) and (2) and the comments after the theorems on inverse operators remain valid.

P13.1(4) The definition of the adjoint of an operator by Eq. (8.33) applies here, i.e., the adjoint of an operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ is defined to be the operator $\widehat{A}^\dagger$ satisfying

$$\langle \widehat{A}^\dagger \vec{\zeta} \mid \vec{\eta} \rangle = \langle \vec{\zeta} \mid \widehat{A}\vec{\eta} \rangle \quad \forall\, \vec{\zeta},\ \vec{\eta} \in \vec{\mathbb{V}}^N. \tag{13.1}$$

By taking the complex conjugates of the scalar products we get, as in Eq. (8.34),

$$\langle \vec{\eta} \mid \widehat{A}^\dagger \vec{\zeta} \rangle = \langle \widehat{A}\vec{\eta} \mid \vec{\zeta} \rangle \quad \forall\, \vec{\zeta},\ \vec{\eta} \in \vec{\mathbb{V}}^N. \tag{13.2}$$

The properties of the adjoint operation listed from Eq. (8.38) to Eq. (8.42) apply. In particular the linear property of the adjoint operation in Eq. (8.38) can be generalised to[1]

$$\left(c_1\widehat{A}_1 + c_2\widehat{A}_2\right)^\dagger = c_1^*\widehat{A}_1^\dagger + c_2^*\widehat{A}_2^\dagger, \quad c_1,\ c_2 \in \mathbb{C}, \tag{13.3}$$

and Eq. (8.41) on the product of two operators remains true, i.e.,

$$\left(\widehat{A}\widehat{B}\right)^\dagger = \widehat{B}^\dagger\widehat{A}^\dagger, \tag{13.4}$$

P13.1(5) Definition 9.4.1(1) on selfadjoint operators applies, i.e., *an operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ is said to be* **selfadjoint** *if it is equal to its adjoint.* This is equivalent to the requirement that $\widehat{A}$ satisfies the following **selfadjointness condition**:

$$\langle \vec{\zeta} \mid \widehat{A}\,\vec{\eta} \rangle = \langle \widehat{A}\vec{\zeta} \mid \vec{\eta} \rangle \quad \forall\, \vec{\zeta},\ \vec{\eta} \in \vec{\mathbb{V}}^N. \tag{13.5}$$

From Eqs. (13.3) and (13.4) we deduce that the sum of two self-adjoint operators is selfadjoint, and the product of two selfadjoint operators is selfadjoint if they commute.[2]

---

[1] Eqs. (13.3) and (13.4) are not generally valid for operators in $\vec{\mathbb{V}}^\infty$. *See* §17.8.

[2] These results are not generally valid for operators in $\vec{\mathbb{V}}^\infty$ because Eqs. (13.3) and (13.4) are not generally valid in $\vec{\mathbb{V}}^\infty$.

P13.1(6) The *quadratic form* $\mathcal{Q}(\widehat{A}, \vec{\eta}\,)$ generated by $\widehat{A}$ on $\vec{\mathbb{V}}^N$ is again defined to be

$$\mathcal{Q}(\widehat{A}, \vec{\eta}\,) := \langle \vec{\eta} \mid \widehat{A}\,\vec{\eta}\,\rangle. \tag{13.6}$$

P13.1(7) The discussion on matrix representation of operators on $\vec{I\!E}^3$ in §8.2.3 remains valid in $\vec{\mathbb{V}}^N$. In particular expressions for matrix representatives in Eq. (8.47) and the matrix representation of operator eigenvalue problem in Eq. (8.50) remain true. In $\vec{\mathbb{V}}^N$ a matrix representation of an operator may contain complex elements and an operator may possess complex eigenvalues.

P13.1(8) Similar to Eq. (12.6) we have

$$\begin{aligned}\langle \vec{\zeta} \mid \widehat{A}\,\vec{\eta}\,\rangle = \frac{1}{2}\Big\{ &\langle \vec{\zeta} + \vec{\eta} \mid \widehat{A}(\vec{\zeta} + \vec{\eta}\,)\rangle - i\langle \vec{\zeta} + i\vec{\eta} \mid \widehat{A}(\vec{\zeta} + i\vec{\eta}\,)\rangle \\ &+(i-1)\big(\langle \vec{\zeta} \mid \widehat{A}\,\vec{\zeta}\,\rangle + \langle \vec{\eta} \mid \widehat{A}\,\vec{\eta}\,\rangle\big)\Big\}.\end{aligned} \tag{13.7}$$

Finally we have the following rather intuitive results.

Theorem 13.1(1)[3]

$$\widehat{A} = \widehat{0} \quad \textit{if and only if} \quad \langle \vec{\xi} \mid \widehat{A}\,\vec{\xi}\,\rangle = 0 \quad \forall\, \vec{\xi} \in \vec{\mathbb{V}}^N. \tag{13.8}$$

Proof Clearly $\widehat{A} = \widehat{0} \Rightarrow \langle \vec{\xi} \mid \widehat{A}\,\vec{\xi}\,\rangle = 0$. To prove the converse we start with $\langle \vec{\xi} \mid \widehat{A}\,\vec{\xi}\,\rangle = 0\ \forall\, \vec{\xi} \in \vec{\mathbb{V}}^N$. This enables us to deduce from Eq. (13.7) that $\langle \vec{\zeta} \mid \widehat{A}\,\vec{\eta}\,\rangle = 0\ \forall\, \vec{\zeta},\ \vec{\eta}$. Equation (12.5) then implies $\widehat{A}\,\vec{\eta} = \widehat{0}\ \forall\, \vec{\eta}$, i.e., $\widehat{A} = \widehat{0}$. **QED**

Corollary 13.1(1)[4] $\widehat{A} = \widehat{B}$ *if and only if*

$$\langle \vec{\zeta} \mid \widehat{A}\,\vec{\zeta}\,\rangle = \langle \vec{\zeta} \mid \widehat{B}\,\vec{\zeta}\,\rangle\ \forall\, \vec{\zeta} \in \vec{\mathbb{V}}^N. \tag{13.9}$$

A useful consequence is that[5]

$$\langle \vec{\zeta} \mid \widehat{A}\,\vec{\zeta}\,\rangle = \langle \vec{\zeta} \mid \vec{\zeta}\,\rangle\ \forall\, \vec{\zeta} \in \vec{\mathbb{V}}^N \quad \Rightarrow \quad \widehat{A} = \widehat{I\!I}. \tag{13.10}$$

The above results tell us that an operator can be characterised by the quadratic form its generates.

---

[3] Roman Vol. 2 p. 536. This result is not valid in a real vector space because it depends on Eq. (13.7). *See* Q13(3) for a counter example. A restricted form of this theorem, i.e., Theorem 9.4.1(1), applies in real vector spaces

[4] This result which is based on Theorem 13.1(1) is not valid in a real vector space. A restricted form of this theorem, i.e., Theorem 9.4.1(2), applies in real vector spaces

[5] *See* Corollary 9.4.1(1) for real vector spaces.

13.2 Subspaces, Projections and Projectors

13.2.1 Subspaces

Following §6.4.4 we shall investigate subspaces, projections and projectors onto subspaces of $\vec{\mathbb{V}}^N$.

Definition 13.2.1(1)

(1) *A subset $\vec{\mathcal{S}}$ of $\vec{\mathbb{V}}^N$ is called a linear subset of $\vec{\mathbb{V}}^N$ if it is closed under addition and scalar multiplication*, i.e.,

$$\vec{\zeta}_1,\ \vec{\zeta}_2\ \in \vec{\mathcal{S}} \quad \Rightarrow \quad c_1\vec{\zeta}_1 + c_2\vec{\zeta}_2\ \in \vec{\mathcal{S}}\ \ \forall\, c_1, c_2\ \in \mathbb{C}. \tag{13.11}$$

(2) *A subset $\vec{\mathcal{S}}$ of $\vec{\mathbb{V}}^N$ is called a subspace of $\vec{\mathbb{V}}^N$ if $\vec{\mathcal{S}}$ is a vector space in its own right.*

(3) *Two subspaces $\vec{\mathcal{S}}_1$ and $\vec{\mathcal{S}}_2$ of $\vec{\mathbb{V}}^N$ are said to have an* **order relation***, to be denoted by $\vec{\mathcal{S}}_1 \subset \vec{\mathcal{S}}_2$, if $\vec{\mathcal{S}}_1$ is a subset of $\vec{\mathcal{S}}_2$.*

The above definitions are introduced with generalisation to infinite-dimensional spaces in mind. It is obvious that a linear subset is a subspace of $\vec{\mathbb{V}}^N$. This is not true for infinite-dimensional Hilbert spaces discussed in §16.2.3. Given a set of n orthonormal vectors $\{\vec{\varepsilon}_j,\ j = 1, 2, \cdots, n \le N\}$ in $\vec{\mathbb{V}}^N$ we can generate an n-dimensional subspace by forming the set of all linear combinations, i.e.,

$$\vec{\mathcal{S}}^n := \Big\{ \sum_{j=1}^{n} c_j\, \vec{\varepsilon}_j, \quad c_j \in \mathbb{C} \Big\}. \tag{13.12}$$

The subspace is also said to be *spanned* by the set of vectors $\vec{\varepsilon}_j$. When $n = N$ we recover the original space $\vec{\mathbb{V}}^N$.[6] We shall often drop the superscript n in $\vec{\mathcal{S}}^n$ for brevity of notation. Some examples are given below.

E13.2.1(1) For the space $\vec{\mathbb{C}}^N$ we can construct n-dimensional subspaces using the basis vectors $\vec{e}_j^{\,c}$ in Eq. (12.14).

[6] An N-dimensional subspace is simply $\vec{\mathbb{V}}^N$ itself.

(1) One-dimensional subspaces An example is the subset consisting vectors of the form

$$\vec{\zeta} := \zeta_1 \vec{e}_1^{\,c} = \begin{pmatrix} \zeta_1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \quad \forall \zeta_1 \in \mathbb{C}. \tag{13.13}$$

(2) Two-dimensional subspaces An example is the subset consisting vectors of the form

$$\vec{\eta} := \eta_1 \vec{e}_1^{\,c} + \eta_2 \vec{e}_2^{\,c} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \quad \forall \eta_1, \eta_2 \in \mathbb{C}. \tag{13.14}$$

E13.2.1(2) For the space $\vec{C}(\Lambda)$ an n-dimensional subspace can be spanned by n orthonormal vectors, e.g., the set in Eq. (12.30).

The notion of orthogonal subspaces and orthogonal complements of subspaces given in Definition 6.4.4 (2) for $\vec{I\!E}^3$ carry over in $\vec{\mathbb{V}}^N$, e.g., the orthogonal complement $\vec{\mathcal{S}}^\perp$ of $\vec{\mathcal{S}}$ is again a subspace and every vector $\vec{\zeta}$ in $\vec{\mathbb{V}}^N$ can be decomposed uniquely as a sum of a vector lying in $\vec{\mathcal{S}}$ and another one lying in $\vec{\mathcal{S}}^\perp$, i.e.,

$$\vec{\zeta} = \vec{\zeta}_{\vec{\mathcal{S}}} + \vec{\zeta}_{\vec{\mathcal{S}}^\perp} \quad \text{where} \quad \vec{\zeta}_{\vec{\mathcal{S}}} \in \vec{\mathcal{S}}, \ \ \vec{\zeta}_{\vec{\mathcal{S}}^\perp} \in \vec{\mathcal{S}}^\perp. \tag{13.15}$$

Following Eqs. (9.39) and (9.40) we can use Eq. (13.15) to define two operators $\widehat{P}_{\vec{\mathcal{S}}}$ and $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ associated with $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^\perp$ by

$$\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\zeta} := \vec{\zeta}_{\vec{\mathcal{S}}} \quad \text{and} \quad \widehat{P}_{\vec{\mathcal{S}}^\perp}\vec{\zeta} := \vec{\zeta}_{\vec{\mathcal{S}}^\perp}. \tag{13.16}$$

13.2.2 Projections and Projectors

Let $\vec{\mathcal{S}}$ be a subspace of $\vec{\mathbb{V}}^N$ and let $\{\vec{\varepsilon}_j, \ j = 1, 2, \cdots, n\}$ be an orthonormal basis for $\vec{\mathcal{S}}$.[7] Let $\vec{\zeta}$ be an arbitrary vector in $\vec{\mathbb{V}}^N$ and let

[7] Basis vectors in $\vec{\mathcal{S}}$ may be different from the basis vectors chosen for $\vec{\mathbb{V}}^N$.

$\vec{\varepsilon}$ be a unit vector in $\vec{\mathbb{W}}^N$. Following Definitions 9.3.1(1) and 9.3.2(1) for $\vec{I\!E}^3$ we introduce the following definitions.

Definition 13.2.2(1)

(1) *The vector* $\vec{\zeta}_{\vec{\varepsilon}} := \langle \vec{\varepsilon} \mid \vec{\zeta} \rangle \vec{\varepsilon}$ *is the projection of* $\vec{\zeta}$ *onto* $\vec{\varepsilon}$*, and the operator* $\widehat{P}_{\vec{\varepsilon}}$ *defined by*

$$\widehat{P}_{\vec{\varepsilon}}\, \vec{\zeta} := \vec{\zeta}_{\vec{\varepsilon}} = \langle \vec{\varepsilon} \mid \vec{\zeta} \rangle\, \vec{\varepsilon} \quad \forall\, \vec{\zeta} \in \vec{\mathbb{W}}^N \tag{13.17}$$

is called the projector onto $\vec{\varepsilon}$*.*

(2) *The vector* $\vec{\zeta}_{\vec{\mathcal{S}}}$ *defined by*

$$\vec{\zeta}_{\vec{\mathcal{S}}} = \sum_{j=1}^{n} \langle \vec{\varepsilon}_j \mid \vec{\zeta} \rangle \vec{\varepsilon}_j \tag{13.18}$$

is the projection of $\vec{\zeta}$ *onto the subspace* $\vec{\mathcal{S}}$*, and the operator* $\widehat{P}_{\vec{\mathcal{S}}}$ *defined by*

$$\widehat{P}_{\vec{\mathcal{S}}}\vec{\zeta} := \vec{\zeta}_{\vec{\mathcal{S}}} = \sum_{j=1}^{n} \langle \vec{\varepsilon}_j \mid \vec{\zeta} \rangle\, \vec{\varepsilon}_j \quad \forall\, \vec{\zeta} \in \vec{\mathbb{W}}^N \tag{13.19}$$

is called the projector onto the subspace $\vec{\mathcal{S}}$*.*

(3) *The dimension of a projector is equal to the dimension of the subspace onto which it projects.*

Definition 13.2.2(2) *Let* $\vec{\mathcal{S}}_1, \vec{\mathcal{S}}_2$ *be two subspaces of* $\vec{\mathbb{W}}^N$.[8]

(1) *The two projectors* $\widehat{P}_{\vec{\mathcal{S}}_1}$ *and* $\widehat{P}_{\vec{\mathcal{S}}_2}$ *onto* $\vec{\mathcal{S}}_1$ *and* $\vec{\mathcal{S}}_2$*, respectively are said to be orthogonal, but not necessarily complementary, if the subspaces* $\vec{\mathcal{S}}_1$ *and* $\vec{\mathcal{S}}_2$ *onto which they project are orthogonal.*

(2) *The two projectors* $\widehat{P}_{\vec{\mathcal{S}}_1}$ *and* $\widehat{P}_{\vec{\mathcal{S}}_2}$ *are said to have an* **order relation***, denoted by* $\widehat{P}_{\vec{\mathcal{S}}_1} \leq \widehat{P}_{\vec{\mathcal{S}}_2}$*, if their associated subspaces* $\vec{\mathcal{S}}_1$ *and* $\vec{\mathcal{S}}_2$ *are related by* $\vec{\mathcal{S}}_1 \subset \vec{\mathcal{S}}_2$ *and vice versa, i.e., we have*

$$\vec{\mathcal{S}}_1 \subset \vec{\mathcal{S}}_2 \quad \Leftrightarrow \quad \widehat{P}_{\vec{\mathcal{S}}_1} \leq \widehat{P}_{\vec{\mathcal{S}}_2}. \tag{13.20}$$

[8] The two subspaces can be of different dimensions. We could have used the notation $\vec{\mathcal{S}}^{n_1}$ and $\vec{\mathcal{S}}^{n_2}$.

(3) *A* **complete orthogonal family of projector** *are as defined by* Definition 9.3.2(3).[9]

(4) *The set of subspaces* $\vec{\mathcal{S}}_\ell$ *corresponding to complete orthogonal family of projectors* $\widehat{P}_{\vec{\mathcal{S}}_\ell}$ *is called a* **complete orthogonal family of subspaces**.

The following comments help to clarify the above definitions:

C13.2.2(1) In Dirac notation the projector $\widehat{P}_{\vec{\varepsilon}}$ is written as $\widehat{P}_{\vec{\varepsilon}} = |\vec{\varepsilon}\,\rangle\langle\vec{\varepsilon}\,|$. The Dirac notation can be extended to represent linear operators of the form $\big|\vec{\zeta}\,\big\rangle\big\langle\vec{\eta}\,\big|$ which is defined by[10]

$$\Big(\big|\vec{\zeta}\,\big\rangle\big\langle\vec{\eta}\,\big|\Big)\,\vec{\xi} = \langle\vec{\eta}\mid\vec{\xi}\,\rangle\,\vec{\zeta} \quad \forall\vec{\xi}\in\vec{\mathbb{V}}^N. \tag{13.21}$$

The adjoint of the operator is $\big|\vec{\eta}\,\big\rangle\big\langle\vec{\zeta}\,\big|$, i.e.,

$$\Big(\big|\vec{\zeta}\,\big\rangle\big\langle\vec{\eta}\,\big|\Big)^\dagger = \big|\vec{\eta}\,\big\rangle\big\langle\vec{\zeta}\,\big|. \tag{13.22}$$

The product of two operators of the form given by Eq. (13.21) is

$$\Big(\big|\vec{\zeta}_1\,\big\rangle\big\langle\vec{\eta}_1\,\big|\Big)\Big(\big|\vec{\zeta}_2\,\big\rangle\big\langle\vec{\eta}_2\,\big|\Big) = \langle\vec{\eta}_1\mid\vec{\zeta}_2\,\rangle\,\big|\vec{\zeta}_1\,\big\rangle\big\langle\vec{\eta}_2\,\big|. \tag{13.23}$$

C13.2.2(2) Projectors are idempotent and they satisfy the selfadjointness condition in Eq. (13.5), i.e., for all $\vec{\zeta},\ \vec{\eta}\in\vec{\mathbb{V}}^N$ we have

$$\widehat{P}^2_{\vec{\mathcal{S}}} = \widehat{P}_{\vec{\mathcal{S}}} \quad \text{and} \quad \langle\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\zeta}\mid\vec{\eta}\,\rangle = \langle\vec{\zeta}\mid\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\eta}\,\rangle. \tag{13.24}$$

It follows that projectors are selfadjoint.

C13.2.2(3) The order relation in Eq. (13.20) is equivalent to any one of the following two statements[11]:

$$\widehat{P}_{\vec{\mathcal{S}}_1}\widehat{P}_{\vec{\mathcal{S}}_2} = \widehat{P}_{\vec{\mathcal{S}}_1} = \widehat{P}_{\vec{\mathcal{S}}_2}\widehat{P}_{\vec{\mathcal{S}}_1}. \tag{13.25}$$

$$||\widehat{P}_{\vec{\mathcal{S}}_1}\vec{\zeta}\,|| \le ||\widehat{P}_{\vec{\mathcal{S}}_2}\vec{\zeta}\,|| \quad \forall\,\vec{\zeta}\in\vec{\mathbb{V}}^N. \tag{13.26}$$

---

[9] The set of projectors $\widehat{P}_{\vec{\varepsilon}_\ell}$ corresponding to an orthonormal basis $\{\vec{\varepsilon}_\ell,\ \ell = 1, 2, \cdots, N\}$ of $\vec{\mathbb{V}}^N$ is a complete orthogonal family of projectors.

[10] Zettili pp. 89–90.

[11] Prugovečki p. 202. Roman Vol. 2 pp. 538, 569. Since $\langle\vec{\zeta}\mid\widehat{P}_{\vec{\mathcal{S}}_1}\vec{\zeta}\,\rangle = ||\widehat{P}_{\vec{\mathcal{S}}_1}\vec{\zeta}\,||$ and $\langle\vec{\zeta}\mid\widehat{P}_{\vec{\mathcal{S}}_2}\vec{\zeta}\,\rangle = ||\widehat{P}_{\vec{\mathcal{S}}_2}\vec{\zeta}\,||$ Eq. (13.26) is the same as

$$\langle\vec{\zeta}\mid\widehat{P}_{\vec{\mathcal{S}}_1}\vec{\zeta}\,\rangle \le \langle\vec{\zeta}\mid\widehat{P}_{\vec{\mathcal{S}}_2}\vec{\zeta}\,\rangle \quad \forall\,\vec{\zeta}\in\vec{\mathbb{V}}^N.$$

Properties P9.3.1(1) to P9.3.1(8) of subspaces and projectors listed in §9.3 remain true. For easy reference we shall summarise some important properties in the following theorem.

Theorem 13.2.2(1)

(1) *For any projector* $\widehat{P}_{\vec{\mathcal{S}}}$ *and any vector* $\vec{\eta}$ *we have*

$$||\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\eta}\,|| \leq ||\vec{\eta}\,|| \quad \textit{and} \quad ||\widehat{P}_{\vec{\mathcal{S}}}\,|| = 1.$$

(2) *An operator is a projector if and only if it is selfadjoint and idempotent.*[12]

(3) *Two projectors* $\widehat{P}_{\vec{\mathcal{S}}_1}$, $\widehat{P}_{\vec{\mathcal{S}}_2}$ *are orthogonal if and only if their product vanishes, i.e., if and only if* $\widehat{P}_{\vec{\mathcal{S}}_1}\widehat{P}_{\vec{\mathcal{S}}_2} = \widehat{0}$.

(4) *The sum of two orthogonal projectors is again a projector. In particular we have*

$$\widehat{P}_{\vec{\mathcal{S}}} + \widehat{P}_{\vec{\mathcal{S}}^\perp} = \widehat{\mathit{I\!I}}. \tag{13.27}$$

(5) *The sum of a complete orthogonal family of projectors corresponding to an orthonormal basis* $\{\vec{\varepsilon}_\ell,\ \ell = 1, 2, \cdots, N\}$ *is equal to the identity operator, i.e.,*

$$\sum_{\ell=1}^{N} \widehat{P}_{\vec{\varepsilon}_\ell} = \widehat{\mathit{I\!I}}, \tag{13.28}$$

and for any given unit vector $\vec{\zeta}$ *we have*[13]

$$\boxed{0 \leq \langle\vec{\zeta} \mid \widehat{P}_{\vec{\varepsilon}_\ell}\vec{\zeta}\,\rangle \leq 1 \quad \text{and} \quad \textstyle\sum_{j=1}^{N} \langle\vec{\zeta} \mid \widehat{P}_{\vec{\varepsilon}_\ell}\vec{\zeta}\,\rangle = 1.} \tag{13.29}$$

(6) *The product of two commuting projectors* $\widehat{P}_{\vec{\mathcal{S}}_1}$ *and* $\widehat{P}_{\vec{\mathcal{S}}_2}$ *is again a projector which projects onto the intersection of the two subspaces* $\vec{\mathcal{S}}_1$ *and* $\vec{\mathcal{S}}_2$.[14]

---

[12] *See* Definition 7.7.5(1) on projection matrices.
[13] As for Eq. (10.4) these are important for generating probability distributions.
[14] Jordan p. 27. Prugovečki pp. 203–204.

13.3 Selfadjoint Operators

13.3.1 Properties

The discussion and results for selfadjoint operators on $\vec{I\!E}^3$ apply to selfadjoint operators on $\vec{\mathbb{V}}^N$:

(1) A linear combination of projectors with real coefficients is selfadjoint.

(2) The properties presented in §9.4.2 remain valid, i.e., Eqs. (9.60) to (9.64) remain true.

(3) The operators $\widehat{A}^\dagger \widehat{A}$ and $\widehat{A}\widehat{A}^\dagger$ are selfadjoint even if $\widehat{A}$ is not.

(4) Selfadjoint operators correspond to selfadjoint matrices.[15]

(5) The discussion on eigenvalues and eigenvectors in §9.4.3 and §9.4.4, including Definition 9.4.4(1) on eigensubspaces and eigenprojectors and Theorem 9.4.4(1), remain valid. *Eigenvalues of selfadjoint operators remain real despite the complex nature of the vector space and their corresponding eigenvectors can be chosen to form an orthonormal basis.*[16]

(6) The notion of degeneracy introduced in §8.2.4 remain valid. In particular a selfadjoint operator is said to be nondegenerate if all its eigenvalues are nondegenerate.

(7) Spectral Theorems 9.4.5(1) and 9.4.5(2) apply here.

Definition 13.3.1(1) *An operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ is said to be* **positive** *if the quadratic form it generates on $\vec{\mathbb{V}}^N$ is real-valued and non-negative, i.e.,*[17]

$$\langle \vec{\eta} \mid \widehat{A}\vec{\eta} \rangle \in I\!R \quad \text{and} \quad \langle \vec{\eta} \mid \widehat{A}\vec{\eta} \rangle \geq 0 \ \ \forall \vec{\eta} \in \vec{\mathbb{V}}^N. \tag{13.30}$$

A related property is stated in following theorem.

---

[15] *See* §13.5 for matrix representation for operators in $\vec{\mathbb{V}}^N$.

[16] Since this is the case for selfadjoint matrices.

[17] Fano p. 88. Definition 9.4.4(2) on positive operators applies to a real vector space like $\vec{I\!E}^3$.

Theorem 13.3.1(1)[18] *An operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ is selfadjoint if and only if the quadratic form generated by the operator on $\vec{\mathbb{V}}^N$ is real-valued, i.e.,*

$$\langle \vec{\eta} \mid \widehat{A}\, \vec{\eta} \rangle \in I\!R \quad \forall\, \vec{\eta} \in \vec{\mathbb{V}}^N. \tag{13.31}$$

A positive operator on $\vec{\mathbb{V}}^N$ is selfadjoint on account of this theorem. A positive operator has no negative eigenvalues.[19]

13.3.2 Spectral Theorem

In view of the importance of the Spectral Theorems 9.4.5(1) and 9.4.5(2) we shall restate them for operators on $\vec{\mathbb{V}}^N$.

Theorem 13.3.2(1) *A selfadjoint operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ can be expressed as a linear combination of a complete orthogonal family of eigenprojectors $\widehat{P}_{\vec{\varepsilon}_\ell}$ generated by a complete orthonormal set of eigenvectors $\vec{\varepsilon}_\ell$, $\ell = 1, 2, \cdots, N$ of $\widehat{A}$, with the corresponding eigenvalues a_ℓ as coefficients, i.e.,*

$$\boxed{\widehat{A} = \sum_{\ell=1}^{N} a_\ell\, \widehat{P}_{\vec{\varepsilon}_\ell}} \tag{13.32}$$

The above decomposition shows that $\widehat{A}$ commutes with all its eigenprojectors. The identity operator can be similarly decomposed, i.e.,

$$\widehat{I\!I} = \sum_{\ell=1}^{N} \widehat{P}_{\vec{\varepsilon}_\ell}, \tag{13.33}$$

In Dirac notation we have

$$\widehat{A} = \sum_{\ell=1}^{N} a_\ell\, |\vec{\varepsilon}_\ell\rangle\langle\vec{\varepsilon}_\ell| \quad \text{and} \quad \widehat{I\!I} = \sum_{\ell=1}^{N} |\vec{\varepsilon}_\ell\rangle\langle\vec{\varepsilon}_\ell|. \tag{13.34}$$

[18] Fano pp. 73, 88. Roman Vol. 2 pp. 532–537. This theorem is not valid in a real vector space which has a real scalar product. The quadratic form generated by an orthogonal operator on $\vec{E}^3$ is real-valued, but the operator is not selfadjoint. This theorem is not valid in complex infinite-dimensional spaces either (*see* comments after Definition 19.1(2).

[19] Fano pp. 88–89.

The decomposition of the identity reflects the completeness of the set of eigenvectors and the set of eigenprojectors. This also leads to the following expression for the scalar product.

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle = \sum_{\ell=1}^{N} \langle \vec{\zeta} \mid \vec{\varepsilon}_\ell \rangle \langle \vec{\varepsilon}_\ell \mid \vec{\eta} \rangle. \tag{13.35}$$

As pointed out in the discussion of Theorem 9.4.5(2) some of the eigenvalues may be degenerate. Let us denote the distinct eigenvalues by a_m where the subscript m goes from 1 to the total number of different eigenvalues M. For a degenerate operator we have $M < N$. We can restate the above theorem in terms of projectors $\widehat{P}^{\hat{A}}(a_m)$ corresponding to distinct eigenvalues a_m as in Theorem 9.4.5(2).

Theorem 13.3.2(2) *A selfadjoint operator $\widehat{A}$ is expressible as a linear combination of its complete orthogonal family of eigenprojectors $\widehat{P}^{\hat{A}}(a_m)$ corresponding distinct eigenvalues a_m, i.e.,*

$$\boxed{\widehat{A} = \sum_{m=1}^{M} a_m \, \widehat{P}^{\hat{A}}(a_m).} \tag{13.36}$$

The above expression is known as the *spectral decomposition* of $\widehat{A}$. The corresponding spectral decomposition of the identity in Eq. (9.80) remains valid.

13.3.3 Functions of a Selfadjoint Operator

Spectral Theorem 13.3.2(2) enables us to retain Definition 9.4.6(1) on functions of a selfadjoint operator $\widehat{A}$.

Definition 13.3.3(1) *Given a function $f(\tau)$ on $\mathbb{R}$ we define a corresponding function of a selfadjoint operator $\widehat{A}$ by*

$$f(\widehat{A}) := \sum_{\ell=1}^{N} f(a_\ell) \, \widehat{P}_{\vec{\varepsilon}_\ell}. \tag{13.37}$$

if $\widehat{A}$ is expressed in terms of Eq. (13.32)*, or*

$$f(\widehat{A}) := \sum_{m=1}^{M} f(a_m) \, \widehat{P}^{\hat{A}}(a_m), \tag{13.38}$$

if $\widehat{A}$ is expressed in terms of Eq. (13.36)*.*

The resulting operators for different functions commute. The function $f(\tau)$ can be real or complex. The operator $f(\widehat{A})$ is selfadjoint if $f(\tau)$ is real-valued. Examples in Eq. (9.85) for the square of $\widehat{A}$ and in Eq. (9.87) for the square root of a positive $\widehat{A}$ apply here. The function $f_{uc}(\tau) = 1$ for all $\tau \in I\!R$, to be referred to as the *unit constant function*,[20] defines an operator function

$$f_{uc}(\widehat{A}) := \sum_{m=1}^{M} \widehat{P}^{\hat{A}}(a_m) = \widehat{I\!I}, \tag{13.39}$$

which is the identity operator.

Another simple but important example of real-valued functions of $\widehat{A}$ is the characteristic function $\chi_{\Lambda}(\tau)$ introduced in Eq. (4.3). Consider the special case in which Λ is an interval $(-\infty, a\,]$ where $a \in I\!R$. The operator $\chi_{(-\infty, a\,]}(\widehat{A})$ possesses the following properties:

P13.3.3(1) It is selfadjoint and idempotent, i.e., it is a projector.

P13.3.3(2) It is a sum of eigenprojectors of $\widehat{A}$ corresponding to eigenvalues less than or equal to a, i.e.,

$$\chi_{(-\infty,a]}(\widehat{A}) := \sum_{m=1}^{M} \chi_{(-\infty,a]}(a_m)\widehat{P}^{\hat{A}}(a_m) \tag{13.40}$$

$$= \sum_{m'} \widehat{P}^{\hat{A}}(a_{m'}), \tag{13.41}$$

where the sum in Eq. (13.41) is over all $a_{m'} \leq a$, since $\chi_{(-\infty,a]}(a_m) = 0$ for all $a_m > a$. For example, we have

$$\chi_{(-\infty,\, a=a_1]}(\widehat{A}) := \widehat{P}^{\hat{A}}(a_1), \tag{13.42}$$

$$\chi_{(-\infty,\, a=a_2]}(\widehat{A}) := \widehat{P}^{\hat{A}}(a_1) + \widehat{P}^{\hat{A}}(a_2). \tag{13.43}$$

An individual eigenprojector can be regarded as a function of $\widehat{A}$, i.e., $\widehat{P}^{\hat{A}}(a_m)$ can be identify with $\chi_{\Lambda}(\widehat{A})$ for an interval Λ which contains only a single eigenvalue a_m. Explicitly we have

$$\chi_{(a_m-0,\, a_m]}(\widehat{A}) := \widehat{P}^{\hat{A}}(a_m), \tag{13.44}$$

[20] This can be compared with the unit step function g_{us} in Eq. (4.7).

since the interval $(a_m - \delta,\, a_m]$ for sufficiently small δ contains a_m but no other eigenvalues.

P13.3.3(3) The properties discussed above can be generalised to $\vec{\mathbb{V}}^{\infty}$.

P13.3.3(4) Complex-valued functions of a selfadjoint operator are not selfadjoint. A complex-valued function $f(\tau)$ and its complex conjugate $f^*(\tau)$ define two operators which are the adjoint of each other, i.e.,

$$f^*(\widehat{A}) = \big(f(\widehat{A})\big)^{\dagger}. \tag{13.45}$$

An important example is the operator defined by a **complex exponential function** $f_e(\tau) = \exp(i\tau)$. We have

$$e^{i\hat{A}} := \sum_{\ell=1}^{N} e^{ia_\ell}\, \widehat{P}_{\vec{\varepsilon}_\ell}, \tag{13.46}$$

or equivalently by

$$e^{i\hat{A}} := \sum_{m=1}^{M} e^{ia_m}\, \widehat{P}^{\hat{A}}(a_m). \tag{13.47}$$

The complex conjugate $f_e^*(\tau) = \exp(-i\tau)$ defines the operator

$$e^{-i\hat{A}} := \sum_{\ell=1}^{N} e^{-ia_\ell}\, \widehat{P}_{\vec{\varepsilon}_\ell}, \tag{13.48}$$

or equivalently

$$e^{-i\hat{A}} := \sum_{m=1}^{M} e^{-ia_m}\, \widehat{P}^{\hat{A}}(a_m). \tag{13.49}$$

The operators in Eqs. (13.46) and (13.48) are adjoints of each other. They are also the inverse of each other, i.e., we have[21]

$$\left(e^{i\hat{A}}\right)^{\dagger} = e^{-i\hat{A}} \quad \text{and} \quad \left(e^{i\hat{A}}\right)^{-1} = e^{-i\hat{A}}. \tag{13.50}$$

In other words, we have

$$\left(e^{i\hat{A}}\right)\left(e^{i\hat{A}}\right)^{\dagger} = e^{i\hat{A}}e^{-i\hat{A}} = \sum_{m=1}^{M} \widehat{P}^{\hat{A}}(a_m) = \widehat{\mathit{II}}. \tag{13.51}$$

This is a very important property. We shall return to study these operators in more detail in Theorem 13.4.3(1).

[21] Fano pp. 91–92.

13.3.4 Commuting Selfadjoint Operators

A selfadjoint operator commutes with its eigenprojectors and the eigenprojectors of a selfadjoint operator commute with each other. Commutation relations play an important role in the study of selfadjoint operators. Let $\widehat{A}$ and $\widehat{B}$ be two selfadjoint operators on $\vec{\mathbb{V}}^N$ with the following spectral decompositions[22]:

$$\widehat{A} = \sum_{m=1}^{M} a_m \, \widehat{P}^{\hat{A}}(a_m) \quad \text{and} \quad \widehat{B} = \sum_{m'=1}^{M'} b_{m'} \, \widehat{P}^{\hat{B}}(b_{m'}). \tag{13.52}$$

The commuting property of $\widehat{A}$ and $\widehat{B}$ are related to the commuting property of their eigenprojectors as stated below.

Theorem 13.3.4(1)[23] *Two selfadjoint operators $\widehat{A}$ and $\widehat{B}$ commute if and only if their eigenprojectors $\widehat{P}^{\hat{A}}(a_m)$ and $\widehat{P}^{\hat{B}}(b_{m'})$ in* Eq. (13.52) *commute for all m and m′.*

This theorem tells us that if $\widehat{A}$ and $\widehat{B}$ commute then:

(1) The products $\widehat{P}^{\hat{A}}(a_m)\widehat{P}^{\hat{B}}(b_{m'})$ are projectors by Theorem 13.2.2(1). The order in the product is unimportant since $\widehat{P}^{\hat{A}}(a_m)\widehat{P}^{\hat{B}}(b_{m'}) = \widehat{P}^{\hat{B}}(b_{m'})\widehat{P}^{\hat{A}}(a_m)$. We have a decomposition of the identity in terms of these projectors, i.e.,

$$\sum_{m,m'=1}^{M,M'} \widehat{P}^{\hat{A}}(a_m)\widehat{P}^{\hat{B}}(b_{m'}) = \widehat{\mathit{I}\!\mathit{I}}. \tag{13.53}$$

(2) The projection $\vec{\eta}_{mm'} = \left(\widehat{P}^{\hat{A}}(a_m)\widehat{P}^{\hat{B}}(b_{m'})\right)\vec{\eta}$ of any vector $\vec{\eta}$ is an eigenvector of both $\widehat{A}$ and $\widehat{B}$, i.e., we have[24]

$$\widehat{A}\vec{\eta}_{mm'} = a_m \, \vec{\eta}_{mm'} \quad \text{and} \quad \widehat{B}\vec{\eta}_{mm'} = b_{m'} \, \vec{\eta}_{mm'}. \tag{13.54}$$

Further properties are seen in the theorems below.

Theorem 13.3.4(2) *Let $\widehat{A}$ be a nondegenerate selfadjoint operator. If another selfadjoint operator $\widehat{B}$ commutes with $\widehat{A}$ then:*

[22] The notation implies that a_m are all different and $b_{m'}$ are all different.
[23] Jordan pp. 53–54, Isham p. 189 and p. 199. The theorem also applies if the eigenprojectors in Eq. (13.32) are used.
[24] Isham p. 99. Jordan pp. 53, 55.

(1) *Each eigenvector of* $\widehat{A}$ *is an eigenvector of* $\widehat{B}$.[25]

(2) $\widehat{B}$ *is a function of* $\widehat{A}$, *i.e., there is a real-valued function* f *on* $\mathbb{R}$ *such that* $\widehat{B} = f(\widehat{A})$.[26]

Proof To prove the first statement let $\vec{\varphi}_\ell$ be an eigenvector of $\widehat{A}$ corresponding to the eigenvalue a_ℓ, i.e., $\widehat{A}\vec{\varphi}_\ell = a_\ell\vec{\varphi}_\ell$. Then

$$\widehat{A}\big(\widehat{B}\,\vec{\varphi}_\ell\big) = \big(\widehat{A}\widehat{B}\,\big)\vec{\varphi}_\ell = \big(\widehat{B}\widehat{A}\,\big)\vec{\varphi}_\ell = a_\ell\big(\widehat{B}\vec{\varphi}_\ell\big). \tag{13.55}$$

This means that $\widehat{B}\vec{\varphi}_\ell$ is an eigenvector of $\widehat{A}$ corresponding to the eigenvalue a_ℓ. Since the eigenvalue is nondegenerate we must have $\widehat{B}\vec{\varphi}_\ell = b_\ell\vec{\varphi}_\ell$ for some constant b_ℓ, i.e., $\vec{\varphi}_\ell$ is an eigenvector $\widehat{B}$.

The second statement of the theorem follows from the first statement. To see this we start with the spectral decomposition of $\widehat{A}$ given by Eq. (13.32) with a_ℓ all being different. Since $\widehat{B}$ commutes with $\widehat{A}$ each eigenvector $\vec{\varepsilon}_\ell$ of $\widehat{A}$ is an eigenvector of $\widehat{B}$ by (1) of the theorem, i.e., we have

$$\widehat{B}\vec{\varepsilon}_\ell = b_\ell\vec{\varepsilon}_\ell, \tag{13.56}$$

where b_ℓ is the corresponding eigenvalue of $\widehat{B}$. It follows that $\vec{\varepsilon}_\ell$ form a complete orthonormal set of eigenvectors of $\widehat{B}$. Applying the spectral theorem to $\widehat{B}$ we get

$$\widehat{B} = \sum_{\ell=1}^{N} b_\ell\,\widehat{P}_{\vec{\varepsilon}_\ell}. \tag{13.57}$$

Let $f(\tau)$ be any real-valued function on $\mathbb{R}$ which maps a_ℓ to b_ℓ, i.e., $f(a_\ell) = b_\ell$.[27] Then we have, by Eq. (13.37),

$$f(\widehat{A}\,) = \sum_{\ell=1}^{N} f(a_\ell)\,\widehat{P}_{\vec{\varepsilon}_\ell} = \sum_{\ell=1}^{N} b_\ell\,\widehat{P}_{\vec{\varepsilon}_\ell} = \widehat{B}. \tag{13.58}$$

QED

[25] Dicke and Wittke p. 96.

[26] Jordan pp. 56–58. This is an example of Theorem 20.6(1) which applies to a complete commuting set of selfadjoint operators. As commented after Definition 20.6(2) a nondegenerate selfadjoint operator constitutes such a complete set in $\vec{\mathbb{V}}^N$.

[27] One can choose a polynomial in τ for the function. The values of $f(\tau)$ for $\tau \neq a_\ell$ do not matter.

The theorem does not apply if $\widehat{A}$ is degenerate. As an example consider Eq. (13.38) where the projector $\widehat{P}_{\vec{\varepsilon}_1}$ commutes with projector $\widehat{P}_{\vec{\varepsilon}_2}$. But $\widehat{P}_{\vec{\varepsilon}_2}$ is not a function of $\widehat{P}_{\vec{\varepsilon}_1}$, since the eigenvalue 0 of $\widehat{P}_{\vec{\varepsilon}_1}$ is degenerate.[28]

The following theorem applies, irrespective of whether the eigenvalues of the operators are degenerate or not.

Theorem 13.3.4(3) *Two selfadjoint operators $\widehat{A}$ and $\widehat{B}$ commute if and only if they share a complete orthonormal set of eigenvectors.*[29]

Theorem 13.3.4(4)[30]

(1) *If two selfadjoint operators $\widehat{A}$ and $\widehat{B}$ commute then they are functions of a third selfadjoint operator, i.e., there exist a selfadjoint operator $\widehat{C}$ and two real-valued functions $f(\tau), g(\tau)$ on IR such that*

$$\widehat{A} = f(\widehat{C}), \quad \widehat{B} = g(\widehat{C}). \tag{13.59}$$

(2) *Let $\{\widehat{A}_j,\ j = 1, 2, \cdots, n\}$ be a finite set of mutually commuting selfadjoint operators. Then $\widehat{A}_j$ are functions of a selfadjoint operator, i.e., there exists a selfadjoint operator $\widehat{C}$ such that*

$$\widehat{A}_j = f_j(\widehat{C}) \quad \forall\, j = 1, 2, \cdots, n. \tag{13.60}$$

As an example consider two eigenprojectors $\widehat{P}_{\vec{\varepsilon}_1}$ and $\widehat{P}_{\vec{\varepsilon}_2}$ of $\widehat{A}$. As already pointed out earlier they are not functions of each other. Instead they both are functions of $\widehat{A}$, i.e., they are characteristic functions of $\widehat{A}$ in accordance with Eq. (13.44). This gives us a two-way relation between a selfadjoint operator and its eigenprojectors:

> *a selfadjoint operator is a linear combination of its eigenprojectors and conversely these eigenprojectors are functions of the selfadjont operator.*

[28] The eigenvalue 0 of a one-dimensional projector is degenerate.

[29] Aulette, Fortunato and Parisi p. 66. Dicke and Wittke pp. 97–98. Fano pp. 84–85, Merzbacher p. 215. This theorem does not say that every eigevector of $\widehat{A}$ is an eigenvector of $\widehat{B}$ since $\widehat{A}$ may not be nondegenerate.

[30] Isham p. 98 for a constructive proof. Fano p. 405. Jordan pp. 56–67. Beltrametti and Gasinelli pp. 19–22. Riesz and Nagy pp. 356–358 for a general proof. *See* Theorem 20.6(1) which applies to infinite-dimensional Hilbert spaces.

The discussion in this section can be extended to infinite-dimensional spaces. This is done in §20.6.

We are now in a position to extend Definition 13.3.3(1) to define functions of two commuting selfadjont operators.

Definition 13.3.4(1) *Given a function* $f(\tau_1, \tau_2)$ *of two independent real variables* τ_1 *and* τ_2 *we can define a corresponding function of two commuting selfadjoint operators* $\widehat{A}$ *and* $\widehat{B}$ *in* Eq. (13.52) *by*[31]

$$f(\widehat{A}, \widehat{B}) := \sum_{m,m'}^{M,M'} f(a_m, b_{m'})\, \widehat{P}^{\hat{A}}(a_m) \widehat{P}^{\hat{B}}(b_{m'}). \tag{13.61}$$

The above expression can be extended to define functions of any finite set of commuting selfadjoint operators.

13.4 Unitary Operators

13.4.1 Definition and Spectral Decomposition

The definition of orthogonal operators given by Eq. (9.2) does not apply to $\vec{\mathbb{V}}^N$ since we do not have an obvious concept of rotations in a complex space of N dimensions. But we can introduce operators using Definition 9.2.2(1) in terms of the preservation of the norm of vectors as a generalisation of the notion of rotations in $\vec{\mathbb{V}}^N$. These new operators which are a generalisation of orthogonal operators on $\vec{I\!E}^3$ to $\vec{\mathbb{V}}^N$ are called *unitary operators*.

Definition 13.4.1(1)[32] *An operator* $\widehat{U}$ *on* $\vec{\mathbb{V}}^N$ *is called a unitary operator on* $\vec{\mathbb{V}}^N$ *if it preserves the norm of all vectors in* $\vec{\mathbb{V}}^N$*, i.e., we have for all* $\vec{\zeta} \in \vec{\mathbb{V}}^N$

$$\langle \widehat{U}\vec{\zeta} \mid \widehat{U}\vec{\zeta} \rangle = \langle \vec{\zeta} \mid \vec{\zeta} \rangle \quad \text{or} \quad ||\widehat{U}\vec{\zeta}|| = ||\vec{\zeta}||. \tag{13.62}$$

Theorem 13.4.1(1)[33] *An operator* $\widehat{U}$ *on* $\vec{\mathbb{V}}^N$ *is unitary if and only if* $\widehat{U}$ *is invertible with its inverse equal to its adjoint.*

---

[31] Jordan p. 55.

[32] Halmos pp. 142–143. Fano pp. 74–75. Note that an operator on $\vec{\mathbb{V}}^N$ means the domain of the operator coincides with $\vec{\mathbb{V}}^N$.

[33] Halmos pp. 62, 64, 142–143.

Proof For a unitary operator $\widehat{U}$ we have

$$\langle \widehat{U}\,\vec{\zeta} \mid \widehat{U}\,\vec{\zeta} \rangle = \langle \vec{\zeta} \mid \widehat{U}^{\dagger}\widehat{U}\,\vec{\zeta} \rangle = \langle \vec{\zeta} \mid \vec{\zeta} \rangle. \tag{13.63}$$

It follows from Eq. (13.10) and the comments right after Theorem 8.2.2(1) that

$$\widehat{U}^{\dagger}\widehat{U} = \widehat{I\!I} \;\Rightarrow\; \widehat{U}^{\dagger} = \widehat{U}^{-1}. \tag{13.64}$$

Next if an operator $\widehat{U}$ is invertible with its inverse $\widehat{U}^{-1}$ equal to its adjoint $\widehat{U}^{\dagger}$ then

$$\langle \widehat{U}\,\vec{\zeta} \mid \widehat{U}\,\vec{\zeta} \rangle = \langle \vec{\zeta} \mid \widehat{U}^{\dagger}\widehat{U}\,\vec{\zeta} \rangle = \langle \vec{\zeta} \mid \widehat{U}^{-1}\widehat{U}\,\vec{\zeta} \rangle = \langle \vec{\zeta} \mid \vec{\zeta} \rangle. \tag{13.65}$$

The operator is therefore unitary. **QED**

Theorem 13.4.1(1) tells us that an operator $\widehat{U}$ in $\vec{\mathbb{V}}^N$ is unitary if

$$\widehat{U}^{\dagger}\widehat{U} = \widehat{U}\,\widehat{U}^{\dagger} = \widehat{I\!I} \quad \text{or} \quad \widehat{U}^{\dagger}\widehat{U} = \widehat{I\!I}. \tag{13.66}$$

The following properties are easily established:

P13.4.1(1) Unitary operators are a generalisation of the notion of rotations in $\vec{\mathbb{V}}^N$ and they preserve scalar product of vectors, i.e.,

$$\langle \widehat{U}\,\vec{\zeta} \mid \widehat{U}\,\vec{\eta} \rangle = \langle \vec{\zeta} \mid \vec{\eta} \rangle. \tag{13.67}$$

P13.4.1(2) If $\widehat{U}$ is unitary then its adjoint $\widehat{U}^{\dagger}$ is also unitary.

P13.4.1(3) Unitary operators possess complex eigenvalues of absolute value 1, i.e., their eigenvalues are of the form[34]

$$u = e^{ia}, \quad a \in I\!R. \tag{13.68}$$

The eigenvalues can be real, e.g., $u = 1$, if $a = 0$ and $u = -1$ if $a = \pi$.

P13.4.1(4) Unitary operators are represented by unitary matrices.

Further properties are given in the form of theorems in the remainder of this section.

[34] Fano pp. 86–87.

As with selfadjoint operators the eigenvectors of a unitary operator also form a complete orthonormal set in $\vec{\mathbb{V}}^N$. These eigenvectors generate a complete orthogonal family of eigenprojectors which can effect a spectral decomposition of the unitary operator in a similar way a complete orthogonal family of eigenprojectors can effect a spectral decomposition of a selfadjoint operator. Let us highlight this property in the following spectral theorem.

Theorem 13.4.1(2) *Let $\widehat{U}$ be a unitary operator on $\vec{\mathbb{V}}^N$.*[35]

(1) *$\widehat{U}$ possesses a complete orthonormal set of eigenvectors $\vec{\varepsilon}_\ell$ corresponding to eigenvalues e^{ia_ℓ}, where $a_\ell \in \mathbb{R}$.*

(2) *$\widehat{U}$ is expressible as a linear combination of its complete orthogonal family of eigenprojectors $\widehat{P}_{\vec{\varepsilon}_\ell}$ with the corresponding eigenvalues e^{ia_ℓ} as coefficients, i.e.,* [36]

$$\boxed{\widehat{U} = \sum_{\ell=1}^{N} e^{ia_\ell}\, \widehat{P}_{\vec{\varepsilon}_\ell} = \sum_{\ell=1}^{N} e^{ia_\ell}\, |\vec{\varepsilon}_\ell\rangle\langle\vec{\varepsilon}_\ell|.} \tag{13.69}$$

The above expression is known as a *spectral decomposition* of $\widehat{U}$. Unitary operators preserve the complete orthonormal nature of a set of vectors as shown in the following theorem.

Theorem 13.4.1(3)[37]

(1) *Let $\widehat{U}$ be a unitary operator on $\vec{\mathbb{V}}^N$, and let $\{\vec{\epsilon}_\ell\}$ be an orthonormal basis for $\vec{\mathbb{V}}^N$. Then $\{\vec{\epsilon}\,'_\ell = \widehat{U}\,\vec{\epsilon}_\ell\}$ is also an orthonormal basis for $\vec{\mathbb{V}}^N$.*

(2) *If $\{\epsilon_\ell\}$ is an orthonormal basis for $\vec{\mathbb{V}}^N$ and if $\{\vec{\epsilon}\,'_\ell\}$ is an orthonormal basis for $\vec{\mathbb{V}}^N$ then there is a unique unitary operator $\widehat{U}$ on $\vec{\mathbb{V}}^N$ such that $\{\vec{\epsilon}\,'_\ell = \widehat{U}\,\vec{\epsilon}_\ell\}$.*

We can express the above unitary operator in Dirac notation as[38]

$$\widehat{U} = \sum_{\ell=1}^{N} \left|\vec{\epsilon}\,'_\ell\right\rangle\left\langle\vec{\epsilon}_\ell\right| \tag{13.70}$$

[35] Fano pp. 86, 91. Jordan pp. 39–41. *See* Theorem 13.4.3(1) and Eq. (13.84).
[36] *See* Eq. (13.84).
[37] Prugovečki p. 215. Roman Vol. 2 p. 559.
[38] *See* Eqs. (13.21) and (13.22).

so that

$$\widehat{U}\,\vec{\epsilon}_m = \Big(\sum_{\ell=1}^{N} \big|\,\vec{\epsilon}\,'_\ell\big\rangle\big\langle\vec{\epsilon}_\ell\,\big|\Big)\vec{\epsilon}_m = \vec{\epsilon}\,'_m. \tag{13.71}$$

13.4.2 Unitary Transformations

Definition 13.4.2(1) *Let $\widehat{U}$ be a unitary operator on $\vec{\mathbb{V}}^N$.*

(1) *The vector $\vec{\zeta}\,' := \widehat{U}\,\vec{\zeta}$ is called the unitary transform of $\vec{\zeta}$ generated by the unitary operator $\widehat{U}$.*

(2) *The operator $\widehat{A}' := \widehat{U}\,\widehat{A}\widehat{U}^\dagger$ is called the unitary transform of $\widehat{A}$ generated by the unitary operator $\widehat{U}$.*

As mentioned in relation to Eq. (12.33) the concept of unitary transformations also applies to any one-to-one and onto mapping between two spaces which preserves the scalar product. We also call such a mapping a unitary operator.

Unitary transformations possess the following properties[39]:

P13.4.2(1) A unitary transformation of vectors preserves the scalar product, i.e.,

$$\vec{\zeta}^{\,\prime} := \widehat{U}\,\vec{\zeta},\;\; \vec{\eta}^{\,\prime} := \widehat{U}\,\vec{\eta} \quad\Rightarrow\quad \langle\vec{\zeta}^{\,\prime}\mid\vec{\eta}^{\,\prime}\rangle = \langle\vec{\zeta}\mid\vec{\eta}\rangle. \tag{13.72}$$

P13.4.2(2) A unitary transformation preserves orthonormal bases, i.e., given an orthonormal basis $\{\vec{\varepsilon}_\ell\}$ in $\vec{\mathbb{V}}^N$ their unitary transforms $\{\vec{\varepsilon}\,'_\ell\}$ generated by a unitary operator $\widehat{U}$ is again an orthonormal basis for $\vec{\mathbb{V}}^N$. Conversely if $\{\vec{\varepsilon}_\ell\}$ and $\{\vec{\varepsilon}\,'_\ell\}$ are two orthonormal bases for $\vec{\mathbb{V}}^N$ then there is a unique unitary operator $\widehat{U}$ such that $\vec{\varepsilon}\,'_\ell = \widehat{U}\,\vec{\varepsilon}_\ell$.[40]

P13.4.2(3) The unitary transform of the product $\widehat{A}\widehat{B}$ of two operators is equal to the product of the unitary transforms of the two operators, i.e., $\widehat{A}\widehat{B} = \widehat{A}'\widehat{B}'$ since

$$\widehat{U}\,(\widehat{A}\widehat{B})\widehat{U}^\dagger = \widehat{U}\,\widehat{A}(\widehat{U}^\dagger\widehat{U})\,\widehat{B}\widehat{U}^\dagger = (\widehat{U}\,\widehat{A}\widehat{U}^\dagger)(\widehat{U}\,\widehat{B}\widehat{U}^\dagger). \tag{13.73}$$

[39] These properties remain valid for unitary operators in an infinite-dimensional space. *See* §18.3.

[40] Prugovečki p. 215.

P13.4.2(4) A unitary transformation of operators preserves commutation relations, i.e.,[41]

$$[\,\widehat{A},\,\widehat{B}\,]=\widehat{C}\quad\Leftrightarrow\quad[\,\widehat{A}',\,\widehat{B}']=\widehat{C}'. \tag{13.74}$$

This has many applications as seen in Eqs. (18.69) and (29.25).

P13.4.2(5) A *simultaneous unitary transformation* of vectors and operators preserves scalar product, i.e.,

$$\vec{\zeta}^{\,\prime}:=\widehat{U}\,\vec{\zeta},\quad \vec{\eta}^{\,\prime}:=\widehat{U}\,\vec{\eta},\quad \widehat{A}':=\widehat{U}\,\widehat{A}\widehat{U}^{\,\dagger} \tag{13.75}$$

$$\Rightarrow\quad \langle\vec{\zeta}^{\,\prime}\mid\widehat{A}'\vec{\eta}^{\,\prime}\rangle=\langle\vec{\zeta}\mid\widehat{A}\,\vec{\eta}\,\rangle. \tag{13.76}$$

In particular we have the preservation of the quadratic form, i.e.,

$$\langle\vec{\zeta}^{\,\prime}\mid\widehat{A}'\vec{\zeta}^{\,\prime}\rangle=\langle\vec{\zeta}\mid\widehat{A}\,\vec{\zeta}\,\rangle. \tag{13.77}$$

P13.4.2(6) Let $\widehat{A}$ be a selfadjoint operator on $\vec{\boldsymbol{V}}^N$. Let a_ℓ be the eigenvalues of $\widehat{A}$ corresponding to eigenvectors $\vec{\varepsilon}_\ell$. Then

> *The unitary transform $\widehat{A}'$ of $\widehat{A}$ generated by a unitary operator $\widehat{U}$ possesses the same eigenvalues as $\widehat{A}$ and the eigenvectors $\vec{\varepsilon}'_\ell$ of $\widehat{A}'$ are the unitary transforms of the eigenvectors $\vec{\varepsilon}_\ell$ of $\widehat{A}$.*

These results mean that in addition to the preservation of the quadratic form in Eq. (13.77) a simultaneous unitary transformation preserves the eigenvalue equation, i.e.,

$$\widehat{A}\vec{\varepsilon}_\ell=a_\ell\vec{\varepsilon}_\ell,\quad \widehat{A}'=\widehat{U}\,\widehat{A}\widehat{U}^{\,\dagger},\quad \vec{\varepsilon}\,'_\ell=\widehat{U}\,\vec{\varepsilon}_\ell \tag{13.78}$$

$$\Rightarrow\quad \widehat{A}'\vec{\varepsilon}\,'_\ell=a_\ell\vec{\varepsilon}\,'_\ell. \tag{13.79}$$

13.4.3 Stone's Theorem

Unitary operators are closely related to selfadjoint operators as seen in the following theorem.

[41] The commutator $[\,\widehat{A},\,\widehat{B}\,]$ is again defined by Eq. (8.19).

Theorem 13.4.3(1)[42] *An operator $\widehat{U}$ on $\vec{\mathbb{V}}^N$ is unitary if and only if it can be expressed in term of a selfadjoint operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ in the form*

$$\widehat{U} = e^{i\hat{A}}. \tag{13.80}$$

This theorem tells us that a unitary operator is a complex exponential function of a selfadjoint operator. This result can be established as follows:

(1) We can start with the spectral decomposition of $\widehat{U}$ in Eq. (13.69). Using the eigenprojectors $\widehat{P}_{\vec{\varepsilon}_\ell}$ and the constants a_ℓ in the spectral decomposition we can construct the following selfadjoint operator:

$$\widehat{A} = \sum_{\ell=1}^{N} a_\ell \, \widehat{P}_{\vec{\varepsilon}_\ell}. \tag{13.81}$$

On account of Eq. (13.46) we can regard $\widehat{U}$ as an exponential function of $\widehat{A}$. In other words, a unitary operator $\widehat{U}$ generates a selfadjoint operator $\widehat{A}$ with $\widehat{P}_{\vec{\varepsilon}_\ell}$ identified with the corresponding eigenprojectors of $\widehat{A}$. The two operators are related by Eq. (13.80). The choice of $\widehat{A}$ is not unique, e.g., an addition of a term $2\pi\,\widehat{\mathbb{I}}$ to the above expression for $\widehat{A}$ will lead to the same unitary operator.[43]

(2) Given a selfadjoint operator $\widehat{A}$ in Eq. (13.46) we can construct a complex exponential function by Eq. (13.46), i.e.,

$$\widehat{U} = e^{i\hat{A}} = \sum_{\ell=1}^{N} e^{ia_\ell} \, \widehat{P}_{\vec{\varepsilon}_\ell}. \tag{13.82}$$

This operator is invertible and its inverse is equal to its adjoint, i.e., the resulting operator is unitary.

Suppose not all the values $u_\ell = e^{ia_\ell}$ are different, i.e., there are only $M \leq N$ different eigenvalues and there are n_m mutually orthonormal eigenvectors $\vec{\varphi}_{mj}$, $j = 1, 2, \cdots, n_m$ corresponding to each distinct

[42] Fano pp. 91–92.
[43] Roman Vol. 2 pp. 660–661.

value u_m. Introduce the following projector associated with the eigenvalue $u_m = e^{ia_m}$:

$$\widehat{P}^{\hat{U}}(u_m) = \sum_{j=1}^{n_m} \widehat{P}_{\vec{\varphi}_{mj}}. \tag{13.83}$$

Then the spectral decomposition in Eq. (13.69) can be rewritten as[44]

$$\boxed{\widehat{U} = \sum_{m=1}^{M} u_m \, \widehat{P}^{\hat{U}}(u_m).} \tag{13.84}$$

The following properties can be established:

P13.4.3(1) Let $\widehat{A}$ be a selfadjoint operator and let

$$\widehat{U}(a) := e^{i\hat{A}a} \quad \text{and} \quad \widehat{U}(b) := e^{i\hat{A}b}, \quad a,\ b \in \mathbb{R}. \tag{13.85}$$

Then we have

$$\widehat{U}(a)\widehat{U}(b) = \widehat{U}(a+b). \tag{13.86}$$

In other words, we have[45]

$$e^{i\hat{A}a} e^{i\hat{A}b} = e^{i\hat{A}(a+b)}. \tag{13.88}$$

P13.4.3(2) The usual expansion of an exponential function applies, i.e., we have[46]

$$e^{i\hat{A}} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(i\widehat{A}\right)^n. \tag{13.89}$$

This expansion needs to be treated with care in an infinite-dimensional space, e.g., we have to consider the domain of $\widehat{A}^n$ as $n \to \infty$. We shall return to this in §19.5.

P13.4.3(3) Equations (13.69) and (13.81) shows that $\widehat{U}$ and $\widehat{A}$ share the same eigenvectors $\vec{\varepsilon}_\ell$ and that the eigenvalues of $\widehat{U}$ are exponential function e^{ia_ℓ} of the eigenvalues a_ℓ of $\widehat{A}$, i.e.,[47]

$$\widehat{A}\vec{\varepsilon}_\ell = a_\ell \vec{\varepsilon}_\ell \quad \text{and} \quad \widehat{U}\,\vec{\varepsilon}_\ell = e^{ia_\ell}\vec{\varepsilon}_\ell. \tag{13.90}$$

---

[44] This is analogous to Theorem 9.4.5(2). *See* also Isham p. 111.

[45] Note that

$$e^{i\hat{A}} e^{i\hat{B}} \neq e^{i(\hat{A}+\hat{B})}, \tag{13.87}$$

unless $\widehat{A}$ and $\widehat{B}$ commute. We shall return to discuss this in §21.1.

[46] The symbol $n! = n \times (n-1) \times (n-2) \times \cdots 3 \times 2 \times 1$ stands for "n factorial".

[47] *See* Eq. (13.68).

P13.4.3(4) A selfadjoint operator $\widehat{A}$ can generate a *one-parameter family of unitary operators* by

$$\widehat{U}(t) = e^{i\hat{A}t}, \quad t \in I\!R. \tag{13.91}$$

This family of unitary operators has the following properties:

$$\widehat{U}(0) = \widehat{I\!I}, \text{ the identity operator,} \tag{13.92}$$

$$\widehat{U}(t_1)\widehat{U}(t_2) = \widehat{U}(t_1 + t_2) \ \forall\, t_1,\ t_2 \ \in I\!R, \tag{13.93}$$

and

$$\langle \vec{\zeta} \mid \widehat{U}(t)\vec{\eta} \rangle \tag{13.94}$$

is a continuous function of t for any pair of vectors $\vec{\zeta}$ and $\vec{\eta}$ in $\vec{\mathbb{W}}^N$.

Definition 13.4.3(1)[48] *A family of unitary operators* $\widehat{U}(t),\ t \in I\!R$, *possessing the properties given in* Eqs. (13.92), (13.93) *and* (13.94) *is called a continuous one-parameter group of unitary operators.*

Theorem 13.4.3(2) Stone's Theorem[49] *Let* $\widehat{U}(t)$ *be a continuous one-parameter group of unitary operators. Then there is a unique selfadjoint operator* $\widehat{A}$*, known as the generator of the group, such that*

$$\widehat{U}(t) = e^{-\dot{\imath}\hat{A}t}, \quad \dot{\imath} = \frac{i}{\hbar} \tag{13.95}$$

and

$$\widehat{A}\,\vec{\zeta} = i\hbar \lim_{t\to 0} \frac{\widehat{U}(t) - \widehat{I\!I}}{t}\,\vec{\zeta}, \quad \vec{\zeta} \ \in \vec{\mathbb{W}}^N. \tag{13.96}$$

The notation $\dot{\imath} = i/\hbar$ will be used through out the book.[50] We can rewrite Eq. (13.96) in the notation of a formal differentiation, i.e.,

$$\widehat{A} = i\hbar \left(\frac{d\widehat{U}(t)}{dt} \right)_{t=0}. \tag{13.97}$$

---

[48] Isham p. 113. Roman Vol. 2 p. 661. Such a set of operators possesses the properties of a group, e.g., each element of the family has an inverse $\widehat{U}^{-1}(t) = \widehat{U}(-t)$.

[49] Isham p. 113. Roman Vol. 2 pp. 662–665. Prugovečki pp. 288, 335. Stone (1903–1989) is an American mathematician.

[50] The theorem is often stated in terms of $\widehat{U} = \exp(i\widehat{A})$. We have replaced i by $-\dot{\imath} = (-i/\hbar)$ for applications to quantum evolution in §29.1.2 and §29.2.

We can also formally differentiate $\widehat{U}$ in Eq. (13.95) as

$$i\hbar \frac{d\widehat{U}(t)}{dt} = \widehat{A}\,\widehat{U}(t). \tag{13.98}$$

Let $\vec{\zeta}(0)$ be an arbitrary vector and let $\vec{\zeta}(t) = \widehat{U}(t)\vec{\zeta}(0)$. We have

$$\begin{aligned} i\hbar \frac{d\vec{\zeta}(t)}{dt} &= i\hbar \left(\frac{d\widehat{U}(t)\vec{\zeta}(0)}{dt} \right) = i\hbar \frac{d\widehat{U}(t)}{dt}\,\vec{\zeta}(0) \qquad (13.99) \\ &= \widehat{A}\,\widehat{U}(t)\vec{\zeta}(0) \\ \Rightarrow \qquad i\hbar \frac{d\vec{\zeta}(t)}{dt} &= \widehat{A}\,\vec{\zeta}(t). \qquad (13.100) \end{aligned}$$

This equation bears a striking resemblance to the time-dependent Schrödinger equation shown in Eq. (10.27) for the evolution of quantum waves, with the generator $\widehat{A}$ playing the role of $\widehat{H}$. In §29.1.2 we will discuss how the Stone's theorem is used to establish the time-dependent Schrödinger equation for quantum evolution.

13.5 Matrix Representation of Operators

The notion of matrix representation of operators discussed in §8.2.3 can be extended to $\vec{\mathbb{V}}^N$. Let $\{\vec{\varepsilon}_\ell\}$ be an orthonormal basis in $\vec{\mathbb{V}}^N$. For an arbitrary operator $\widehat{A}$ we have $\widehat{A}\,\vec{\varepsilon}_\ell = \vec{\varepsilon}\,'_\ell$. The output vector $\vec{\varepsilon}\,'_\ell$ can be expressed in basis $\{\vec{\varepsilon}_\ell\}$ as

$$\vec{\varepsilon}\,'_\ell = \sum_{k=1}^{N} M_{\hat{A}k\ell}\,\vec{\varepsilon}_k, \quad M_{\hat{A}k\ell} = \langle \vec{\varepsilon}_k \mid \vec{\varepsilon}\,'_\ell \rangle = \langle \vec{\varepsilon}_k \mid \widehat{A}\,\vec{\varepsilon}_\ell \rangle. \tag{13.101}$$

For any arbitrary input vector $\vec{\zeta}$ we have $\widehat{A}\,\vec{\zeta} = \vec{\zeta}\,'$. Expressing the input vector $\vec{\zeta}$ in basis $\{\vec{\varepsilon}_\ell\}$ as

$$\vec{\zeta} = \sum_{\ell} \zeta_\ell \vec{\varepsilon}_\ell, \quad \zeta_\ell = \langle \vec{\varepsilon}_\ell \mid \vec{\zeta}\,\rangle \tag{13.102}$$

we get

$$\begin{aligned} \widehat{A}\,\vec{\zeta} &= \sum_{\ell=1}^{N} \zeta_\ell\,\widehat{A}\,\vec{\varepsilon}_\ell = \sum_{\ell=1}^{N} \zeta_\ell\,\vec{\varepsilon}\,'_\ell = \sum_{\ell=1}^{N} \zeta_\ell \left(\sum_{k=1}^{N} M_{\hat{A}k\ell}\vec{\varepsilon}_k \right) \\ &= \sum_{k=1}^{N} \left(\sum_{\ell=1}^{N} \zeta_\ell\, M_{\hat{A}k\ell} \right) \vec{\varepsilon}_k. \qquad (13.103) \end{aligned}$$

Next, expressing the output vector $\vec{\zeta}\,'$ in basis $\{\vec{\varepsilon}_\ell\}$ as

$$\vec{\zeta}\,' = \sum_k \zeta'_k \vec{\varepsilon}_k, \quad \zeta'_k = \langle \vec{\varepsilon}_k \mid \vec{\zeta}\,' \rangle \tag{13.104}$$

we get

$$\widehat{A}\vec{\zeta} = \vec{\zeta}\,' = \sum_{k=1}^{N} \zeta'_k \vec{\varepsilon}_k. \tag{13.105}$$

Comparing Eqs. (13.103) and (13.105) we get

$$\sum_{\ell=1}^{N} \zeta_\ell \, M_{\hat{A}k\ell} = \zeta'_k \qquad \text{or} \qquad \sum_{\ell=1}^{N} M_{\hat{A}k\ell} \, \zeta_\ell = \zeta'_k. \tag{13.106}$$

Finally let $\boldsymbol{C}_{\vec{\zeta}}$ be the column vector with elements ζ_ℓ and $\boldsymbol{C}_{\vec{\zeta}'}$ be the column vector with elements ζ'_m, and let $\boldsymbol{M}_{\hat{A}}$ be the matrix with matrix elements

$$M_{\hat{A}k\ell} := \langle \vec{\varepsilon}_k \mid \widehat{A}\,\vec{\varepsilon}_\ell \rangle. \tag{13.107}$$

Then Eqs. (13.105) and (13.106) correspond to the matrix equation

$$\begin{pmatrix} M_{\hat{A}11} & M_{\hat{A}12} & \cdots \\ M_{\hat{A}21} & M_{\hat{A}22} & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} \zeta'_1 \\ \zeta'_2 \\ \cdot \\ \cdot \end{pmatrix}, \tag{13.108}$$

or

$$\boldsymbol{M}_{\hat{A}} \boldsymbol{C}_{\vec{\zeta}} = \boldsymbol{C}_{\vec{\zeta}'}, \tag{13.109}$$

with the matrix $\boldsymbol{M}_{\hat{A}}$ corresponding to the operator $\widehat{A}$, and the column vectors $\boldsymbol{C}_{\vec{\zeta}}$ and $\boldsymbol{C}_{\vec{\zeta}'}$ corresponding to the vectors $\vec{\zeta}$ and $\vec{\zeta}\,'$. It can be easily shown that this correspondence possesses the following properties:

P13.5(1) Preservation of vector equations involving operators:

$$\widehat{A}\,\vec{\zeta} = \vec{\zeta}' \quad \Leftrightarrow \quad \boldsymbol{M}_{\hat{A}} \boldsymbol{C}_{\vec{\zeta}} = \boldsymbol{C}_{\vec{\zeta}'}. \tag{13.110}$$

P13.5(2) Preservation of eigenvalues:

$$\widehat{A}\,\vec{\eta} = a\,\vec{\eta} \quad \Leftrightarrow \quad \boldsymbol{M}_{\hat{A}} \boldsymbol{C}_{\vec{\eta}} = a\,\boldsymbol{C}_{\vec{\eta}}. \tag{13.111}$$

P13.5(3) Preservation of selfadjointness:

$$\widehat{A} = \widehat{A}^\dagger \quad \Leftrightarrow \quad \boldsymbol{M}_{\hat{A}} = \boldsymbol{M}_{\hat{A}}^\dagger. \tag{13.112}$$

P13.5(4) Preservation of addition and multiplication:

$$\widehat{A}_1 + \widehat{A}_2 \;\Leftrightarrow\; \boldsymbol{M}_{\hat{A}_1} + \boldsymbol{M}_{\hat{A}_2}, \tag{13.113}$$

$$\widehat{A}_1 \widehat{A}_2 \;\Leftrightarrow\; \boldsymbol{M}_{\hat{A}_1}\, \boldsymbol{M}_{\hat{A}_2}. \tag{13.114}$$

P13.5(5) Preservation of commutation relations:

$$[\widehat{A}_1, \widehat{A}_2] = \widehat{A}_3 \quad \Leftrightarrow \quad [\boldsymbol{M}_{\hat{A}_1}\,,\, \boldsymbol{M}_{\hat{A}_2}] = \boldsymbol{M}_{\hat{A}_3}. \tag{13.115}$$

An important application is the matrix representation of spin operators, as seen in Chapter 14. This is further discussed in §36.3.6.2. It is often convenient to drop the subscript $\hat{A}$ if doing so does not cause any confusion.

Exercises and Problems

Q13(1) Prove Eqs. (13.3) and (13.4).

Q13(2) Verify Eq. (13.7).

Q13(3) Consider the xy-plane as a two-dimensional real vector space in its own right. In basis $\{\vec{i},\ \vec{j}\,\}$ an arbitrary vector $\vec{v}$ on the xy-plane has the matrix representation

$$\boldsymbol{C}_{\vec{v}} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}. \tag{13.116}$$

In basis $\{\vec{i},\ \vec{j}\,\}$ the operator $\widehat{R}_p(\pi/2)$ which rotates any vector $\vec{u}$ on the xy-plane about the origin by an angle of $\pi/2$ has the following matrix representation in accordance with Eq. (7.136):

$$\boldsymbol{R}_p(\pi/2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{13.117}$$

Show that $\langle \vec{v} \mid \widehat{R}_p(\pi/2)\vec{v}\,\rangle = 0$ and explain why this result does not satisfy Theorem 13.1(1).

Q13(4) Prove that every vector $\vec{\zeta}$ in $\vec{\mathbb{V}}^N$ can be decomposed uniquely as a sum of a vector lying in a given subspace $\vec{\mathcal{S}}$ and another one lying in its orthogonal complement $\vec{\mathcal{S}}^{\perp}$ as shown in Eq. (13.15).

Q13(5) Show that Eq. (13.20) is equivalent to Eq. (13.25) or Eq. (13.26) in defining the order relation of projectors.

Q13(6) Prove Eqs. (13.22) and (13.23).

Q13(7) Prove statement (3) of Theorem 13.2.2(1).

Q13(8) Prove the two expressions in Eq. (13.29).

Q13(9) Prove Theorem 13.3.1(1) on selfadjoint operators.

Q13(10) Prove Spectral Theorems 13.3.2(1) and 13.3.2(2).

Q13(11) Show that the eigenprojectors of a selfadjoint operator commutes with each other and that a selfadjoint operator commutes with its eigenprojectors.

Q13(12) For a selfadjoint operator $\widehat{A}$ with its spectral decomposition given by Eq. (13.36) show that

$$\widehat{A}\,\widehat{P}^{\hat{A}}(a_m) = a_m\,\widehat{P}^{\hat{A}}(a_m). \tag{13.118}$$

Q13(13) Prove Eq. (13.50).

Q13(14) Prove Theorem 13.3.4(1).

Q13(15) Prove Eqs. (13.53) and (13.54).

Q13(16) Prove properties P13.4.1(1) to P13.4.1(4) of a unitary operator listed right after Theorem 13.4.1(1).

Q13(17) Show that the operator in Eq. (13.69) is unitary.

Q13(18) Show that the operator $\widehat{U}$ in Eq. (13.70) is unitary.

Q13(19) Prove Eq. (13.74) on the preservation of commutation relations.

Q13(20) Prove Eq. (13.77) on the preservation of the quadratic form.

Q13(21) [51]Let $\widehat{B}$ a selfadjoint operator and let $\widehat{B}'$ be the unitary transform of $\widehat{B}$ generated by a unitary operator $\widehat{U}$. Show that $\widehat{B}$ and $\widehat{B}'$ possess the same set of eigenvalues and that their corresponding eigenvectors are unitary transforms of each other.[52]

Q13(22) Let $\widehat{U}(t)$ be a continuous one-parameter group of unitary operators. Let $\widehat{A}$ be the generator of $\widehat{U}(t)$ in accordance with Theorem 13.4.3(2) of Stone. Let $\vec{\xi}(0)$ be an eigenvector of $\widehat{A}$ corresponding to a non-denegerate eigenvalue a. Show that

$$\widehat{U}(t)\vec{\xi}(0) = e^{-iat}\,\vec{\xi}(0), \tag{13.119}$$

and that

$$\vec{\xi}(t) = \widehat{U}(t)\vec{\xi}(0), \tag{13.120}$$

is a solution of Eq. (13.100).

[51]This question is on the important property stated in P13.4.2(6).

[52]First show that $\widehat{B}'$ possesses the same eigenvalues as $\widehat{B}$ and that the eigenvectors $\widehat{B}'$ are the corresponding unitary transforms of the eigenvectors of $\widehat{B}$. Then show that $\widehat{B}$ possesses the same eigenvalues as $\widehat{B}'$ and that the eigenvectors $\widehat{B}$ are the corresponding unitary transforms of the eigenvectors of $\widehat{B}'$.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 14

Model Theories Based on Complex Vector Spaces $\vec{\mathbb{V}}^N$

The discussion in Chapter 10 on the generation of probability distributions and the formulation of probabilistic theories can be applied to a complex vector space $\vec{\mathbb{V}}^N$. We shall follow Chapter 10 to establish model theories of spin 1/2 in $\vec{\mathbb{V}}^2$ and of spin 1 in $\vec{\mathbb{V}}^3$.

14.1 Model Theories of Spin

14.1.1 Electron Spin

For the familiar electron spin there are three basic observables, i.e., S_x, S_y and S_z, the spin components along the positive directions of x, y and z axes. These observables can take only two values, $\pm\hbar/2$. In §5.3 we introduce the state α_z^s to correspond to S_z possessing the value $\hbar/2$. This is traditionally called the spin-up state which correspond to S_z aligned along the positive z-direction. The state β_z^s corresponds to S_z possessing the value $-\hbar/2$. This is the spin-down state which corresponds to S_z aligned along the negative z-direction. Following §10.2.2 a probability theory of electron spin should be set up in a two-dimensional complex vector space $\vec{\mathbb{V}}^2$.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

14.1.1.1 On states

Take $\vec{\mathbb{V}}^2$ as the state space.[1] An arbitrary state η^s is described by a unit vector $\vec{\eta} \in \vec{\mathbb{V}}^2$. We then

(1) choose an orthonormal basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ for $\vec{\mathbb{V}}^2$, and
(2) describe the spin-up and spin-down states α_z^s and β_z^s by the two orthonormal basis vectors $\vec{\alpha}_z$ and $\vec{\beta}_z$.

The matrix representations of $\vec{\alpha}_z, \vec{\beta}_z$ in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ are, in accordance with Eqs. (13.102), (13.107) and (13.108) and the related discussion,

$$\boldsymbol{C}_{\vec{\alpha}_z} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\beta}_z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{14.1}$$

An arbitrary unit vector $\vec{\eta}$ is expressible in terms of these basis vectors, i.e., we have

$$\vec{\eta} = c_+ \vec{\alpha}_z + c_- \vec{\beta}_z, \quad c_+, c_- \in \mathbb{C}, \ |c_+|^2 + |c_-|^2 = 1. \tag{14.2}$$

This provides the mathematical description of the superposition principle stated in QMP5.3(4) in §5.3. The vector $\vec{\eta}$ has the following matrix representation in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$:

$$\boldsymbol{C}_{\vec{\eta}} = \begin{pmatrix} c_+ \\ c_- \end{pmatrix}. \tag{14.3}$$

14.1.1.2 On observable S_z

An observable is represented by a selfadjoint operator on the state space $\vec{\mathbb{V}}^2$. The basic observables S_x, S_y and S_z are then described by three selfadjoint operators $\widehat{S}_x$, $\widehat{S}_y$ and $\widehat{S}_z$ on $\vec{\mathbb{V}}^2$. The selfadjoint operator $\widehat{S}_z$ must admit $\vec{\alpha}_z$ and $\vec{\beta}_z$ as its eigenvectors corresponding to eigenvalues $\hbar/2$ and $-\hbar/2$. It follows that $\widehat{S}_z$ has following spectral decomposition:

$$\widehat{S}_z = \frac{1}{2}\hbar \widehat{P}_{\vec{\alpha}_z} - \frac{1}{2}\hbar \widehat{P}_{\vec{\beta}_z} \tag{14.4}$$

$$= \frac{1}{2}\hbar \, |\vec{\alpha}_z\rangle\langle\vec{\alpha}_z| - \frac{1}{2}\hbar \, |\vec{\beta}_z\rangle\langle\vec{\beta}_z|. \tag{14.5}$$

[1] Here we just consider a model theory. A systematic theory of spin is disccused in §36.3.

The matrix representation of the selfadjoint operator $\widehat{S}_z$ in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ is, in accordance with Eq. (13.107),

$$\boldsymbol{M}_{\hat{S}_z} = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{14.6}$$

The matrix $\boldsymbol{M}_{\hat{S}_z}$ admits the column vectors $\boldsymbol{C}_{\vec{\alpha}_z}$ and $\boldsymbol{C}_{\vec{\beta}_z}$ as eigenvectors with eigenvalues $\hbar/2$ and $-\hbar/2$, respectively.

The matrix representations of projectors $\widehat{P}_{\vec{\alpha}_z}$ and $\widehat{P}_{\vec{\beta}_z}$ in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ are[2]

$$\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_z}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{M}_{\hat{P}_{\vec{\beta}_z}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{14.7}$$

The spectral decomposition of $\boldsymbol{M}_{\hat{S}_z}$ takes the form:

$$\boldsymbol{M}_{\hat{S}_z} = \frac{\hbar}{2}\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_z}} - \frac{\hbar}{2}\boldsymbol{M}_{\hat{P}_{\vec{\beta}_z}}. \tag{14.8}$$

14.1.1.3 On probability distributions

Following Eqs. (10.2), (10.3) and (10.7) we can obtain the probability mass function for the probabilities of the measured values of S_z in state η^s by

$$\wp^{S_z}(\eta^s, \hbar/2) = \mathcal{Q}(\widehat{P}_{\vec{\alpha}_z}, \vec{\eta}) = \langle \vec{\eta} \mid \widehat{P}_{\vec{\alpha}_z}\vec{\eta} \rangle \tag{14.9}$$

$$= |\langle \vec{\alpha}_z \mid \vec{\eta} \rangle|^2 = |c_+|^2, \tag{14.10}$$

$$\wp^{S_z}(\eta^s, -\hbar/2) = \mathcal{Q}(\widehat{P}_{\vec{\beta}_z}, \vec{\eta}) = \langle \vec{\eta} \mid \widehat{P}_{\vec{\beta}_z}\vec{\eta} \rangle \tag{14.11}$$

$$= |\langle \vec{\beta}_z \mid \vec{\eta} \rangle|^2 = |c_-|^2. \tag{14.12}$$

All the scalar product expressions can also be calculated by their matrix representations, i.e.,

$$\langle \vec{\eta} \mid \widehat{P}_{\vec{\alpha}_z}\vec{\eta} \rangle = \boldsymbol{C}^{\dagger}_{\vec{\eta}} \cdot (\boldsymbol{M}_{\vec{\alpha}_z}\boldsymbol{C}_{\vec{\eta}}), \tag{14.13}$$

$$\langle \vec{\eta} \mid \widehat{P}_{\vec{\beta}_z}\vec{\eta} \rangle = \boldsymbol{C}^{\dagger}_{\vec{\eta}} \cdot (\boldsymbol{M}_{\vec{\beta}_z}\boldsymbol{C}_{\vec{\eta}}), \tag{14.14}$$

$$\langle \vec{\alpha}_z \mid \vec{\eta} \rangle = \boldsymbol{C}^{\dagger}_{\vec{\alpha}_z} \cdot \boldsymbol{C}_{\vec{\eta}}, \tag{14.15}$$

$$\langle \vec{\beta}_z \mid \vec{\eta} \rangle = \boldsymbol{C}^{\dagger}_{\vec{\beta}_z} \cdot \boldsymbol{C}_{\vec{\eta}}. \tag{14.16}$$

[2] These are the projection matrices in Eqs. (7.182) and (7.183).

It must be pointed out that the two unit vectors $\vec{\alpha}_z$ and $\vec{\beta}_z$ are abstract mathematical description of the z-component spin-up state and spin-down states and they cannot not be identified with the geometric directions of the spin, i.e., $\vec{\alpha}_z$ and $\vec{\beta}_z$ cannot be identified with the unit vectors $\vec{k}$ and $-\vec{k}$ along the positive and the negative directions of the z-axis.

14.1.1.4 On observable S_x

There must be a state α_x^s for the spin aligned along the positive x-direction, and a state β_x^s for the spin aligned along the negative x-direction. These states would correspond to two orthonormal vectors $\vec{\alpha}_x$ and $\vec{\beta}_x$ in $\vec{\mathbb{V}}^2$. These new vectors can be expressed in terms of the basis vectors $\vec{\alpha}_z$ and $\vec{\beta}_z$, i.e.,

$$\vec{\alpha}_x = c_{x+}\vec{\alpha}_z + c_{x-}\vec{\beta}_z, \quad \vec{\beta}_x = c'_{x+}\vec{\alpha}_z + c'_{x-}\vec{\beta}_z. \tag{14.17}$$

where the coefficients c_{x+}, c_{x-}, c'_{x+} and c'_{x-} may be complex. They satisfy

$$|c_{x+}|^2 + |c_{x-}|^2 = 1 \quad \text{and} \quad |c'_{x+}|^2 + |c'_{x-}|^2 = 1. \tag{14.18}$$

When the system is in state α_x^s the probability of a measured result of S_z being $\hbar/2$ or $-\hbar/2$ are given respectively by

$$\wp^{S_z}(\alpha_x^s, \hbar/2) = |c_{x+}|^2 \quad \text{and} \quad \wp^{S_z}(\alpha_x^s, -\hbar/2) = |c_{x-}|^2. \tag{14.19}$$

in accordance to Eqs. (14.10) and (14.12).

Because of the geometric symmetry one would expect the two probabilities to be the same, i.e., the coefficients c_{x+}, c_- should satisfy

$$|c_{x+}|^2 = |c_{x-}|^2. \tag{14.20}$$

For a measurement of S_z in state β_x^s we have similar results, i.e.,

$$\wp^{S_z}(\beta_x^s, \hbar/2) = |c'_{x+}|^2 \quad \text{and} \quad \wp^{S_z}(\beta_x^s, -\hbar/2) = |c'_{x-}|^2, \tag{14.21}$$

where

$$|c'_{x+}|^2 = |c'_{x-}|^2. \tag{14.22}$$

A natural choice of coefficients satisfying Eqs. (14.18), (14.20) and (14.22) is

$$c_{x+} = 1/\sqrt{2}, \quad c_{x-} = 1/\sqrt{2}, \tag{14.23}$$

$$c'_{x+} = 1/\sqrt{2}, \quad c'_{x-} = -1/\sqrt{2}. \tag{14.24}$$

As a result we obtain the following expressions for $\vec{\alpha}_x$ and $\vec{\beta}_x$

$$\vec{\alpha}_x = \frac{1}{\sqrt{2}}(\vec{\alpha}_z + \vec{\beta}_z), \quad \vec{\beta}_x = \frac{1}{\sqrt{2}}(\vec{\alpha}_z - \vec{\beta}_z). \tag{14.25}$$

Their matrix representations in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ are

$$\boldsymbol{C}_{\vec{\alpha}_x} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\beta}_x} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{14.26}$$

The operator $\widehat{S}_x$ for S_x must admit $\vec{\alpha}_x$ and $\vec{\beta}_x$ as eigenvectors corresponding to eigenvalues $\hbar/2$ and $-\hbar/2$, respectively. It follows that the operator has the following spectral decomposition:

$$\widehat{S}_x = \frac{1}{2}\hbar\widehat{P}_{\vec{\alpha}_x} - \frac{1}{2}\hbar\widehat{P}_{\vec{\beta}_x} \tag{14.27}$$

$$= \frac{1}{2}\hbar\,|\vec{\alpha}_x\rangle\langle\vec{\alpha}_x| - \frac{1}{2}\hbar\,|\vec{\beta}_x\rangle\langle\vec{\beta}_x|. \tag{14.28}$$

The corresponding matrix representation of $\widehat{P}_{\vec{\alpha}_x}$, $\widehat{P}_{\vec{\beta}_x}$ and $\widehat{S}_x$ are[3]

$$\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{M}_{\hat{P}_{\vec{\beta}_x}} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{14.29}$$

$$\boldsymbol{M}_{\hat{S}_x} = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{14.30}$$

The spectral decomposition of $\boldsymbol{M}_{\hat{A}}$ takes the form:

$$\boldsymbol{M}_{\hat{S}_x} = \frac{\hbar}{2}\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}} - \frac{\hbar}{2}\boldsymbol{M}_{\hat{P}_{\vec{\beta}_x}}. \tag{14.31}$$

So far we have managed without complex numbers or complex vectors, giving an impression that we may be able describe electron spin in a two-dimensional real vector space such as $\vec{E}^2$. This impression turns out to be wrong when we come to consider S_y.

[3] $\widehat{P}_{\vec{\alpha}_x}$, $\widehat{P}_{\vec{\beta}_x}$ are the projection matrices in Eqs. (7.178) and (7.179). All the representations are in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$.

14.1.1.5 On observable S_y

Let us follow the arguments for S_x here. First there must be a state α_y^s for the spin aligned along the positive y-direction, and a state β_y^s for the spin aligned along the negative y-direction. These states would correspond to two orthonormal vectors $\vec{\alpha}_y$ and $\vec{\beta}_y$ in $\vec{\mathbb{V}}^2$. Again these vectors are expressible as linear combinations of basis vectors $\vec{\alpha}_z$ and $\vec{\beta}_z$, i.e., we have

$$\vec{\alpha}_y = c_{y+}\vec{\alpha}_z + c_{y-}\vec{\beta}_z, \quad \vec{\beta}_y = c'_{y+}\vec{\alpha}_z + c'_{y-}\vec{\beta}_z. \tag{14.32}$$

The symmetry argument on the measurement of S_z in states α_x and β_x in the discussion of S_x applies equally well to states α_y^s and β_y^s. This means that Eqs. (14.18), (14.20) and (14.22) apply, i.e., we have

$$|c_{y+}|^2 + |c_{y-}|^2 = 1, \quad |c'_{y+}|^2 + |c'_{y-}|^2 = 1. \tag{14.33}$$

$$|c_{y+}|^2 = |c_{y-}|^2, \quad |c'_{y+}|^2 = |c'_{y-}|^2. \tag{14.34}$$

We need to choose the coefficients in Eq. (14.32) to be different from those in Eqs. (14.23) and (14.24) in order to produce two new vectors $\vec{\alpha}_x$ and $\vec{\beta}_x$. A natural choice of a such a set of coefficients is

$$c_{y+} = 1/\sqrt{2}, \quad c_{y-} = i/\sqrt{2}, \tag{14.35}$$

$$c'_{y+} = 1/\sqrt{2}, \quad c'_{y-} = -i/\sqrt{2}, \tag{14.36}$$

which would result in the following two new complex vectors to describe the states α_y^s and β_y^s:

$$\vec{\alpha}_y = \frac{1}{\sqrt{2}}(\vec{\alpha}_z + i\vec{\beta}_z), \quad \vec{\beta}_y = \frac{1}{\sqrt{2}}(\vec{\alpha}_z - i\vec{\beta}_z). \tag{14.37}$$

The operator $\widehat{S}_y$ for S_y must admit $\vec{\alpha}_y$ and $\vec{\beta}_y$ as eigenvectors corresponding to eigenvalues $\hbar/2$ and $-\hbar/2$, respectively. It follows that the operator has the following spectral decomposition

$$\widehat{S}_y = \frac{\hbar}{2}\,\widehat{P}_{\vec{\alpha}_y} - \frac{\hbar}{2}\,\widehat{P}_{\vec{\beta}_y} \tag{14.38}$$

$$= \frac{\hbar}{2}\,|\vec{\alpha}_y\rangle\langle\vec{\alpha}_y| - \frac{\hbar}{2}\,|\vec{\beta}_y\rangle\langle\vec{\beta}_y|. \tag{14.39}$$

The matrix representations $\vec{\alpha}_y$, $\vec{\beta}_y$ and $\widehat{S}_y$ in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ are

$$\boldsymbol{C}_{\vec{\alpha}_y} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\beta}_y} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}, \tag{14.40}$$

$$\boldsymbol{M}_{\hat{S}_y} = \frac{1}{2}\hbar\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{14.41}$$

The matrix representations $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_y}}$ and $\boldsymbol{M}_{\hat{P}_{\vec{\beta}_y}}$ of the projectors $\widehat{P}_{\vec{\alpha}_y}$ and $\widehat{P}_{\vec{\alpha}_y}$ in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ agree with the projection matrices in Eqs. (7.180) and (7.181),[4]

We conclude our discussion with the following comments:

C14.1.1(1) The description of electron spin requires a two-dimensional complex-vector space as state space. Such a state space can be represented by $\vec{\mathbb{C}}^2$.

C14.1.1(2) The matrix representations of spin operators, known as **spin matrices**, are related to the Pauli matrices in Eq. (7.9) by

$$\boldsymbol{M}_{\hat{S}_x} = \frac{\hbar}{2}\sigma_x, \quad \boldsymbol{M}_{\hat{S}_y} = \frac{\hbar}{2}\sigma_y, \quad \boldsymbol{M}_{\hat{S}_z} = \frac{\hbar}{2}\sigma_z. \tag{14.42}$$

C14.1.1(3) The spin matrices $\boldsymbol{M}_{\hat{S}_x}$, $\boldsymbol{M}_{\hat{S}_y}$, $\boldsymbol{M}_{\hat{S}_z}$ obey the following commutation relations:

$$[\,\boldsymbol{M}_{\hat{S}_x}, \boldsymbol{M}_{\hat{S}_y}\,] = i\hbar\boldsymbol{M}_{\hat{S}_z}, \tag{14.43}$$

$$[\,\boldsymbol{M}_{\hat{S}_z}, \boldsymbol{M}_{\hat{S}_x}\,] = i\hbar\boldsymbol{M}_{\hat{S}_y}, \tag{14.44}$$

$$[\,\boldsymbol{M}_{\hat{S}_y}, \boldsymbol{M}_{\hat{S}_z}\,] = i\hbar\boldsymbol{M}_{\hat{S}_x}. \tag{14.45}$$

C14.1.1(4) The spin operators $\widehat{S}_x$, $\widehat{S}_y$ and $\widehat{S}_z$ must also satisfy the same commutation relations, i.e.,[5]

$$[\widehat{S}_x, \widehat{S}_y] = i\hbar\widehat{S}_z, \quad [\widehat{S}_z, \widehat{S}_x] = i\hbar\widehat{S}_y, \quad [\widehat{S}_y, \widehat{S}_z] = i\hbar\widehat{S}_x. \tag{14.46}$$

These commutation relations are characteristic of angular momentum operators, e.g., orbital angular momentum operators also obey these commutation relations as will be shown in §27.4.

[4] *See* Q14(2).

[5] *See* Eq. (13.115) which shows that operators and their matrix representations have the same commutation relations.

14.1.2 Spin-1 Particles

For the description of the spin of a spin-1 particle we would need a three-dimensional complex vector space, i.e., $\vec{\mathbb{V}}^3$.[6] Let the spin operators along the x, y and z directions be denoted by $\widehat{S}_{(1)x}$, $\widehat{S}_{(1)y}$ and $\widehat{S}_{(1)z}$, respectively. These operators act on a three-dimensional vector space $\vec{\mathbb{V}}^3$. We would expect these operators to obey the commutation relations in Eq. (14.46), i.e.,

$$[\widehat{S}_{(1)x}, \widehat{S}_{(1)y}] = i\hbar\widehat{S}_{(1)z}, \quad [\widehat{S}_{(1)z}, \widehat{S}_{(1)x}] = i\hbar\widehat{S}_{(1)y},$$
$$[\widehat{S}_{(1)y}, \widehat{S}_{(1)z}] = i\hbar\widehat{S}_{(1)x}. \tag{14.47}$$

The commutation relations which contain the imaginary number i imply that in any matrix representation of the three spin-1 operators the matrices cannot be all real, as seen in the following well-known matrix representation of the above operators[7]:

$$\boldsymbol{M}_{\hat{S}_{(1)x}} = \frac{\hbar}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{14.48}$$

$$\boldsymbol{M}_{\hat{S}_{(1)y}} = \frac{\hbar}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \tag{14.49}$$

$$\boldsymbol{M}_{\hat{S}_{(1)z}} = \hbar\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{14.50}$$

These matrices possess real eigenvalues $\hbar$, 0, $-\hbar$. The corresponding normalised eigenvectors are

(1) For the matrix $\boldsymbol{M}_{\hat{S}_{(1)x}}$:

$$\frac{1}{2}\begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{2} \\ -1 \end{pmatrix}. \tag{14.51}$$

[6] Zettili pp. 188–204 contains many examples of finite-dimensional complex vector spaces.

[7] Zettili pp. 277–278. Gasiorowicz pp. 150–151. Schiff pp. 145–146.

(2) For the matrix $\boldsymbol{M}_{\hat{S}_{(1)y}}$:

$$\frac{1}{2}\begin{pmatrix} -i \\ \sqrt{2} \\ i \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} i \\ 0 \\ i \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} i \\ \sqrt{2} \\ -i \end{pmatrix}. \tag{14.52}$$

(3) For the matrix $\boldsymbol{M}_{\hat{S}_{(1)z}}$:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{14.53}$$

Clearly the three-dimensional real vector space $\vec{I\!E}^3$ is not able to form the state space for the description of spin-1 particles.

14.2 Generating Probability Distribution Functions on $\vec{\mathbb{W}}^N$

Clearly we can go on to establish a probability theory in $\vec{\mathbb{W}}^N$ for model systems with properties as described in §10.2.1 for observables A which can assumes N values $a_1, a_2, \cdots, a_N$. Such a theory can be established as follows:

(1) Take the state space to be $\vec{\mathbb{W}}^N$. A state η^s is described by a unit vector $\vec{\eta}$ in $\vec{\mathbb{W}}^N$.

(2) Choose an appropriate complete orthonormal set $\{\vec{\varepsilon}_\ell, \ell = 1, 2, \cdots, N\}$ of vectors in $\vec{\mathbb{W}}^N$ together with the set of values $a_1, a_2, \cdots, a_N$ of A to construct a selfadjoint operator $\widehat{A}$ by

$$\widehat{A} := \sum_{\ell=1}^{N} a_\ell \, \widehat{P}_{\vec{\varepsilon}_\ell}. \tag{14.54}$$

(3) Describe observable A by the selfadjoint operator $\widehat{A}$.

(4) The complete orthogonal family of projectors $\widehat{P}_{\vec{\varepsilon}_\ell} = |\vec{\varepsilon}_\ell\rangle\langle\vec{\varepsilon}_\ell|$ generates a probability mass function on the sample space

$$S_{am} := \{a_1, a_2, \cdots, a_N\} \tag{14.55}$$

for each unit vector $\vec{\eta}$ by

$$a_\ell \to \wp^{\hat{A}}(\vec{\eta}, a_\ell) := \langle \vec{\eta} \mid \widehat{P}_{\vec{\varepsilon}_\ell} \vec{\eta} \rangle. \tag{14.56}$$

We can then employ probability mass function $\wp^{\hat{A}}(\vec{\eta}, a_\ell)$ to describe the probability distribution of the set of values $\{a_\ell\}$ of A in state η^s.

(5) The expectation value is then given by

$$\mathcal{E}(\widehat{A}, \vec{\eta}) := \sum_{\ell=1}^{N} a_\ell \wp^{\hat{A}}(\vec{\eta}, a_\ell) = \langle \vec{\eta} \mid \widehat{A} \vec{\eta} \rangle. \tag{14.57}$$

We can go further to introduce a probability distribution function in accordance with Eq. (3.38), i.e., we have

$$\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau) := \begin{cases} 0 & \text{if } \tau < a_1 \\ \wp^{\hat{A}}(\vec{\eta}, a_1) & \text{if } a_1 \le \tau < a_2 \\ \wp^{\hat{A}}(\vec{\eta}, a_1) + \wp^{\hat{A}}(\vec{\eta}, a_2) & \text{if } a_2 \le \tau < a_3 \\ \cdots & \cdots \\ \sum_{\ell=1}^{n} \wp^{\hat{A}}(\vec{\eta}, a_\ell) & \text{if } a_n \le \tau < a_{n+1} \\ \cdots & \cdots \\ \sum_{\ell=1}^{N} \wp^{\hat{A}}(\vec{\eta}, a_\ell) = 1 & \text{if } a_N \le \tau \end{cases}. \tag{14.58}$$

This is a piecewise-constant function with discontinuous jumps occurring at $\tau_\ell = a_\ell$ with the jump at each τ_ℓ equal to

$$\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau_\ell) - \mathcal{F}^{\hat{A}}(\vec{\eta}, \tau_\ell - 0) = \wp^{\hat{A}}(\vec{\eta}, a_\ell). \tag{14.59}$$

This is equal to the probability of a measured value of A in state η^s to be a_ℓ. The expectation value can be written in terms of a Stieltjes integral in accordance with Eq. (4.90):

$$\mathcal{E}(\widehat{A}, \vec{\eta}) = \int_{-\infty}^{\infty} \tau \, d\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau). \tag{14.60}$$

The probability distribution function being discrete there is no probability density function, except in the sense of Dirac delta functions in Eq. (4.73).

We can define a corresponding probability measure by Eq. (4.89), i.e.,

$$\mathcal{M}^{\hat{A}}(\vec{\eta}, \Lambda) := \int_{\Lambda} d\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau). \tag{14.61}$$

Later on we shall extend our studies to physical systems with observables which take on a continuous set of values. It is then necessary to employ probability distribution functions and probability measures to describe the distribution of their values. Probability distribution functions and probability measures provide a uniform treatment of both discrete and continuous cases.

Exercises and Problems

Q14(1) Verify the matrix representation of $\boldsymbol{M}_{\hat{S}_x}$ in Eq. (14.30) and of $\boldsymbol{M}_{\hat{S}_y}$ in Eq. (14.41).

Q14(2) Find the matrix representation of projectors $\widehat{P}_{\vec{\alpha}_y}$ and $\widehat{P}_{\vec{\beta}_y}$ in Eq. (14.38) in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$.

Q14(3) Find the probability distribution function for the x-component spin values in the state given by the vector $\vec{\alpha}_z$. What is the corresponding expectation value?

Q14(4) In the two-dimensional vector space $\vec{\mathbb{V}}^2$ we have the following matrix representations in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$:

(1) The matrix representations of the projectors $\widehat{P}_{\vec{\alpha}_x}$ and $\widehat{P}_{\vec{\beta}_x}$ are given by $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}}$ and $\boldsymbol{M}_{\hat{P}_{\vec{\beta}_x}}$ in Eq. (14.29).

(2) The matrix representations of projectors $\widehat{P}_{\vec{\alpha}_y}$ and $\widehat{P}_{\vec{\beta}_y}$ are given by the matrices $\boldsymbol{P}_{\vec{\alpha}_y}$ and $\boldsymbol{P}_{\vec{\beta}_y}$ in Eqs. (7.180) and (7.181). In the notation of this chapter these matrices are relabelled as $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_y}}$ and $\boldsymbol{M}_{\hat{P}_{\vec{\beta}_y}}$.

(3) The vector $\vec{\eta}$ is represented by the column vector $\boldsymbol{C}_{\vec{\eta}}$ in Eq. (14.3).

Using the expression for $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}} = \boldsymbol{C}_{\vec{\alpha}_x}\boldsymbol{C}^{\dagger}_{\vec{\alpha}_x}$ in Eq. (7.178) evaluate $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}}\boldsymbol{C}_{\vec{\eta}}$. Show that the same result is obtained using explicit matrix representations of $\boldsymbol{C}_{\vec{\alpha}_x}$ and $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}}$ in Eqs. (14.3), (14.26) and (14.29). Explain how the result confirms the projection nature of the matrix $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_x}}$.
Carry out a similar evaluation of $\boldsymbol{M}_{\hat{P}_{\vec{\beta}_x}}\boldsymbol{C}_{\vec{\eta}}$, $\boldsymbol{M}_{\hat{P}_{\vec{\alpha}_y}}\boldsymbol{C}_{\vec{\eta}}$ and $\boldsymbol{M}_{\hat{P}_{\vec{\beta}_y}}\boldsymbol{C}_{\vec{\eta}}$.

Q14(5) Write down a formal expression of the probability density function $w^{\hat{A}}(\vec{\eta}, \tau)$ for the piecewise-constant probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau)$ in Eq. (14.58) in terms of Dirac delta functions.[8] Write down the Stieltjes integral in Eq. (14.60) for the expectation value in terms of the probability density function $w^{\hat{A}}(\vec{\eta}, \tau)$ obtained above and evaluate the integral.

[8] *See* Eq. (4.74).

Chapter 15

Spectral Theory in $\vec{\mathbb{W}}^N$ in Terms of Stieltjes Integrals

In $\vec{\mathbb{W}}^N$ the spectral theorem for selfadjoint operators is summarised in Theorem 13.3.2(2). A selfadjoint operator in an infinite-dimensional vector space $\vec{\mathbb{W}}^\infty$ may be expected to have a countably infinite number of eigenvalues with a corresponding countably infinite set of eigenvectors and eigenprojectors. One may also expect a similar spectral theorem to apply. Unfortunately this is not generally true. The eigenvalue equation for some selfadjoint operators $\widehat{A}$ in $\vec{\mathbb{W}}^\infty$ do not admit any solutions, i.e., there may not exist any vector $\vec{\eta} \in \vec{\mathbb{W}}^\infty$ satisfying the eigenvalue equation

$$\widehat{A}\,\vec{\eta} = a\,\vec{\eta} \quad \text{for some constant } a. \tag{15.1}$$

Examples in Eqs. (18.9) to (18.12) in §18.1 serve to demonstrate this fact. A new approach is required to recover a notion of eigenvalues and eigenvectors and a form of spectral theorem for such operators. As discussed in the relation to the examples cited above a generalised concept of eigenvalues can be established with the result that these generalised eigenvalues can form a continuous, rather than a discrete, set of values. Classical probability theory for discrete sample spaces discussed in Chapter 3 fails when the sample

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

space is continuous. A probability distribution function rather than by a probability mass function has to be introduced. We can also introduce a probability distribution function in the discrete case to obtain a uniform treatment of both discrete and continuous sample spaces, as done in Chapter 4. A similar situation exists with operators. A selfadjoint operator on $\vec{\mathbb{V}}^N$ possesses a complete orthogonal family of projectors. But a selfadjoint operator in $\vec{\mathbb{V}}^\infty$ may not possess a complete orthogonal family of projectors in the same way. Instead it has a **spectral function** associated with it. Spectral functions are a generalisation of a complete orthogonal family of projectors.

In this chapter we shall introduce spectral functions for selfadjoint operators in $\vec{\mathbb{V}}^N$ in a way which can be generalised to $\vec{\mathbb{V}}^\infty$.

15.1 Spectral Functions and Spectral Measures

Equation (13.36) for the spectral decomposition of a selfadjoint operator $\widehat{A}$ can be rewritten in terms of Stieltjes integrals in the same way a discrete sum is expressible as a Stieltjes integral in Eq. (4.72). To do so we need to introduce the concept of projector-valued functions $\widehat{F}(\tau)$[1]:

Definition 15.1(1)

(1) *A projector-valued function of a real variable τ is a mapping from the reals $\mathbb{R}$ into the set of projectors on a given vector space, i.e., the function assigns a projector to each value of the variable τ.*

(2) *A projector-valued set function on the Borel sets of $\mathbb{R}$ assigns a projector on a given vector space to each Borel set.*

As an example consider a nondegenerate selfadjoint operator $\widehat{A}$ on a given vector space with a complete orthonormal set of eigenvectors $\vec{\epsilon}_\ell$ corresponding to eigenvalues a_ℓ. Define the following projector-valued function in terms of the complete orthogonal family of

[1] Prugovečki pp. 231–237.

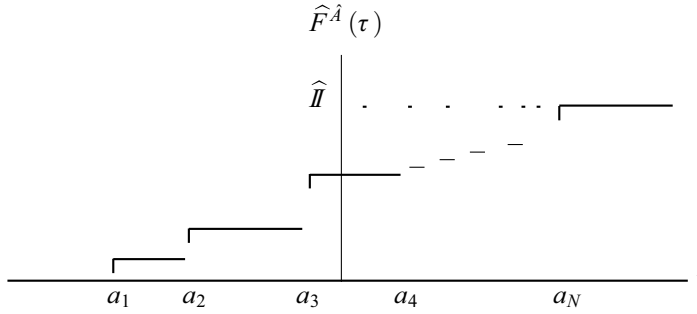


Figure 15.1 Projector-valued function.

eigenprojectors $\widehat{P}_{\vec{\varepsilon}_\ell}$:

$$\widehat{F}^{\hat{A}}(\tau) := \begin{cases} \widehat{0}, & \tau < a_1 \\ \widehat{P}_{\vec{\varepsilon}_1}, & a_1 \le \tau < a_2 \\ \widehat{P}_{\vec{\varepsilon}_1} + \widehat{P}_{\vec{\varepsilon}_2}, & a_2 \le \tau < a_3 \\ \cdots & \cdots \\ \sum_{\ell=1}^{n} \widehat{P}_{\vec{\varepsilon}_\ell}, & a_n \le \tau < a_{n+1} \\ \cdots & \cdots \\ \sum_{\ell=1}^{N} \widehat{P}_{\vec{\varepsilon}_\ell} = \widehat{\mathit{II}}, & a_N \le \tau \end{cases} . \tag{15.2}$$

This function is closely related to the probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau)$ in Eq. (14.58). On account of Eq. (14.56) we can see that $\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau)$ is equal to the quadratic form generated by the above projector-valued function, i.e.,

$$\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau) = \langle \vec{\eta} \mid \widehat{F}^{\hat{A}}(\tau)\, \vec{\eta} \rangle. \tag{15.3}$$

In other words, the *cumulative probability* is given by the quadratic form generated by the projector-valued function defined by Eq. (15.2).[2]

Figure 15.1 shows the similarity of $\widehat{F}^{\hat{A}}(\tau)$ to the function in Eq. (3.38) depicted in Figure 3.6, e.g., the present projector-valued function is also piecewise-constant.[3]

[2] As mentioned in PP3.6(1) cumulative probabilities are probabilities of events which contain many outcomes

[3] The eigenvalues are arranged in an ascending order, i.e., $a_1 < a_2 < \cdots < a_N$.

The projector-valued function $\widehat{F}^{\hat{A}}(\tau)$ possesses the following properties:

P15.1(1) $\widehat{F}^{\hat{A}}(\tau)$ increases from the zero operator $\widehat{0}$ to the identity operator $\widehat{\mathbb{I}}$ as τ increases from $-\infty$ to ∞, i.e.,[4]

$$\widehat{F}^{\hat{A}}(-\infty) = \widehat{0}, \quad \widehat{F}^{\hat{A}}(\infty) = \widehat{\mathbb{I}}. \tag{15.4}$$

$$\widehat{F}^{\hat{A}}(\tau_1) \leq \widehat{F}^{\hat{A}}(\tau_2) \quad \text{if} \quad \tau_1 \leq \tau_2. \tag{15.5}$$

P15.1(2) $\widehat{F}^{\hat{A}}(\tau)$ is continuous from the right, i.e., in terms of the notation used in Eqs.(3.39) to (3.43) we have

$$\widehat{F}^{\hat{A}}(\tau + 0) = \widehat{F}^{\hat{A}}(\tau) \quad \forall \tau \in I\!R. \tag{15.6}$$

In particular we have

$$\widehat{F}^{\hat{A}}(a_\ell + 0) = \widehat{F}^{\hat{A}}(a_\ell) \quad \forall a_\ell. \tag{15.7}$$

The limits of a projector-valued function means[5]

$$\widehat{F}(\tau) \to \widehat{F}(\tau_0) \quad \Leftrightarrow \quad ||\big(\widehat{F}(\tau) - \widehat{F}(\tau_0)\big)\vec{\eta}\,|| \to 0 \tag{15.8}$$

for every $\vec{\eta} \in \vec{\mathbb{V}}^N$ as $\tau \to \tau_0$ from the right.

P15.1(3) The function $\widehat{F}^{\hat{A}}(\tau)$ remains unchanged for all τ except when τ is equal to an eigenvalues where it undergoes a discontinuous jump.

P15.1(4) Each discontinuous jump of $\widehat{F}^{\hat{A}}(\tau)$ is associated with an eigenvalue. It follows that *we can regard eigenvalues as the values of* τ *at which* $\widehat{F}^{\hat{A}}(\tau)$ *is discontinuous.* The jump at an eigenvalue a_ℓ is equal to the corresponding eigenprojector $\widehat{P}_{\vec{\varepsilon}_\ell}$, i.e.,

$$\widehat{F}^{\hat{A}}(a_\ell) - \widehat{F}^{\hat{A}}(a_\ell - 0) = \widehat{P}_{\vec{\varepsilon}_\ell}. \tag{15.9}$$

These properties are similar to that of distribution functions introduced in Definition 3.6(2).

P15.1(5) For a degenerate selfadjoint operator $\widehat{A}$ with degenerate eigenvalues a_m, $m = 1, 2, \cdots, M < N$ the eigenprojectors in

[4] The $\widehat{F}^{\hat{A}}(\tau)$ increases in the sense of the order relation of projectors defined by Eq. (13.20) in Definition 13.2.2(2).

[5] Wan p. 93 for a general discussion on the convergence of a family of operators.

Eq. (15.2) should be replaced by $\widehat{P}^{\hat{A}}(a_m)$ with N replaced by M.[6] The projector $\widehat{P}^{\hat{A}}(a_m)$ which projects onto the eigensubspace of the eigenvalue a_m may not be one-dimensional.

P15.1(6) The jump at an eigenvalue a_ℓ is equal to the corresponding eigenprojector $\widehat{P}^{\hat{A}}(a_\ell)$, i.e.,

$$\widehat{F}^{\hat{A}}(a_\ell) - \widehat{F}^{\hat{A}}(a_\ell - 0) = \widehat{P}^{\hat{A}}(a_\ell). \tag{15.10}$$

For a nondegenerate eigenvalue a_ℓ the eigenprojector is equal to the projector generated by the corresponding eigenvector $\vec{\varepsilon}_\ell$, i.e., $\widehat{P}^{\hat{A}}(a_\ell) = \widehat{P}_{\vec{\varepsilon}_\ell} = |\vec{\varepsilon}_\ell\rangle\langle\vec{\varepsilon}_\ell|$.

Let us formalise this type of projector-valued functions in the following definition.

Definition 15.1(2)[7] *A projector-valued function on the real line $\mathbb{R}$ which assigns a projector $\widehat{F}(\tau)$ on $\vec{\mathbb{V}}^N$ to each $\tau \in \mathbb{R}$ is called a* **spectral function** *if it satisfies the following properties:*

SF15.1(1)	*Non-decreasing*	$\widehat{F}(\tau_1) \leq \widehat{F}(\tau_2)$ if $\tau_1 \leq \tau_2$.
SF15.1(2)	*Values at infinities*	$\widehat{F}(-\infty) = \widehat{0}$, $\widehat{F}(\infty) = \widehat{\mathbb{I}}$.
SF15.1(3)	*Continuity from the right*	$\widehat{F}(\tau + 0) = \widehat{F}(\tau)$.

The projector-valued function $\widehat{F}^{\hat{A}}(\tau)$ defined by Eq. (15.2) is a spectral function. For a spectral function Eq. (13.25) implies that[8]

$$\widehat{F}(\tau_1)\widehat{F}(\tau_2) = \widehat{F}(\tau_2)\widehat{F}(\tau_1) = \widehat{F}(\tau_1) \quad \text{if} \quad \tau_1 \leq \tau_2. \tag{15.11}$$

It follows from Theorem 13.2.2(1) that the difference $\widehat{F}(\tau_2) - \widehat{F}(\tau_1)$, where $\tau_2 > \tau_1$, is also a projector.

Next we shall introduce *spectral measures* with properties similar to that of probability measures given in Definition 3.4(2). Let $\widehat{\mathbb{I}}$ be the identity operator on $\vec{\mathbb{V}}^N$ and let $\Lambda_1, \Lambda_2, \cdots$ be any sequence

[6] Here all a_m are different and M is the total number of different eigenvalues which is less than N since some of the eigenvalues are degenerate.

[7] Prugovečki §5.4 on p. 231, §5.5 on p. 235, Akhiezer and Glazman Vol. 2 §61. Roman Vol. 2 §13.4b on p. 633. Wan pp. 142–148. The order relation in SF15.1(1) is defined by Eq. (13.20).

[8] Roman Vol. 2 p. 569. The result means $\widehat{F}(\tau_1)$ and $\widehat{F}(\tau_2)$ commute.

of mutually disjoint Borel sets of $\mathbb{R}$. Then we have the following definition.

Definition 15.1(3) *A* **spectral measure** $\widehat{M}$ *is a projector-valued set function on the Borel sets* $\mathbb{B}$ *of* $\mathbb{R}$ *which assigns a projector* $\widehat{M}(\Lambda)$ *on* $\vec{\mathbb{W}}^N$ *to each Borel set* Λ *of* $\mathbb{R}$ *with the following properties*[9]*:*

SM15.1(1) *Normalisation* $\widehat{M}(\mathbb{R}) = \widehat{\mathbb{I}}$.

SM15.1(2) *Countable Additivity* $\widehat{M}(\Lambda_1 \cup \Lambda_2 \cup \cdots) = \widehat{M}(\Lambda_1) + \widehat{M}(\Lambda_2) + \cdots$ *where* $\Lambda_1, \Lambda_2, \cdots$ *are mutually disjoint.*

Spectral measures possess the following properties[10]:

$$\widehat{M}(\{\emptyset\}) = \widehat{0}, \tag{15.12}$$

$$\widehat{M}(\Lambda_1)\widehat{M}(\Lambda_2) = \widehat{M}(\Lambda_2)\widehat{M}(\Lambda_1), \tag{15.13}$$

$$\widehat{M}(\Lambda_1)\widehat{M}(\Lambda_2) = \widehat{0}, \quad \text{if} \quad \Lambda_1 \cap \Lambda_2 = \emptyset, \tag{15.14}$$

$$\widehat{M}(\Lambda_1 \cup \Lambda_2) = \widehat{M}(\Lambda_1) + \widehat{M}(\Lambda_2) \quad \text{if} \quad \Lambda_1 \cap \Lambda_2 = \emptyset, \tag{15.15}$$

$$\widehat{M}(\Lambda_1 \cup \Lambda_2) = \widehat{M}(\Lambda_1) + \widehat{M}(\Lambda_2) - \widehat{M}(\Lambda_1 \cap \Lambda_2), \tag{15.16}$$

$$\widehat{M}(\Lambda_1 \cap \Lambda_2) = \widehat{M}(\Lambda_1)\widehat{M}(\Lambda_2). \tag{15.17}$$

Probability distribution functions and probability measures are related by Eqs. (4.84), (4.87) and (4.88). Similarly there is a one-to-one correspondence between spectral functions and spectral measures as seen in the following theorem.

Theorem 15.1(1)[11]

(1) *A spectral measure* $\widehat{M}$ *generates a unique spectral function* $\widehat{F}$ *by*

$$\widehat{F}(\tau) := \widehat{M}\big((-\infty, \tau]\big). \tag{15.18}$$

(2) *A spectral function* $\widehat{F}$ *generates a unique spectral measure* $\widehat{M}$ *by*

$$\widehat{M}\big((\tau_1, \tau_2]\big) := \widehat{F}(\tau_2) - \widehat{F}(\tau_1), \tag{15.19}$$

with the value $\widehat{M}\big(\{\tau_0\}\big)$ *for a singleton set* $\{\tau_0\}$ *given by*[12]

$$\widehat{M}\big(\{\tau_0\}\big) := \widehat{F}(\tau_0) - \widehat{F}(\tau_0 - 0). \tag{15.20}$$

[9] Prugovečki p. 231. SM15.1(1) and SM15.1(2) can be compared with those of probability measures in Definition 4.1.2(1).
[10] Prugovečki pp. 232, 236. Blank, Exner and Havlíček p. 151.
[11] Prugovečki p. 236.
[12] Blank, Exner and Havlíček p. 172.

While a spectral measure clearly determines a spectral function by Eq. (15.18) the converse is less obvious. Using the properties of spectral measures and Eqs. (15.19) and (15.20) we can extend the set function from half-open intervals $(\tau_1, \tau_2]$ to all Borel sets.[13] It is not difficult to show that Eq. (15.19) can be extended to more general intervals. First Eq. (15.19) agrees with Eq. (15.18) for a semi-infinite interval $(-\infty, \tau]$. For other finite intervals we have[14]

$$\widehat{M}\big([\tau_1, \tau_2]\big) = \widehat{F}(\tau_2) - \widehat{F}(\tau_1 - 0), \tag{15.21}$$

$$\widehat{M}\big([\tau_1, \tau_2)\big) = \widehat{F}(\tau_2 - 0) - \widehat{F}(\tau_1 - 0), \tag{15.22}$$

$$\widehat{M}\big((\tau_1, \tau_2)\big) = \widehat{F}(\tau_2 - 0) - \widehat{F}(\tau_1). \tag{15.23}$$

15.2 Spectral Measures in Terms of Riemann-Stieltjes Integrals

To keep up with a close analogy with the relationship between probability distribution functions and probability measures in Eq. (4.89) we can rewrite Eq. (15.19) in the form of an integral[15]:

$$\widehat{M}\big((\tau_1, \tau_2]\big) = \int_{\tau_1}^{\tau_2} d_\tau \widehat{F}(\tau). \tag{15.24}$$

Such an integral expression can be extended to an arbitrary Borel set Λ with

$$\widehat{M}(\Lambda) = \int_\Lambda d_\tau \widehat{F}(\tau). \tag{15.25}$$

The above integral expressions should be understood in terms of their action on vectors.[16] In other words, Eq. (15.25) means

$$\widehat{M}(\Lambda)\vec{\eta} = \int_\Lambda d_\tau \Big(\widehat{F}(\tau)\vec{\eta}\Big) \quad \forall\, \vec{\eta} \in \vec{\mathbb{V}}^N, \tag{15.26}$$

or equivalently and in accordance with Eq. (13.9),

$$\langle \vec{\eta} \mid \widehat{M}(\Lambda)\vec{\eta} \rangle = \int_\Lambda d_\tau \langle \vec{\eta} \mid \widehat{F}(\tau)\vec{\eta} \rangle \quad \forall\, \vec{\eta} \in \vec{\mathbb{V}}^N. \tag{15.27}$$

---

[13] Prugovečki p. 236.

[14] These results are similar to those for the Lebesgue-Stieltjes measure in Eqs. (4.18) to (4.23).

[15] The notation d_τ serves as a reminder that the differentiation is with respect to τ, not to other variables.

[16] Roman Vol. 2 p. 636, Fano p. 346, Naimark Part 2 p. 15.

Later on we shall encounter integrals of the form

$$\widehat{f} = \int_{-\infty}^{\infty} f(\tau) d_\tau \widehat{F}(\tau), \tag{15.28}$$

where $f(\tau)$ is a real-valued function on $\mathbb{R}$. Following Eq. (15.27) we consider Eq. (15.28) to mean

$$\langle \vec{\eta} \mid \widehat{f}\, \vec{\eta} \rangle = \int_{-\infty}^{\infty} f(\tau)\, d_\tau \langle \vec{\eta} \mid \widehat{F}(\tau) \vec{\eta} \rangle \quad \forall\, \vec{\eta} \in \vec{\mathbb{W}}^N. \tag{15.29}$$

The integral on the right-hand-side is a Riemann-Stieljes integral. In view of its importance we shall examine the meaning of integrals of this kind more carefully.

Following Eq. (15.26) we can write down the action of the integral expression in Eq. (15.28) on vectors, i.e., $\forall\, \vec{\eta} \in \vec{\mathbb{W}}^N$ we have

$$\left(\int_{-\infty}^{\infty} f(\tau) d_\tau \widehat{F}(\tau) \right) \vec{\eta} = \int_{-\infty}^{\infty} f(\tau)\, d_\tau \left(\widehat{F}(\tau) \vec{\eta} \right). \tag{15.30}$$

The right hand side contains an integral which can be defined to be the limit of a sum as for Riemann-Stieltjes integrals introduced in §4.2.3. To understand the limiting process let the real line $\mathbb{R}$ be divided into a large number of intervals by a set of values $\tau_j \in \mathbb{R}$, i.e.,

$$\cdots < \tau_{-2} < \tau_{-1} < \tau_0 < \tau_1 < \tau_2 < \cdots . \tag{15.31}$$

Let m be a large positive integer. Introduce the following finite sums:

$$\widehat{f}_m := \sum_{j=-m}^{m} f(\tau_j') \left(\widehat{F}^{\hat{A}}(\tau_j) - \widehat{F}(\tau_{j-1}) \right), \quad \tau_j' \in (\tau_{j-1}, \tau_j]. \tag{15.32}$$

Given any $\vec{\eta} \in \vec{\mathbb{W}}^N$ let

$$\vec{\eta}_m = \widehat{f}_m \vec{\eta} \tag{15.33}$$

$$= \sum_{j=-m}^{m} f(\tau_j') \left(\widehat{F}^{\hat{A}}(\tau_j) - \widehat{F}^{\hat{A}}(\tau_{j-1}) \right) \vec{\eta}. \tag{15.34}$$

It can be shown that, as $m \to \infty$ and all the intervals $(\tau_{j-1}, \tau_j]$ vanishing, $\widehat{f}_m$ tends to a selfadjoint operator $\widehat{f}$ and $\vec{\eta}_m$ tends to the vector $\widehat{f}\vec{\eta}$, in the sense that for all $\vec{\eta} \in \vec{\mathbb{V}}^N$ we have[17]

$$\lim_{m\to\infty} ||\widehat{f}\,\vec{\eta} - \widehat{f}_m\vec{\eta}\,|| \to 0\,, \quad \text{or} \tag{15.35}$$

$$\lim_{m\to\infty} ||(\widehat{f} - \widehat{f}_m)\vec{\eta}\,|| \;\to\; 0. \tag{15.36}$$

These results are the basis for interpretation that the integrals in Eqs. (15.28) and (15.30). We can symbolise all these results by equating the expression in Eq. (15.28) to the selfadjoint operator $\widehat{f}$, i.e., we write

$$\widehat{f} = \int_{-\infty}^{\infty} f(\tau) d_\tau \widehat{F}(\tau). \tag{15.37}$$

Let us demonstrate this type of integrals with two important examples:

E15.2(1) Setting $f(\tau) = 1$ we get the integral

$$\int_{-\infty}^{\infty} d_\tau \widehat{F}(\tau). \tag{15.38}$$

From the limiting values of $\widehat{F}(\tau)$ we can see that this integral is equal to the identity operator, i.e.,

$$\int_{-\infty}^{\infty} d_\tau \widehat{F}(\tau) = \widehat{\mathit{II}}. \tag{15.39}$$

This integral expression is known as a **spectral decomposition of the identity**.

E15.2(2) Setting $f(\tau) = \tau$ we get the integral

$$\int_{-\infty}^{\infty} \tau\, d_\tau \widehat{F}(\tau). \tag{15.40}$$

The significance of this example will become obvious in the discussion of the spectral theorem in the next section.

---

[17] Roman Vol. 2 p. 640.

15.3 Spectral Theorem and Spectrum

A selfadjoint operator has associated with it a complete orthogonal family of eigenprojectors. These eigenprojectors determine a spectral function by Eq. (15.2). Conversely every spectral function defines a selfadjoint operator by Eq. (15.40). The one-to-one relation between spectral functions and selfadjoint operators is summarised in the theorem below.

Theorem 15.3(1) The Spectral Theorem

(1) *To every selfadjoint operator* $\widehat{A}$ *on* $\vec{\mathbb{W}}^N$ *there corresponds a unique spectral function, known as the spectral function of* $\widehat{A}$ *and denoted by* $\widehat{F}^{\hat{A}}(\tau)$*, such that*

$$\boxed{\widehat{A} = \int_{-\infty}^{\infty} \tau \, d_\tau \widehat{F}^{\hat{A}}(\tau).} \tag{15.41}$$

(2) *Every spectral function on* $\vec{\mathbb{W}}^N$ *defines a selfadjoint operator by* Eq. (15.41)*.*

The above theorem tells us that every spectral function on $\vec{\mathbb{W}}^N$ is the spectral function of a selfadjoint operator.[18] Hence every spectral function on $\vec{\mathbb{W}}^N$ is piecewise-constant of the form of Eq. (15.2). The integral expression for $\widehat{A}$ in Eq. (15.41), known as the **spectral decomposition** of $\widehat{A}$, reduces to a linear combination of eigenprojectors of $\widehat{A}$ shown in Eq. (13.36). In analogy to a similar situation of a numerical Riemann-Stieltjes integral in Eq. (4.72) this reduction is due to the fact that the spectral function $\widehat{F}^{\hat{A}}(\tau)$ is piecewise-constant. It follows that the present theorem agrees with Theorem 13.3.2(2).

Definition 15.3(1) *The spectral measure* $\widehat{M}^{\hat{A}}$ *determined by the spectral function* $\widehat{F}^{\hat{A}}(\tau)$ *of* $\widehat{A}$ *by* Theorem 15.1(1) *is called the* **spectral measure** *of* $\widehat{A}$*. The projector* $\widehat{M}^{\hat{A}}(\Lambda)$ *for any Borel set* Λ *is called a* **spectral projector** *of* $\widehat{A}$ *and the subspace* $\vec{\mathcal{S}}^{\hat{A}}(\Lambda)$ *onto which* $\widehat{M}^{\hat{A}}(\Lambda)$ *projects is called a* **spectral subspace** *of* $\widehat{A}$*.*

[18] Roman Vol. 2 p. 638.

Spectral measures of selfadjoint operators on $\vec{\mathbb{W}}^N$ possess the following properties:

P15.3(1) From Figure 15.1 we can see that

(1) The spectral projector $\widehat{M}^{\hat{A}}(\{\tau\})$ of a singleton set $\{\tau\}$, which is equal to $\widehat{F}^{\hat{A}}(\{\tau\}) - \widehat{F}^{\hat{A}}(\{\tau - 0\})$ by Eq. (15.20), is zero except at a discontinuity of the spectral function $\widehat{F}^{\hat{A}}(\tau)$.

(2) Discontinuities of $\widehat{F}^{\hat{A}}(\tau)$ occur at the eigenvalues a_ℓ of $\widehat{A}$. The spectral projector at a discontinuity $\widehat{M}^{\hat{A}}(\{\tau = a_\ell\})$ is equal to the eigenprojector $\widehat{P}^{\hat{A}}(a_\ell)$ corresponding to the eigenvalue a_ℓ in accordance with Eq. (15.10), i.e.,[19]

$$\widehat{M}^{\hat{A}}(\{\tau = a_\ell\}) = \widehat{F}^{\hat{A}}(a_\ell) - \widehat{F}^{\hat{A}}(a_\ell - 0) = \widehat{P}^{\hat{A}}(a_\ell). \qquad (15.42)$$

P15.3(2) Following Theorem 13.3.4(1) we can deduce that two selfadjoint operators $\widehat{A}$ and $\widehat{B}$ commute if their respective spectral projectors $\widehat{M}^{\hat{A}}(\Lambda_1)$ and $\widehat{M}^{\hat{B}}(\Lambda_2)$ commute for all Borel sets Λ_1 and Λ_2 of $\mathbb{R}$, and the converse is also true, i.e., $\forall\ \Lambda_1,\ \Lambda_2$ we have

$$[\,\widehat{A},\ \widehat{B}\,] = \widehat{0} \quad \Leftrightarrow \quad [\,\widehat{M}^{\hat{A}}(\Lambda_1),\ \widehat{M}^{\hat{B}}(\Lambda_2)] = \widehat{0}. \qquad (15.43)$$

P15.3(3) Spectral projectors and spectral subspaces are generalisation of the notion of eigenprojectors and eigensubspaces introduced in Definition 9.4.4(1). While an eigensubspace is associated with a single eigenvalue *a spectral subspace is generally related to a set of eigenvalues*. As will be discussed in §18.1 selfadjoint operators in infinite-dimensional vector spaces may not possess eigenprojectors and eigensubspaces. However, they always possess spectral projectors and spectral subspaces.

There is an important theorem which relate the spectral measures of two unitarily related selfajoint operators.

Theorem 15.3(2)[20] *Let $\widehat{U}$ be a unitary operator and let $\widehat{A}$ be a selfadjoint operator. Then the spectral measure $\widehat{M}^{\hat{A}'}$ of the unitary transform $\widehat{A}' = \widehat{U}\,\widehat{A}\widehat{U}^\dagger$ of $\widehat{A}$ is equal to the unitary transform of the*

[19] Roman Vol. 2 pp. 650–651.
[20] Prugovečki pp. 269, 331.

spectral measure $\widehat{M}^{\hat{A}}$ *of* $\widehat{A}$*, i.e.,*

$$\widehat{M}^{\hat{A}'} = \widehat{U}\,\widehat{M}^{\hat{A}}\widehat{U}^{\dagger}. \tag{15.44}$$

The spectral functions of two unitarily related operators are related in the same way. This theorem remains valid an infinite-dimensional space.

Finally let us re-introduce the *spectrum* of $\widehat{A}$ in a way which can be generalised to infinite-dimensional spaces.

Definition 15.3(2)[21] *The set of values* τ *is called the spectrum of a selfadjoint operator* $\widehat{A}$ *on* $\vec{\mathbb{W}}^N$*, to be denoted by* $sp(\widehat{A})$*, if for every* $\tau \in sp(\widehat{A})$ *the spectral projector* $\widehat{M}^{\hat{A}}(\Lambda_\tau)$ *for every open interval* $\Lambda_\tau = (\tau - \delta, \tau + \delta)$ *containing* τ *does not vanish.*

The following comments serve to clarify the concept:

C15.3(1) The interval $\Lambda_\tau = (\tau - \delta, \tau + \delta)$ can be arbitrarily small. This means that an element of the spectrum is a point of change of the spectral function. Not every point is a point of change. Figure 15.1 shows that for any τ_0 lying between a_1 and a_2, i.e., $a_1 < \tau_0 < a_2$, the spectral function remains constant in Λ_{τ_0} so that according to Eq. (15.23) we have, for sufficiently small δ,

$$\widehat{M}^{\hat{A}}(\Lambda_{\tau_0}) = \widehat{F}^{\hat{A}}(\tau_0 + \delta) - \widehat{F}^{\hat{A}}(\tau_0 - \delta) = \widehat{0}. \tag{15.45}$$

It follows that not every value of $\tau \in I\!R$ is an element of the spectrum of $\widehat{A}$.

C15.3(2) Intuitively we can see that there are two ways of change:

(1) *Continuous change* A point $\tau_0 \in sp(\widehat{A})$ is a *point of continuous growth* if

$$\widehat{M}^{\hat{A}}\left(\Lambda_{\tau_0}\right) \neq \widehat{0} \quad \text{but} \quad \widehat{M}^{\hat{A}}\left(\Lambda_{\tau_0}\right) \to \widehat{0} \quad \text{as} \quad \delta \to 0. \tag{15.46}$$

Figure 15.1 tells us that a point of continuous growth does not exist.

[21] Prugovečki p. 253. Weidmann p. 200. Wan p. 157. The spectrum can also be defined in terms of the resolvent set. Our present definition is for selfadjoint operators which have a real spectrum.

(2) *Discontinuous change* We call a point $\tau_0 \in sp(\widehat{A})$ a *point of discontinuous growth* or a *jump point* of $\widehat{F}^{\hat{A}}(\tau)$, if the spectral projector at τ_0 is not zero, i.e.,

$$\widehat{M}^{\hat{A}}(\{\tau_0\}) = \widehat{F}^{\hat{A}}(\tau_0 + 0) - \widehat{F}^{\hat{A}}(\tau_0 - 0) \neq \widehat{0}. \tag{15.47}$$

This implies that at a jump point the spectral projector $\widehat{M}^{\hat{A}}(\Lambda_{\tau_0})$ for every open interval Λ_{τ_0} containing τ_0 does not vanish. Figure 15.1 shows that τ_0 must be an eigenvalue of $\widehat{A}$ and the spectral projector $\widehat{M}^{\hat{A}}(\{\tau_0\})$ is equal to the eigenprojector corresponding to the eigenvalue.[22]

For a selfadjoint operator $\widehat{A}$ on $\vec{\mathbb{V}}^N$ we can conclude that

(1) The spectrum does not contain any point of continuous growth.

(2) The spectrum consists of the set of jump points. It follows that $sp(\widehat{A})$ coincides with the set of eigenvalues of $\widehat{A}$, showing that the present definition agrees with previous definition.

(3) The spectral projector at each jump point coincides with the eigenprojector for the corresponding eigenvalue.

The spectrum of a selfadjoint operator on $\vec{\mathbb{V}}^N$ is called *purely discrete* for obvious reason. The corresponding spectral function is also said to be *purely discrete*. Later on we shall study selfadjoint operators in an infinite-dimensional vector space. The spectrum of a selfadjoint operator may well contain only points of continuous growth without any jump points. Such spectrum is referred to as a *purely continuous spectrum* to contrast the present discrete ones. A discussion is given in §15.3 with examples given in §20.4.

15.4 Functions of Commuting Selfadjoint Operators

The expression for functions of a selfadjoint operators in Definition 13.3.3(1) can be rewritten in terms of the spectral function $\widehat{F}^{\hat{A}}(\tau)$ of $\widehat{A}$ in the form of an integral shown in Eq. (15.28).

[22] Prugovečki p. 253. Roman Vol. 2 pp. 650–651.

Definition 15.4(1) *Given a real-valued function $f(\tau)$ on $\mathbb{R}$ we define a corresponding function of a sefadjoint operator $\widehat{A}$ in terms of its spectral function $\widehat{F}^{\hat{A}}(\tau)$ by*[23]

$$f(\widehat{A}) := \int_{-\infty}^{\infty} f(\tau)\, d_\tau \widehat{F}^{\hat{A}}(\tau). \tag{15.48}$$

By extending the above integral into a multiple integral the expression for functions of commuting selfadjoint operators given in Definition 13.3.4(1) can be rewritten in the form of a multiple integral in the same way.

Definition 15.4(2)[24] *Let $\widehat{A}_1$ and $\widehat{A}_2$ be two commuting selfadjoint operators with respective spectral functions $\widehat{F}^{\hat{A}_1}(\tau_1)$ and $\widehat{F}^{\hat{A}_2}(\tau_2)$. Given a real-valued function $f(\tau_1, \tau_2)$ of two independent real varaibles τ_1 and τ_2 we define a corresponding function of $\widehat{A}_1$ and $\widehat{A}_2$ by*

$$f(\widehat{A}_1, \widehat{A}_2) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\tau_1, \tau_2)\, d_{\tau_1} d_{\tau_2} \left(\widehat{F}^{\hat{A}_1}(\tau_1) \widehat{F}^{\hat{A}_2}(\tau_2) \right). \tag{15.49}$$

This can be extended in a straightforward manner to define functions of three or more commuting selfadjoint operators. The definitions on functions of commuting selfadjoint operators in terms of spectral functions presented here apply to infinite-dimensional spaces.

15.5 Probability Distribution and Expectation Values

The importance of selfadjoint operators together with their spectral functions and spectral measures lie in their relation with probability distribution functions and probability measures. As shown in Eq. (15.3) a spectral function $\widehat{F}(\tau)$ together with a unit vector $\vec{\eta}$ generates a probability distribution function $\mathcal{F}(\vec{\eta}, \tau)$. Its corresponding spectral measure $\widehat{M}$ would generate a probability measure $\mathcal{M}(\vec{\eta}, \Lambda)$,

[23] Complex-valued functions of $\widehat{A}$ can be defined in the same way.
[24] Jordan p. 54. *See* Definition 13.3.4(1) for operators with a discrete spectrum.

i.e., we have

$$\mathcal{F}(\vec{\eta}, \tau) := \langle \vec{\eta} \mid \widehat{F}(\tau)\vec{\eta} \rangle \quad \text{and} \quad \mathcal{M}(\vec{\eta}, \Lambda) := \langle \vec{\eta} \mid \widehat{M}(\Lambda)\vec{\eta} \rangle. \quad (15.50)$$

Since a selfadjoint operator $\widehat{A}$ possesses a unique spectral function $\widehat{F}^{\hat{A}}$ and spectral measure $\widehat{M}^{\hat{A}}$ we arrive at the following theorem.

Theorem 15.5(1)[25] *A selfadjoint operator $\widehat{A}$ together with a unit vector $\vec{\eta}$ in $\vec{\mathbb{W}}^N$ generates a unique probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau)$ and a unique probability measure $\mathcal{M}^{\hat{A}}(\vec{\eta}, \Lambda)$ defined in terms of its spectral function $\widehat{F}^{\hat{A}}(\tau)$ and spectral measure $\widehat{M}^{\hat{A}}$ by*

$$\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau) := \langle \vec{\eta} \mid \widehat{F}^{\hat{A}}(\tau)\vec{\eta} \rangle, \quad (15.51)$$

$$\mathcal{M}^{\hat{A}}(\vec{\eta}, \Lambda) := \langle \vec{\eta} \mid \widehat{M}^{\hat{A}}(\Lambda)\vec{\eta} \rangle. \quad (15.52)$$

These are known respectively as the probability distribution function and the probability measure generated by the spectral function $\widehat{F}^{\hat{A}}(\tau)$ and the spectral measure $\widehat{M}^{\hat{A}}(\Lambda)$ of $\widehat{A}$ in $\vec{\eta}$. The resulting expectation value is given by Eq. (4.90).

Corollary 15.5(1) *The expectation value of the probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\eta}, \tau)$, written as $\mathcal{E}(\widehat{A}, \vec{\eta})$, can be given directly in terms of $\widehat{A}$, i.e., we have*

$$\mathcal{E}(\widehat{A}, \vec{\eta}) = \langle \vec{\eta} \mid \widehat{A}\vec{\eta} \rangle. \quad (15.53)$$

To establish this corollary we observe that, according to Eq. (4.90),

$$\begin{aligned} \mathcal{E}(\widehat{A}, \vec{\eta}) &= \int_{-\infty}^{\infty} \tau \, d_\tau \mathcal{F}^A(\vec{\eta}, \tau) = \int_{-\infty}^{\infty} \tau \, d_\tau \langle \vec{\eta} \mid \widehat{F}^{\hat{A}}(\vec{\eta}, \tau)\vec{\eta} \rangle \\ &= \langle \vec{\eta} \mid \left(\int_{-\infty}^{\infty} \tau d_\tau \widehat{F}^{\hat{A}}(\vec{\eta}, \tau) \right) \vec{\eta} \rangle = \langle \vec{\eta} \mid \widehat{A}\vec{\eta} \rangle. \end{aligned} \quad (15.54)$$

The last step is based on the spectral decomposition of $\widehat{A}$ given by the spectral theorem.

Following Eq. (10.10) the uncertainty arising from the probability distribution function can be similarly expressed.

[25] Wan §2.7 to §2.11.

Exercises and Problems

Q15(1) Given a spectral function $\widehat{F}(\tau)$ show that $\widehat{F}(\tau_2) - \widehat{F}(\tau_1)$ for $t_2 > t_1$ is a projector. Is $\widehat{F}(\tau_1) - \widehat{F}(\tau_2)$ also a projector?

Q15(2) Using Eqs. (15.15), (15.19) and (15.20) prove Eqs. (15.21), (15.22) and (15.23).

Q15(3) Show that $\mathcal{F}(\vec{\eta}, \tau)$ and $\mathcal{M}(\vec{\eta}, \Lambda)$ in Eq. (15.50) define a probability distribution function and a probability measure, respectively.

Chapter 16

Infinite-Dimensional Complex Vectors and Hilbert Spaces

16.1 Infinite-Dimensional Vector Spaces

For a quantum system with observables capable of assuming an infinite number of different values we would need an infinite-dimensional complex vector space to serve as its state space.

16.1.1 The Space $\vec{\ell}^{\,2}$

Consider an extension of the space $\vec{\mathbb{C}}^{N}$ to infinite dimensions, i.e., to $\vec{\mathbb{C}}^{\infty}$. The extended space is formed by column vectors having an infinite number of elements, i.e., column vectors of the form

$$\vec{\zeta} := \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ . \\ . \\ . \end{pmatrix}, \quad \vec{\eta} := \begin{pmatrix} \eta_1 \\ \eta_2 \\ . \\ . \\ . \end{pmatrix}, \cdots, \quad \zeta_\ell, \eta_\ell, \cdots \in \mathbb{C}. \tag{16.1}$$

The resulting space is infinite-dimensional, i.e., there are infinite number of linearly independent vectors such as

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

$$\vec{e}_1^{\,c} := \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ . \end{pmatrix}, \quad \vec{e}_2^{\,c} := \begin{pmatrix} 0 \\ 1 \\ . \\ . \\ . \end{pmatrix}, \quad \vec{e}_3^{\,c} := \begin{pmatrix} 0 \\ 0 \\ 1 \\ . \\ . \end{pmatrix} \quad \cdots . \tag{16.2}$$

These vectors form a basis in the above set of infinite-dimensional vectors, i.e., every vector $\vec{\zeta}$ can be written as a linear combination of the above basis vectors with ζ_ℓ as components. We cannot define a scalar product using Eq. (12.11) since a sum over an infinite number of terms may not converge.[1]

Let $\vec{\boldsymbol{\ell}}^{\,2} = \{\vec{\zeta},\ \vec{\eta},\ \cdots\}$ be a subset of $\vec{\mathbb{C}}^{\,\infty}$ satisfying the condition[2]

$$\sum_{\ell=1}^{\infty} |\zeta_\ell|^2 < \infty, \quad \sum_{\ell=1}^{\infty} |\eta_\ell|^2 < \infty, \quad \cdots . \tag{16.3}$$

It is necessary that the components of each vector must tend to zero as ℓ tends to infinity for the infinite sum to converge. The above convergence condition implies that $\vec{\boldsymbol{\ell}}^{\,2}$ is an infinite-dimensional vector space which can also be endowed with a scalar product. This can be proved as follows:

(1) There are two inequalities associated with any two complex numbers $z = a + ib$ and $w = c + id$, where a, b, c, d are real:

$$|z+w|^2 \le 2\Big(|z|^2+|w|^2\Big) \ \text{ and } \ |z^*w| \le \frac{1}{2}\Big(|z|^2+|w|^2\Big). \tag{16.4}$$

(2) Applying the above inequalities to ζ_ℓ and η_ℓ are get

$$|\zeta_\ell + \eta_\ell|^2 \le 2\Big(|\zeta_\ell|^2 + |\eta_\ell|^2\Big), \tag{16.5}$$

$$|\zeta_\ell^* \eta_\ell| \le \frac{1}{2}\Big(|\zeta_\ell|^2 + |\eta_\ell|^2\Big). \tag{16.6}$$

[1] An infinite sum of positive terms may diverge, i.e., producing an infinite value.
[2] The notation means that the infinite sums all converge to a finite number.

(3) The convergence of the sums in Eq. (16.3) and the first inequality above implies the following convergence[3]:

$$\sum_{\ell=1}^{\infty} |\zeta_\ell + \eta_\ell|^2 \le \sum_{\ell=1}^{\infty} 2\Big(|\zeta_\ell|^2 + |\eta_\ell|^2\Big)$$
$$= 2\sum_{\ell=1}^{\infty} |\zeta_\ell|^2 + 2\sum_{\ell=1}^{\infty} |\eta_\ell|^2 < \infty. \qquad (16.7)$$

It follows that given any pair $\vec{\zeta}$ and $\vec{\eta}$ in $\vec{\ell}^{\,2}$ their linear combinations, e.g., $\vec{\zeta} + \vec{\eta}$, are also in $\vec{\ell}^{\,2}$, i.e., the set $\vec{\ell}^{\,2}$ is closed under scalar multiplication and addition, and is therefore a vector space in its own right.[4]

(4) The convergence of the sums in Eq. (16.3) and the second inequality in item (2) above implies

$$\sum_{\ell=1}^{\infty} \zeta_\ell^* \eta_\ell < \infty. \qquad (16.10)$$

This enables us to define a scalar product for any $\vec{\zeta}, \vec{\eta}$ in $\vec{\ell}^{\,2}$ by

$$\langle \vec{\zeta} \mid \vec{\eta} \rangle = \sum_{\ell=1}^{\infty} \zeta_\ell^* \eta_\ell. \qquad (16.11)$$

16.1.2 Spaces of Complex-Valued Functions

Let Λ be an interval of the real line $\mathbb{R}$. The interval Λ may be bounded, e.g., $\Lambda = [0, L]$, semi-infinite, e.g., $\Lambda = [0, \infty)$, or infinite, i.e., $\Lambda = (-\infty, \infty)$. Let us consider functions $f(x)$, both real-valued and complex-valued, defined on Λ. A function $f(x)$ on Λ is said to be differentiable if its derivative is defined at every point in Λ. For a finite interval the function must have right hand derivative at $\tau = 0$ and the left hand derivative at $\tau = L$.[5] As pointed out in §4.2.2

[3] *See* Spiegel (2) p. 141 on relevant theorems on convergence of complex series.

[4] This result can also be directly proved from the following Cauchy and Minkowski inequalities (*see* Kreyszig p. 14):

$$\sum_{\ell=1}^{\infty} |\zeta_\ell^* \eta_\ell| \le \Big(\sum_{m=1}^{\infty} |\zeta_m|^2\Big)^{1/2} \Big(\sum_{n=1}^{\infty} |\eta_n|^2\Big)^{1/2}. \qquad (16.8)$$

$$\Big(\sum_{\ell=1}^{\infty} |\zeta_\ell^* + \eta_\ell|^2\Big)^{1/2} \le \Big(\sum_{m=1}^{\infty} |\zeta_m|^2\Big)^{1/2} + \Big(\sum_{n=1}^{\infty} |\eta_n|^2\Big)^{1/2}. \qquad (16.9)$$

[5] Spiegel (1) p. 57.

a continuous function is not necessarily differentiable. A function may be *once differentiable* everywhere but not *twice differentiable* everywhere, i.e., a function may have a first order derivative df/dx everywhere but its second order derivative d^2f/dx^2 may not exist everywhere. A function on Λ is said to be **infinitely differentiable** if it can be differentiated everywhere for as many times as we wish. Such functions are also said to be **smooth**. Being smooth is quite a stringent requirement.

16.1.2.1 Continuous functions

The following is a list of examples.

E16.1.2.1(1) $C(\Lambda)$ This is the set of continuous complex-valued functions $\phi(x)$ defined on a closed and bounded $\Lambda = [0, L]$ introduced in §12.3.3. This set defines a vector space $\vec{C}(\Lambda)$ with each function $\phi(x)$ defining a vector $\vec{\phi}$. This relationship is denoted by $\vec{\phi} := \phi(x)$ and $\vec{C}(\Lambda) := C(\Lambda)$, a notation introduced in Eq. (12.24). The vector space $\vec{C}(\Lambda)$ is infinite-dimensional. This space becomes a scalar product space with scalar product defined by Eq. (12.26). Many of the infinite-dimensional vector spaces used in quantum mechanics are spaces of functions. We shall often employ the symbols $\vec{\phi}, \vec{\psi}, \cdots$ to denote infinite-dimensional vectors, whether they are defined by functions or not.

E16.1.2.1(2) $C(I\!R^+)$ This is the set of continuous complex-valued functions defined on the half-real line $I\!R^+ = [0, \infty)$. The corresponding vector space $\vec{C}(I\!R^+)$ is infinite-dimensional. However, we cannot define a scalar product for an arbitrary pairs of functions in $C(I\!R^+)$ since we would have to extend the range of the integral in Eq. (12.26) from $[0, L]$ to $[0, \infty)$. The resulting Riemann integrals

$$\int_0^\infty \phi^*(x)\psi(x)dx \tag{16.12}$$

over an infinite interval may not converge to a finite number.

E16.1.2.1(3) $C(I\!R)$ This is the set of continuous complex-valued functions defined on the entire real line $I\!R = (-\infty, \infty)$ with the corresponding vector space denoted by $\vec{C}(I\!R)$. Again we cannot have

a scalar product in terms of integrals of the form of Eq. (12.26) for an arbitrary pair of functions in $C(I\!R)$.

E16.1.2.1(4) $C(I\!R^2)$ This is the set of continuous complex-valued functions defined on $I\!R^2$ with the corresponding vector space denoted by $\vec{C}(I\!R^2)$. Again we cannot have a scalar product in terms of integrals of the form of Eq. (12.26) for an arbitrary pair of functions in $C(I\!R^2)$. Continuous functions can be defined on $I\!R^3$ leading to the sets $C(I\!R^3)$ with the corresponding vector space denoted by $\vec{C}(I\!R^3)$.

16.1.2.2 Absolutely continuous functions

Absolutely continuous functions are introduced in Eq. (4.49). These functions are once-differentiable almost everywhere.

E16.1.2.2(1) $AC(\Lambda)$ The set of absolutely continuous functions defined on a closed and bounded interval $\Lambda = [0, L]$.

E16.1.2.2(2) $AC(I\!R^+)$ The set of absolutely continuous functions defined on $I\!R^+$.

E16.1.2.2(3) $AC(I\!R)$ The set of absolutely continuous functions defined on $I\!R$.

E16.1.2.2(4) $AC(I\!R^2)$ This is the set of absolutely continuous functions defined on $I\!R^2$, i.e., $AC(I\!R^2)$ consists of functions of the Cartesian coordinates x and y which are absolutely continuous in x and y. We can also have functions on $I\!R^2$ absolutely continuous in x only, in y only. We can similarly define the set $AC(I\!R^3)$ of absolutely continuous functions on $I\!R^3$.

These are subsets of $C(\Lambda)$, $C(I\!R^+)$, $C(I\!R)$, $C(I\!R^2)$ and $C(I\!R^3)$, respectively. Their corresponding infinite-dimensional vector spaces are denoted by $\vec{AC}(\Lambda)$, $\vec{AC}(I\!R^+)$, $\vec{AC}(I\!R)$, $\vec{AC}(I\!R^2)$ and $\vec{AC}(I\!R^3)$, respectively. The same problem arises when we try to define a scalar product for these spaces, except for $\vec{AC}(\Lambda)$.

16.1.2.3 Smooth functions

E16.1.2.3(1) $C^\infty(\Lambda)$ This is the set of smooth, i.e., infinitely differentiable and hence the superscript ∞, functions defined on a closed and bounded interval $\Lambda = [0, L]$.

E16.1.2.3(2) $C^\infty(I\!R^+)$ This is the set of *smooth* functions defined on $I\!R^+$.

E16.1.2.3(3) $C^\infty(I\!R)$ This is the set of *smooth* functions defined on $I\!R$.

E16.1.2.3(4) $C^\infty(I\!R^2)$ This is the set of *smooth* functions defined on $I\!R^2$, i.e., functions of Cartesian coordinates x, y which are smooth in both x and y. We can similarly define the set $C^\infty(I\!R^3)$ of smooth functions on $I\!R^3$.

These are subsets of $AC(\Lambda)$, $AC(I\!R^+)$, $AC(I\!R)$, $AC(I\!R^2)$ and $AC(I\!R^3)$, respectively. Their corresponding infinite-dimensional vector spaces are denoted by $\vec{C}^\infty(\Lambda)$, $\vec{C}^\infty(I\!R^+)$, $\vec{C}^\infty(I\!R)$, $\vec{C}^\infty(I\!R^2)$ and $\vec{C}^\infty(I\!R^3)$. There is also a problem defining scalar product.

16.1.2.4 Schwartz functions

Often we are interested in smooth functions $f(x)$ on $I\!R$ which, together with their derivatives $d^n f/dx^n$, tend to zero faster than any power of x at infinity, i.e., $f(x) \in C^\infty(I\!R)$ such that for any $n, m = 1, 2, 3, \ldots$

$$f(x) \text{ and } x^n \frac{d^m f(x)}{dx^m} \to 0 \text{ as } |x| \to \infty. \tag{16.13}$$

These are called **Schwartz functions** on $I\!R$ or *functions of rapid decrease*. These functions define a vector space called a **Schwartz space** and is denoted by $\vec{S}_s(I\!R)$, with the set of Schwarts functions denoted by $S_s(I\!R)$.[6] A polynomial in x is clearly not a Schwartz function while an exponentially decreasing function in x such as $\exp(-x^2)$ is. The product of a polynomial in x and an exponentially decaying function in x would also be a Schwartz function. Let us consider the following well-known examples of such products:

[6] Amrein, Jauch and Sinha p. 32. Gallone p. 55. Conway pp. 336, 342. Gallone p. 55. Moretti pp. 72–73. Schwartz (1915–2002) was a French mathematician. A $f(x) \in C^\infty(I\!R)$ may not tend to zero at infinity. The symbol $\vec{S}_s$ stands for Schwartz space (*see* Definition 16.2.3(1) and the comments after).

1. Hermite polynomials[7] These are polynomials $H_n(y)$ of a real variable y defined by

$$H_0(y) = 1. \tag{16.14}$$

$$H_n(y) := (-1)^n e^{y^2} \frac{d^n}{dy^n} e^{-y^2}, \quad n = 1, 2, 3, \ldots. \tag{16.15}$$

These polynomials possess the following properties:[8]

$$\frac{dH_n(y)}{dy} = 2nH_{n-1}. \tag{16.16}$$

$$H_{n+1}(y) = 2y\, H_n(y) - 2n\, H_{n-1}(y). \tag{16.17}$$

2. Hermite functions These are functions in x defined in terms of an exponentially decreasing function and Hermite polynomials by[9]

$$\varphi_{H_n}(x) := \left(\frac{\sqrt{\lambda}}{\sqrt{\pi}\, 2^n n!}\right)^{1/2} e^{-\lambda x^2/2}\, H_n(\sqrt{\lambda}\, x), \tag{16.18}$$

where λ is a real and positive constant. For application to the quantised harmonic oscillator of mass m and angular frequency ω this constant is assigned the value $\lambda = m\omega/\hbar$.[10]

Hermite functions are orthonormal in the sense that

$$\int_{-\infty}^{\infty} \varphi_{H_n}(x)\varphi_{H_n}(x)\, dx = \delta_{nm}. \tag{16.19}$$

The index n can be raised or lowered by the following operators

$$\widehat{A} := \frac{1}{\sqrt{2}}\left(y + \frac{d}{dy}\right), \quad \widehat{A}^\star := \frac{1}{\sqrt{2}}\left(y - \frac{d}{dy}\right), \tag{16.20}$$

where $y = \sqrt{\lambda}\, x$. Using Eq. (16.16) we can verify that

$$\widehat{A}\varphi_{H0}(x) = 0, \tag{16.21}$$

$$\widehat{A}\varphi_{H_n}(x) = \sqrt{n}\, \varphi_{H(n-1)}(x), \quad n = 1, 2, 3 \ldots, \tag{16.22}$$

$$\widehat{A}^\star\varphi_{H_n}(x) = \sqrt{n+1}\, \varphi_{H(n+1)}(x), \quad n = 0, 1, 2, 3 \ldots. \tag{16.23}$$

---

[7] Greiner pp. 116–117. Marzbacher p. 86. Zettili p. 233. Jauch p. 45. *See* also §35.2 of this book. Hermite (1822–1901) was a French mathematician. Hermite polynomials were introduced before quantum mechanics was formulated.

[8] Jauch p. 213.

[9] Greiner p. 119 and Zettili pp. 232–234 for example, *see* §35.2 for applications.

[10] *See* §35.2.1.

These results are directly relevant to the study of the quantum harmonic oscillator in §35.2.
We can similarly define Schwartz functions $f(x, y)$ on $\mathbb{R}^2$ and Schwartz space $\vec{S}_s(\mathbb{R}^2)$, Schwartz functions $f(x, y, z)$ on $\mathbb{R}^3$ and Schwartz space $\vec{S}_s(\mathbb{R}^3)$.[11] For applications see §27.4 and §35.2.

16.1.2.5 Smooth functions of compact support

A function on $\mathbb{R}$ is said to be a function of *compact support* if there is a closed and bounded interval outside which the function takes only the value zero.[12] A similar terminology applies to functions on $\mathbb{R}^+$, $\mathbb{R}^2$ and $\mathbb{R}^3$.

E16.1.2.5(1) $C_D^\infty(\Lambda)$ This is the set of *smooth* functions f on Λ which take the value zero at the boundary, i.e., $f(0) = f(L) = 0$.[13]

E16.1.2.5(2) $C_c^\infty(\mathbb{R}^+)$ This is the set of smooth functions on $\mathbb{R}^+$ of *compact support*, i.e., everyone of these functions vanishes outside a closed and bounded interval in $\mathbb{R}^+$. Different functions may vanish outside different intervals.

E16.1.2.5(3) $C_c^\infty(\mathbb{R})$ This is the set of smooth functions on $\mathbb{R}$ of compact support, i.e., each function in $C_c^\infty(\mathbb{R})$ vanishes outside a closed and bounded interval in $\mathbb{R}$.[14] As an example let $\Lambda = (a, b)$ be a open and bounded interval and let $\Lambda_0 = [a_0, b_0]$ be a closed interval inside Λ. Let $\xi(x)$ be a smooth real-valued function such that

$$\xi(x) := \begin{cases} \xi(x) = 1, & x \in \Lambda_0 \\ \xi(x) \in (0, 1), & x \in \Lambda - \Lambda_0 \\ \xi(x) = 0, & x \notin \Lambda \end{cases} . \tag{16.24}$$

This function, shown in the figure below, is known as a *localising function*.[15] It is clearly a member of $C_c^\infty(\mathbb{R})$.

[11] Amrein, Jauch and Sinha pp. 32–33. Weidmann p. 289.
[12] Williamson p. 14. The *support* of a function is basically the set of points on which the function is not zero. A closed and bounded interval is known to be a *compact set*. We shall not delve into the general definition of compact sets.
[13] Subscript D indicates the imposition of the *Dirichlet boundary condition* given in Eq. (17.24).
[14] Weidmann p. 25. Wan p. 9. Symbols $C_0^\infty(\mathbb{R})$ and $C_0^\infty(\mathbb{R}^+)$ are also used.
[15] Wan p. 192.

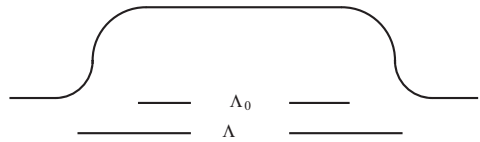


Figure 16.1 Localising function.

E16.1.2.5(4) $C_c^\infty(I\!R^2)$ This is the set of smooth functions defined on $I\!R^2$ vanishing outside a closed and bounded square in $I\!R^2$. We can similarly define the set $C_c^\infty(I\!R^3)$ of smooth functions on $I\!R^3$ vanishing outside a closed and bounded cube in $I\!R^3$.

The corresponding infinite-dimensional vector spaces are denoted $\vec{C}_D^\infty(\Lambda)$, $\vec{C}_c^\infty(I\!R^+)$, $\vec{C}_c^\infty(I\!R)$, $\vec{C}_c^\infty(I\!R^2)$ and $\vec{C}_c^\infty(I\!R^3)$. Since each function vanishes outside a bounded interval the following integral

$$\int \phi^*(x)\psi(x)dx \tag{16.25}$$

over the supports of the functions would converge and can be used to defined a scalar product.

16.1.2.6 Riemann square-integrable functions

E16.1.2.6(1) $\mathcal{L}^2(I\!R)$ This is the set of functions $\phi(x)$ on $I\!R$, not necessarily continuous, which are *Riemann square-integrable* in the sense that the following Riemann integral

$$\int_{-\infty}^{\infty} |\phi(x)|^2\, dx \tag{16.26}$$

converges to a finite value.[16] We shall indicate this by

$$\int_{-\infty}^{\infty} |\phi(x)|^2\, dx < \infty. \tag{16.27}$$

This is not a subset of $C(I\!R)$ since there is no requirement for functions in $\mathcal{L}^2(I\!R)$ to be continuous. These functions form a vector space since any linear combination of two square-integrable functions on $I\!R$ is also square integrable over $I\!R$.[17]

[16] Integrals over an infinite interval, known as *improper integrals*, are defined as the limiting value of the integral over finite intervals. Spiegel (1) p. 260.

[17] This is similar to Eq. (16.7) for $\vec{\ell}^{\,2}$.

The corresponding infinite-dimensional vector space is denoted by $\vec{\mathcal{L}}^2(I\!R)$. Given any $\phi(x)$ and $\psi(x)$ in $\mathcal{L}^2(I\!R)$ we can show that[18]

$$\int_{-\infty}^{\infty} \phi^*(x)\psi(x)dx < \infty. \tag{16.28}$$

A subtle complication arises when we try to use this integral to define a scalar product on $\vec{\mathcal{L}}^2(I\!R)$. Property CSP11.2.2(3) of scalar product is not satisfied, since there are functions $\phi(x)$ such that[19]

$$\int_{-\infty}^{\infty} |\phi(x)|^2\, dx = 0 \quad \not\Rightarrow \quad \phi(x) = 0 \ \forall\, x \in I\!R. \tag{16.29}$$

An example would be the function $f(x)$ given in Eq. (4.13) which is equal to zero for all $x \in I\!R$ except at a single point at $x = 0$ where it is equal to the value 1. Such a situation is discussed earlier in §4.2.2 in the context of Eq. (4.48). A function like $f(x)$ in Eq. (4.13) is said to be *zero almost everywhere*. We can circumvent this difficulty in the following manner:

(1) First we identify all these functions which are zero *almost everywhere* with the *zero function* which is zero everywhere. All these functions are then taken to correspond to the zero vector in $\vec{\mathcal{L}}^2(I\!R)$.
(2) Any given function $\phi(x) \in \mathcal{L}^2(I\!R)$ is taken to correspond a vector $\vec{\phi}$ in $\vec{\mathcal{L}}^2(I\!R)$. We further identity with $\phi(x)$ any function which differs only on a set of measure zero from $\phi(x)$. All these functions which are equal almost everywhere to $\phi(x)$ are then taken to correspond to the same vector $\vec{\phi}$ in $\vec{\mathcal{L}}^2(I\!R)$. This relationship is denoted by

$$\boxed{\vec{\phi} := \phi(x).} \tag{16.30}$$

This enables us to define a scalar product with the notation[20]

$$\langle \vec{\phi} \mid \vec{\psi} \rangle := \int_{-\infty}^{\infty} \phi^*(x)\psi(x)dx. \tag{16.31}$$

[18] This is similar to Eq. (16.10) for $\vec{\ell}^{\,2}$.
[19] Isham p. 30, Fano p. 248.
[20] This is similar to Eq. (16.11) for $\vec{\ell}^{\,2}$.

E16.1.2.6(2) $\mathcal{L}^2(I\!R^+)$ This is the set of functions $\phi(x)$ on $I\!R^+$, not necessarily continuous, which are *Riemann square-integrable* over the range $[0, \infty)$.[21] We can define a scalar product on the corresponding vector space $\vec{\mathcal{L}}^2(I\!R^+)$ by

$$\langle \vec{\phi} \mid \vec{\psi} \rangle := \int_0^\infty \phi^*(x)\psi(x)dx \quad \forall \phi, \psi \in \mathcal{L}^2(I\!R^+). \tag{16.32}$$

E16.1.2.6(3) $\mathcal{L}^2(\Lambda)$ This is formed by functions $\phi(x)$ defined on a closed and bounded interval, e.g., $\Lambda = [0, L]$, not necessarily continuous, which are Riemann square-integrable over the range $[0, L]$, i.e.,

$$\int_0^L |\phi(x)|^2 \, dx < \infty. \tag{16.33}$$

We can define a scalar product on the corresponding vector space $\vec{\mathcal{L}}^2(\Lambda)$ by

$$\langle \vec{\phi} \mid \vec{\psi} \rangle := \int_0^L \phi^*(x)\psi(x)dx \quad \text{for} \quad \phi, \psi \in \mathcal{L}^2(\Lambda) \tag{16.34}$$

16.1.2.7 Lebesgue square-integrable functions

With Riemann integrals for square integrability and the scalar product replaced by Lebesque integrals we would obtain a larger set of functions which are Lebesque square-integrable. The expressions for scalar product in Eqs. (16.31), (16.32) and (16.34) can apply. The corresponding vector spaces are larger than previous ones. The followings are a list of examples.

E16.1.2.7(1) $\vec{L}^2(I\!R)$ is the vector space corresponding to the *space of Lebesgue square-integrable functions on the real line* $I\!R$. It should be pointed out that

(1) Since almost all the functions used in practical applications are Riemann integrable we shall adopt the notation used in Eqs. (16.26) to (16.31) for all integrals. For brevity we shall drop the prefix *Lebesgue* from now on, e.g., we will call $\vec{L}^2(I\!R)$ the **space of square-integrable functions on** $I\!R$.

[21] What has been said earlier about functions which differ only on a set of measure zero also apply here.

(2) Functions which are equal almost everywhere are deemed to be equal. It is in this sense we say that functions in these spaces are defined *almost everywhere*.

These comments apply to all the other spaces listed below.

E16.1.2.7(2) $\vec{L}^2(\mathbb{R}^+)$ is the vector space corresponding to the *space of square-integrable functions on the half real line* $\mathbb{R}^+$.

E16.1.2.7(3) $\vec{L}^2(\Lambda)$ is the vector space corresponding the *space of square-integrable functions on the interval* $\Lambda = [0, L]$.

E16.1.2.7(4) $\vec{L}^2(\mathbb{R}^2)$ is the vector space corresponding to the *space of square-integrable functions on* $\mathbb{R}^2$. This space corresponds to functions $\phi(x, y)$ on the x-y plane which are square-integrable with respect to x and y, i.e.,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\phi(x, y)|^2 \, dxdy < \infty. \tag{16.35}$$

E16.1.2.7(5) $\vec{L}^2(\mathbb{R}^3)$ is the vector space corresponding to the *space of square-integrable functions on* $\mathbb{R}^3$. This space corresponds to functions $\phi(x, y, z)$ on the physical space which are square-integrable with respect to x, y and z, i.e.,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\phi(x, y, z)|^2 \, dxdydz < \infty. \tag{16.36}$$

16.1.2.8 Functions on a circle and $\vec{L}^2(\mathcal{C}_a)$

Consider a circle $\mathcal{C}_a$ of radius a on the x-y plane centred at the origin. In the usual polar coordinates (r, θ) on the x-y plane the circle is specified by

$$\mathcal{C}_a := \Big\{(r, \theta) : \; r = a, \; \theta = [\,0, 2\pi\,]\Big\}, \tag{16.37}$$

where $\theta = 0$ and $\theta = 2\pi$ refer to the same point, taken to be the point of intersection of the circle and the x-axis in the positive direction. The position of a particle on the circle can be specified in two ways:

(1) *Angular position* This is specified by the corresponding polar angle θ, referred to as an *angular position variable*.

(2) *Linear position* This is specified by a variable s, which is the arc length along the circumference from the intersection of the circle with the x-axis to the position the particle. This is referred to as a *linear position variable*.

(3) The two position variables are related by

$$s = a\theta. \tag{16.38}$$

When we wish to emphasise these variables are for the circle $\mathcal{C}_a$ we can use the notation $s(\mathcal{C}_a)$ and $\theta(\mathcal{C}_a)$.[22]

A function, real or complex, defined on the circle can be specified as a function of θ, i.e., we have $\phi = \phi(\theta)$ for $\theta = [\,0,\, 2\pi\,]$. Functions defined on $\mathcal{C}_a$ are similar to functions defined on an interval $\Lambda = [0,\, L]$. A fundamental difference is that $\theta = 0,\, 2\pi$ refer to the same point on the circle. It follows that single-valued functions on $\mathcal{C}_a$, as our functions are, must satisfy the following *periodic boundary* condition[23]:

$$\phi(0) = \phi(2\pi). \tag{16.39}$$

These functions can form a vector space in the same way functions on Λ do. As vectors, to be denoted by $\vec{\phi}$ and $\vec{\psi}$ and so on, we can define a scalar product by integration with respect to the angular position variable θ, i.e.,[24]

$$\langle \vec{\phi} \mid \vec{\psi} \rangle := \int_0^{2\pi} \phi^*(\theta)\psi(\theta)\, d\theta. \tag{16.40}$$

The set $L^2(\mathcal{C}_a)$ of square-integrable functions $\phi(\theta)$ on the circle $\mathcal{C}_a$ defines a vector space $\vec{L}^2(\mathcal{C}_a)$ with the vectors corresponding to the function $\phi(\theta)$ denoted by $\vec{\phi}(\theta)$.[25]

A circle is topologically different from an interval. A single coordinate variable θ is not able to cover the circle properly since

[22] *See* §27.8.
[23] *See* Eq. (17.26).
[24] We may define the scalar product integrating with respect to the linear position variable $a\theta$ with the limits of integration from 0 to $2\pi a$.
[25] As before the integrals are meant to be Lebesque integrals.

$\theta = 0, 2\pi$ refer to the same point on the circle.[26] A proper coordinate covering of the circle can be achieved as follows:

(1) Cover the circle $\mathcal{C}_a$ with overlapping open arcs $S_1, S_2, \ldots$. Each point in the circle is contained in an arc, e.g., there is an arc containing the point $\theta = 0$.[27]
(2) Since each arc S_j is topologically equivalent to an open interval we can have a linear position coordinate s_j or the corresponding angular position variable $\theta_j = s_j/a$ in each arc.
(3) In the overlapping region of two arcs S_1 and S_2 the two coordinates s_1 and s_2 should be smoothly related, i.e., they should be smooth functions of each other.[28] This condition applies to all overlapping regions.

We can then describe things locally in each arc S_j in terms of coordinate θ_j. For example, the differentiability of function at the point $\theta = 0$ is defined in terms of the differentiability of the function with respect to the position variable in the arc containing the point $\theta = 0$. In particular:

(1) A function on the circle is smooth if it is smooth on each arc.
(2) A function on the circle is absolutely continuous if it is absolutely continuous on each arc. The set of absolutely continuous functions on the circle $\mathcal{C}_a$ is denoted by $AC(\mathcal{C}_a)$.

As will be seen in §17.3 and §19.3 we often employ the variable θ to cover the circle in many practical applications.

16.1.2.9 Functions on a unit sphere

Let $\mathcal{S}_u$ be the sphere of unit radius centred at the origin of a Cartesian coordinate system (x, y, z). A sphere can be specified in terms of spherical coordinates (r, θ, φ) which are related to the Cartesian

[26] A coordinate system should assign a unique set of coordinates to each point in the geometric space.
[27] An open arc is one which corresponds to an open interval of a linear position variable.
[28] Wan pp. 57–60 for more details.

coordinates (x, y, z) by

$$x = r\sin\theta\cos\varphi, \quad y = r\sin\theta\sin\varphi, \quad z = r\cos\theta, \tag{16.41}$$

where

$$r \in [0, \infty), \quad \theta \in [0, \pi], \quad \varphi \in [0, 2\pi]. \tag{16.42}$$

The unit sphere $\mathcal{S}_u$ is specified by $(r = 1, \theta, \varphi)$. Functions, real or complex, on the sphere are functions of the two angle variables θ and φ. A function $\phi(\theta, \varphi)$ is square-integrable on $\mathcal{S}_u$ if [29]

$$\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} |\phi(\theta, \varphi)|^2 \sin\theta d\theta d\varphi < \infty. \tag{16.43}$$

The set $L^2(\mathcal{S}_u)$ of square-integrable functions $\phi(\theta, \varphi)$, $\psi(\theta, \varphi), \ldots$ on the sphere defines to a vector space $\vec{L}^2(\mathcal{S}_u)$, with each function $\phi(\theta, \varphi)$ corresponding to a vector $\vec{\phi}$. The scalar product of two vectors $\vec{\phi}$ and $\vec{\psi}$ is defined by

$$\langle \vec{\phi} \mid \vec{\psi} \rangle := \int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} \phi^*(\theta, \varphi)\psi(\theta, \phi)\sin\theta d\theta d\varphi. \tag{16.44}$$

16.2 Hilbert Spaces

16.2.1 Cauchy Sequences and Separability

By definition an infinite-dimensional scalar product space $\vec{\mathbb{V}}^\infty$ admits an infinite number of linearly independent vectors $\vec{\varphi}_\ell$, $\ell = 1, 2, 3, \ldots$. The question is whether such a set can form a basis for the space. Before we can consider the notion of basis in $\vec{\mathbb{V}}^\infty$ we must know how to treat a linear combination of an infinite number of linearly independent vectors, i.e., we must know the meaning of the following infinite sum

$$\sum_{\ell=1}^{\infty} c_\ell \vec{\varphi}_\ell, \quad c_\ell \in \mathbb{C}. \tag{16.45}$$

[29] Integrations on the unit sphere is with respect to the volume element $\sin\theta d\theta d\varphi$. Similar comments on coordinate covering of a circle due to its topology apply to the sphere.

To understand the above infinite sum we first construct the following well-defined finite sums

$$\vec{\phi}_n = \sum_{\ell=1}^{n} c_\ell \, \vec{\varphi}_\ell, \quad \text{where } n \text{ is a positive integer.} \tag{16.46}$$

These finite sums generate an infinite sequence of vectors $\vec{\phi}_n$ with

$$\vec{\phi}_1 = c_1\vec{\varphi}_1, \quad \vec{\phi}_2 = c_1\vec{\varphi}_1 + c_2\vec{\varphi}_2, \quad \cdots. \tag{16.47}$$

The question is whether such a sequence would converge in some well-defined sense to another vector in $\vec{\mathbb{V}}^\infty$ as n tends to infinity. To gain an intuition let us consider the following three examples of infinite sequences of real numbers:

$$1, \ -1, \ 1, \ -1, \ \ldots, \ -(-1)^n, \ \ldots, \tag{16.48}$$

$$1, \ 2, \ 3, \ 4, \ \ldots, \ n, \ \ldots, \tag{16.49}$$

$$1, \ 1/2, \ 1/3, \ 1/4, \ \ldots, \ 1/n, \ \ldots. \tag{16.50}$$

The first sequence does not converge to any value. It oscillates all the way, even though the numbers of the sequence do not become larger and larger. The second sequence diverges to infinity, while the third sequence converges to 0. It is more complicated if complex numbers are involved. It is necessary to make clear the meaning of convergence. This is done as follows:

A sequence of possibly complex numbers c_n is said to converge to a complex number c, known as the limit of the sequence, as $n \to \infty$ if the absolute values $|\, c_n - c \,|$ of the complex numbers $c_n - c$ become vanishingly small as $n \to \infty$, i.e.,[30]

$$|\, c_n - c \,| \ \to \ 0 \ \ as \ \ n \ \to \ \infty. \tag{16.51}$$

There are many tests and criteria for convergence. The most basic one is the following

Cauchy Convergence Criterion[31] *A sequence of numbers c_n converges to a limit c if and only if for each small positive real number*

[30] The number c is meant to be finite, i.e., $|c| < \infty$. Otherwise the sequence is said to diverge.

[31] Spiegel (1) p. 43. Spiegel (2) p. 141. The numbers may be real or complex.

ϵ there exists an integer M such that

$$| c_m - c_n | < \epsilon \quad \text{for all} \quad m, n > M. \tag{16.52}$$

Intuitively we can appreciate that if the sequence converges to a limit c, then both c_m and c_n for large m and n must be very close to c. Hence their difference $| c_m - c_n |$ must be vanishingly small. A sequence $\{c_n\}$ of numbers satisfying the Cauchy convergence criterion is called a *Cauchy sequence*, and the above criterion tells us that every Cauchy sequence of numbers c_n converges. This criterion enables us to test whether a sequence converges or not.

We can generalise the above discussion to vectors. This is done by defining convergence in terms of the length of vectors. The length of a vector $\vec{\phi}$ is measured in terms of its norm, i.e., the length is given by its norm $||\vec{\phi}||$. Intuitively we can appreciate that a vector $\vec{\phi}$ is said to be close to another vector $\vec{\psi}$ if their difference $\vec{\phi} - \vec{\psi}$ has a small length, i.e., if $||\vec{\phi} - \vec{\psi}||$ is very small.

Definition 16.2.1(1)[32] *A sequence of vectors $\vec{\phi}_n$ in a scalar product space $\vec{\mathbb{V}}^\infty$ is said to converge to a vector $\vec{\phi}$ if*

$$||\vec{\phi}_n - \vec{\phi}|| \to 0 \ \text{ as } \ n \to \infty. \tag{16.53}$$

The vector $\vec{\phi}$ is called the limit of the sequence.

The relationship in Eq. (16.53) is often denoted by

$$\vec{\phi}_n \to \vec{\phi} \quad \text{or} \quad \lim_{n\to\infty} \vec{\phi}_n = \vec{\phi}. \tag{16.54}$$

We can now interpret the infinite linear combination in Eq. (16.45) as the limit of the sequence in Eq. (16.46), when the limit exists. The sum is undefined if the limit does not exist.

An obvious question presents itself: what is the criterion for the convergence of a sequence of vectors? Intuitively we wish the Cauchy criterion would apply. So, let us define Cauchy sequences.

[32] Roman Vol. 2 pp. 459–461. There are other concepts of convergence, e.g., we can have a concept of *weak convergence*. To emphasise the difference the convergence in Definition 16.2.1(1) is also called *strong convergence*.

Definition 16.2.1(2) *A sequence of vectors* $\vec{\phi}_1, \vec{\phi}_2, \vec{\phi}_3, \ldots$ *in a scalar product space* $\vec{\mathbb{V}}$ *is called a Cauchy sequence if for each small positive real number* ϵ *there exists an integer M such that*

$$||\vec{\phi}_m - \vec{\phi}_n|| < \epsilon \quad \text{for all } m, n > M. \tag{16.55}$$

While every Cauchy sequence of vectors in a finite-dimensional complex vector space converges to a limit vector in the space the same is not true in an arbitrary infinite-dimensional complex scalar product space.[33] Fortunately infinite-dimensional scalar product spaces of physical importance are such that all Cauchy sequences do converge. Such spaces are called *Hilbert spaces*.[34]

16.2.2 Hilbert spaces

Definition 16.2.2(1) *A scalar product space is said to be a Hilbert space if every Cauchy sequence of vectors in the space converges to a limit vector in the space.*

Definition 16.2.2(2)

(1) *A set linearly independent of vectors* $\vec{\varphi}_\ell$ *in a Hilbert space* $\vec{\mathcal{H}}$ *is said to be a countable basis for* $\vec{\mathcal{H}}$ *if every vector* $\vec{\phi}$ *in* $\vec{\mathcal{H}}$ *can be expressed as a linear combination of vectors in the set, i.e., if* [35]

$$\vec{\phi} = \sum_{\ell=1}^{\infty} c_\ell \, \vec{\varphi}_\ell, \quad c_\ell \in \mathbb{C}. \tag{16.56}$$

The vectors $\vec{\varphi}_\ell$ *are called basis vectors and the basis is said to be orthonormal if* $\vec{\varphi}_\ell$ *are normalised and mutually orthogonal.*

(2) *A Hilbert space is said to be separable if it admits a countable basis.*

The space $\vec{\ell}^{\,2}$ is an infinite-dimensional Hilbert space.[36] It is also separable as it admits a countable basis $\{\vec{e}^{\,c}_\ell\}$ given by Eq. (16.2). We

[33] Fano p. 271 and Roman Vol. 2 p. 439 for counter examples. The adjective "complex" would be omitted from now on.

[34] Hilbert (1862–1943) was a German mathematician, considered to be one of the most influential mathematician of early 20th century.

[35] The set $\{\vec{\varphi}_\ell\}$, called a *complete orthonormal set*, may be finite or infinite.

[36] Jordan pp. 9–10.

have the following expression for an arbitrary vector $\vec{\zeta}$ in $\vec{\ell}^{\,2}$:

$$\vec{\zeta} = \sum_{\ell=1}^{\infty} \zeta_\ell \vec{e}_\ell^{\,c}, \quad \zeta_\ell = \langle \vec{e}_\ell^{\,c} \mid \vec{\zeta} \rangle. \tag{16.57}$$

When an orthonormal basis $\{\vec{\varphi}_\ell\}$ exists we have, for any $\vec{\phi} \in \vec{\mathcal{H}}$,

$$\vec{\phi} = \sum_{\ell=1}^{\infty} c_\ell \vec{\varphi}_\ell, \quad c_\ell = \langle \vec{\varphi}_\ell \mid \vec{\phi} \rangle, \tag{16.58}$$

in the sense of Eq. (16.53) with

$$\vec{\phi}_n = \sum_{\ell=1}^{n} c_\ell \vec{\varphi}_\ell, \quad c_\ell = \langle \vec{\varphi}_\ell \mid \vec{\phi} \rangle. \tag{16.59}$$

An infinite-dimensional Hilbert space may not be separable.[37] Hilbert spaces relevant to quantum mechanics are all separable. From now on all Hilbert spaces are assumed to be separable.

Since every Cauchy sequence in a finite-dimensional scalar product space converges they are automatically Hilbert spaces. The situation in infinite-dimensional spaces is different. Take the example of the set $C(\Lambda)$ of continuous functions on the interval $\Lambda = [0, 2]$. Cauchy sequences in this space may not converge to members of $C(\Lambda)$. A sequence of continuous functions can easily converge to a discontinuous function which lies outside $C(\Lambda)$. We can illustrate this by an example. Let n_0 be a large positive integer and let $n = n_0, n_0 + 1, n_0 + 2, \ldots$ be a sequence of integers. Let $f_n(x)$ be a sequence of continuous real-valued functions defined on Λ which are zero for $x \in [0, 1 - 1/n]$ and rise linearly to reach the value 1 at $x = 1$ and then maintains the value 1 for $x \in (1, 2]$. We can write down an expression of $f_n(x)$ as follows[38]:

$$f_n(x) = \begin{cases} f_n(x) = 0, & x \in [0, 1 - 1/n] \\ f_n(x), & \text{rises linearly from 0 to 1} \\ & \text{for } x \in (1 - 1/n, 1) \\ f_n(x) = 1, & x \in [1, 2] \end{cases}. \tag{16.60}$$

[37] Fano p. 271 for a counter example.
[38] Fano p. 251.

The function is shown in the following diagram:

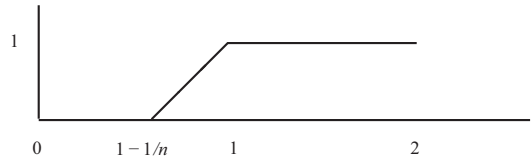


Figure 16.2 A function in $C(\Lambda)$.

The sequence $\{ f_n(x) \}$ is a Cauchy sequence. Intuitively we can see that this sequence of functions would converge to the step function

$$f_s(x) = \begin{cases} 0, & x \in [0, 1) \\ 1, & x \in [1, 2] \end{cases}, \tag{16.61}$$

which is discontinuous and lies outside $C(\Lambda)$. It follows that the corresponding vector space $\vec{C}(\Lambda)$ is not a Hilbert space.

To obtain a Hilbert space we must go to a large set of functions. The following examples of Hilbert spaces serve to illustrate the situation[39]:

E16.2.2(1) The space $\vec{L}^2(\Lambda)$ To enlarge $C(\Lambda)$ in order to arrive at a Hilbert space we need to include discontinuous functions so that all Cauchy sequences of functions within the larger set do converge within the set. The set $\mathcal{L}^2(\Lambda)$ in E16.1.2.6(3) also proves to be insufficiently large.[40] As pointed out in §4.2.1 and §4.2.2 Riemann integrable functions lack the desired convergence property. Lebesgue-integrable functions do possess the convergence property required for the formation of a Hilbert space. Hence, it is $\vec{L}^2(\Lambda)$ which constitutes a Hilbert space. The space also admits countable bases, e.g., we have the set of vectors $\vec{\varphi}_\ell$ in Eq. (12.29) forming an orthonormal basis in $\vec{L}^2(\Lambda)$. It follows $\vec{L}^2(\Lambda)$ is a separable Hilbert space.

E16.2.2(2) The space $\vec{L}^2(\mathcal{C}_a)$ This is the Hilbert space defined by the square-integrable functions defined on a circle $\mathcal{C}_a$ of radius $r = a$. The circle is coordinated by polar coordinates θ. The space $\vec{L}^2(\mathcal{C}_a)$

[39] These examples are described in §16.1.2.7, §16.1.2.8 and §16.1.2.9.
[40] $\mathcal{L}^2(\Lambda)$ and $L^2(\Lambda)$ are introduced in §16.1.2.

corresponds to functions of θ which are square-integrable with respect to the angular position variable θ. This space is separable, e.g., the following functions

$$\varphi_\ell(\theta) := \frac{1}{\sqrt{2\pi}} e^{i\ell\theta}, \quad \ell = 0, \pm 1, \pm 2, \ldots \tag{16.62}$$

define an orthonormal basis $\{\vec{\varphi}_\ell\}$ for $\vec{L}^2(\mathcal{C}_a)$.

E16.2.2(3) The space $\vec{L}^2(\mathcal{S}_u)$ This is the Hilbert space defined by the set $L^2(\mathcal{S}_u)$ of square-integrable functions defined on a unit sphere $\mathcal{S}_u$. This space is separable, e.g., there is an orthonormal basis defined by a well-known set of orthonormal functions known as the **spherical harmonics** which are functions of the angle variables θ and ϕ. Spherical harmonics are specified by two related indices ℓ and m_ℓ, i.e.,[41]

$$Y_{\ell,m_\ell}(\theta, \varphi), \quad \ell = 0, 1, 2 \cdots ; \; m_\ell = 0, \pm 1, \pm 2, \cdots . \tag{16.63}$$

The two indices ℓ, m_ℓ are related by $-\ell \leq m_\ell \leq \ell$. These functions are a product of an exponential function of φ and a trigonometric function of θ, i.e., they are of the form $\exp(i m_\ell \varphi) \times F_{\ell,m_\ell}(\sin\theta, \cos\theta)$ as shown by the following examples[42]:

$$Y_{0,0}(\theta, \varphi) := \frac{1}{\sqrt{4\pi}}, \tag{16.64}$$

$$Y_{1,-1}(\theta, \varphi) := \sqrt{\frac{3}{8\pi}}\, e^{-i\varphi} \sin\theta, \tag{16.65}$$

$$Y_{1,0}(\theta, \varphi) := \sqrt{\frac{3}{4\pi}} \cos\theta, \tag{16.66}$$

$$Y_{1,1}(\theta, \varphi) := -\sqrt{\frac{3}{8\pi}}\, e^{i\varphi} \sin\theta. \tag{16.67}$$

Using the integral for the scalar product in Eq. (16.44) one can verify that these functions are orthonormal. The vectors in $\vec{L}^2(\mathcal{S}_u)$ defined by these functions are denoted by $\vec{Y}_{\ell,m_\ell}$. We shall also refer to these vectors as spherical harmonics.

[41] Merzbacher p. 248. *See* §36.1.2.

[42] *See* Q36(1) and Zettili p. 293 more examples and for the expressions in Cartesian coordinates.

E16.2.2(4) The spaces $\vec{L}^2(I\!R)$ and $\vec{L}^2(I\!R^+)$ The set $\mathcal{L}^2(I\!R)$ in E16.1.2.6(1) is not big enough to lead to a Hilbert space. It is $\vec{L}^2(I\!R)$ which constitutes a Hilbert space. This space is also separable. A well-known example of the set of vectors $\vec{\varphi}_{Hn}$ defined by the Hermite functions $\varphi_{Hn}(x)$ in Eq. (16.18). This set of vectors is known to form a complete orthonormal set, i.e., an orthonormal basis, for $\vec{L}^2(I\!R)$.[43] The space $\vec{L}^2(I\!R^+)$, rather than $\mathcal{L}^2(I\!R^+)$ in E16.1.2.6(2), can also be shown to form a separable Hilbert space.

Infinite-dimensional Hilbert spaces possess many of the properties of finite-dimensional scalar product spaces. For example, the Gram-Schmit orthogonalisation procedure can be applied to obtain an orthonormal basis.[44] and Pythagoras theorem also applies.

For easy reference we shall summarise many of the previous results here. Let $\{\vec{\varphi}_\ell,\ \ell = 1, 2, 3, \ldots\}$ be an orthonormal basis in $\vec{\mathcal{H}}$, and let $\vec{\phi}$ and $\vec{\psi}$ be two arbitrary vectors for $\vec{\mathcal{H}}$. We have the following easily verified properties:

P16.2.2(1) Expansion in an orthonormal basis:

$$\vec{\phi} = \sum_{\ell=1}^{\infty} c_\ell \vec{\varphi}_\ell, \quad c_\ell = \langle \vec{\varphi}_\ell \mid \vec{\phi} \rangle, \tag{16.68}$$

$$\vec{\psi} = \sum_{\ell=1}^{\infty} d_\ell \vec{\varphi}_\ell, \quad d_\ell = \langle \vec{\varphi}_\ell \mid \vec{\psi} \rangle. \tag{16.69}$$

P16.2.2(2) No non-zero vector can be orthogonal to all $\vec{\varphi}_\ell$:

$$\langle \vec{\phi} \mid \vec{\varphi}_\ell \rangle = 0 \quad \forall\, \ell \ \Rightarrow\ \vec{\phi} = \vec{0}. \tag{16.70}$$

$$\langle \vec{\phi} \mid \vec{\varphi}_\ell \rangle = \langle \vec{\psi} \mid \vec{\varphi}_\ell \rangle\ \forall\, \ell \ \Rightarrow\ \vec{\phi} = \vec{\psi}. \tag{16.71}$$

P16.2.2(3) Pythagoras theorem:

$$||\vec{\phi}\,||^2 = \langle \vec{\phi} \mid \vec{\phi} \rangle = \sum_{\ell=1}^{\infty} c_\ell^* c_\ell = \sum_{\ell=1}^{\infty} \langle \vec{\phi} \mid \vec{\varphi}_\ell \rangle \langle \vec{\varphi}_\ell \mid \vec{\phi} \rangle. \tag{16.72}$$

$$||\vec{\phi}\,||^2 = 1 \quad \Rightarrow \quad \sum_{\ell=1}^{\infty} c_\ell^* c_\ell = \sum_{\ell=1}^{\infty} |c_\ell|^2 = 1. \tag{16.73}$$

[43] Takhtajan pp. 107–108. Merzbacher p. 88 for an intuitive proof.
[44] Prugovečki p. 22.

$$\langle \vec{\phi} \mid \vec{\psi} \rangle = \sum_{\ell} c_\ell^* d_\ell = \sum_{\ell=1}^{\infty} \langle \vec{\phi} \mid \vec{\varphi}_\ell \rangle \langle \vec{\varphi}_\ell \mid \vec{\psi} \rangle. \tag{16.74}$$

P16.2.2(4) Inequalities in Eqs. (6.38), (6.39) apply:

(1) The Schwarz Inequality

$$|\langle \vec{\phi} \mid \vec{\psi} \rangle| \leq ||\vec{\phi}|| \, ||\vec{\psi}||. \tag{16.75}$$

(2) Triangle Inequalities

$$||\vec{\phi} + \vec{\psi}|| \leq ||\vec{\phi}|| + ||\vec{\psi}||. \tag{16.76}$$

$$\big| \, ||\vec{\phi}|| - ||\vec{\psi}|| \, \big| \leq ||\vec{\phi} - \vec{\psi}||. \tag{16.77}$$

P16.2.2(5) The results on representing vectors by column vectors presented in §7.5 apply here. The column vectors would generally have an infinite number of elements.

P16.2.2(6) Theorem 8.1(1) of Riesz can be generalised to an infinite-dimensional Hilbert space $\vec{\mathcal{H}}$ by introducing the notion of continuity of linear functionals on $\vec{\mathcal{H}}$:

(1) A *linear functional* F on $\vec{\mathcal{H}}$ is a mapping of $\vec{\mathcal{H}}$ into $\mathbb{C}$

$$F : \vec{\mathcal{H}} \to \mathbb{C} \tag{16.78}$$

with the property

$$F(c_1\vec{\phi}_1 + c_2\vec{\phi}_2) := c_1 F(\vec{\phi}_1) + c_2 F(\vec{\phi}_2), \quad c_1, \ c_2 \in \mathbb{C}. \tag{16.79}$$

(2) A linear functional on $\vec{\mathcal{H}}$ is *continuous* if[45]

$$\vec{\phi}_n \to \vec{\phi} \quad \Rightarrow \quad F(\vec{\phi}_n) \to F(\vec{\phi}). \tag{16.80}$$

(3) A vector $\vec{\varphi} \in \vec{\mathcal{H}}$ generates a continuous linear functional on $\vec{\mathcal{H}}$ by

$$F_{\vec{\varphi}}(\vec{\phi}) := \langle \vec{\varphi} \mid \vec{\phi} \rangle. \tag{16.81}$$

We have the following theorem.[46]

[45] The convergence, written as $\vec{\phi}_n \to \vec{\phi}$ here and in Eq. (16.54), is defined by Definition 16.2.1(1).

[46] Jordan p. 13. Prugovečki p. 184.

Theorem 16.2.2 (1) Riesz Theorem *Every continuous linear functional on a Hilbert space is generated by a unique vector* $\vec{\varphi}$ *by*

$$F_{\vec{\varphi}}(\vec{\phi}) := \langle \vec{\varphi} \mid \vec{\phi} \rangle. \tag{16.82}$$

This theorem establishes a one-to-one correspondence between continuous linear functionals and vectors in a Hilbert space.

16.2.3 Subspaces, Projections and Projectors

The concepts of subspaces, projections and projectors presented in §6.4.4, §13.2.1 and §13.2.2 can be generalised here.

Definition 16.2.3(1)

(1) *A linear subset* $\vec{\mathcal{S}}$ *of* $\vec{\mathcal{H}}$ *is said to be a* **subspace** *of* $\vec{\mathcal{H}}$ *if every Cauchy sequence within* $\vec{\mathcal{S}}$ *converges to a vector in* $\vec{\mathcal{S}}$*.*

(2) *The orthogonal complement of a subspace* $\vec{\mathcal{S}}$*, denoted by* $\vec{\mathcal{S}}^{\perp}$*, is the set of all the vectors in* $\vec{\mathcal{H}}$ *which are orthogonal to every vector in* $\vec{\mathcal{S}}$*, i.e.,*

$$\vec{\mathcal{S}}^{\perp} = \left\{ \vec{\psi} \in \vec{\mathcal{H}} : \langle \vec{\psi} \mid \vec{\phi} \rangle = 0, \forall \vec{\phi} \in \vec{\mathcal{S}} \right\}. \tag{16.83}$$

To be a subspace the subset $\vec{\mathcal{S}}$ must be a Hilbert space in its own right. Not all linear subsets are subspaces. In $\vec{L}^2(\mathbb{R})$ the subset $\vec{S}_s(\mathbb{R})$ corresponding to the set $S_s(\mathbb{R})$ of Schwartz functions on $\mathbb{R}$ introduced in §16.1.2.4 is not a subspace. Despite this we still call $\vec{S}_s(\mathbb{R})$ a *Schwartz space*.[47] The orthogonal complement $\vec{\mathcal{S}}^{\perp}$ can be shown to be a Hilbert space in its own right also, i.e., $\vec{\mathcal{S}}^{\perp}$ also forms a subspace. The results in Eq. (6.59) for $\vec{\mathbb{E}}^3$ and Eq. (13.15) for $\vec{\mathbb{V}}^N$ remain true in $\vec{\mathcal{H}}$. Every vector $\vec{\phi}$ in $\vec{\mathcal{H}}$ can be decomposed uniquely as a sum of a vector $\vec{\phi}_{\vec{\mathcal{S}}}$ in the subspace $\vec{\mathcal{S}}$ and another vector $\vec{\phi}_{\vec{\mathcal{S}}^{\perp}}$ in its orthogonal complement $\vec{\mathcal{S}}^{\perp}$, i.e.,

$$\vec{\phi} = \vec{\phi}_{\vec{\mathcal{S}}} + \vec{\phi}_{\vec{\mathcal{S}}^{\perp}}, \quad \text{where} \quad \vec{\phi}_{\vec{\mathcal{S}}} \in \vec{\mathcal{S}}, \quad \vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{S}}^{\perp}. \tag{16.84}$$

[47] As in Eq. (16.30) this relationship is conveniently denoted by $\vec{S}_s(\mathbb{R}) := S_s(\mathbb{R})$. The set of vectors corresponding to the set $S_s(\mathbb{R}^3)$ of Schwartz functions on $\mathbb{R}^3$ is denoted by $\vec{S}_s(\mathbb{R}^3)$

The vectors $\vec{\phi}_{\vec{\mathcal{S}}}$ and $\vec{\phi}_{\vec{\mathcal{S}}^\perp}$ are defined to be the *projections* of $\vec{\phi}$ onto $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^\perp$, respectively. The projection of any vector onto any subspace can be described in terms of projectors as in §13.2.2, i.e., the decomposition of a vector $\vec{\phi}$ into projections onto a subspace $\vec{\mathcal{S}}$ and its orthogonal complement $\vec{\mathcal{S}}^\perp$ in Eq. (16.84) may be regarded as effected by two operators $\widehat{P}_{\vec{\mathcal{S}}}$ and $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ defined on every $\vec{\phi} \in \vec{\mathcal{H}}$ by

$$\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\phi} = \vec{\phi}_{\vec{\mathcal{S}}} \quad \text{and} \quad \widehat{P}_{\vec{\mathcal{S}}^\perp}\,\vec{\phi} = \vec{\phi}_{\vec{\mathcal{S}}^\perp}. \tag{16.85}$$

Definition 16.2.3(2) *The operators* $\widehat{P}_{\vec{\mathcal{S}}}$ *and* $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ *in* Eq. (16.85) *are called the* **projectors** *onto* $\vec{\mathcal{S}}$ *and* $\vec{\mathcal{S}}^\perp$*, respectively.*

If $\{\vec{\varphi}_j\}$ is an orthonormal basis for $\vec{\mathcal{S}}$, then, following Eq. (13.19), we have

$$\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\phi} = \sum_j \langle \vec{\varphi}_j \mid \vec{\phi} \rangle\, \vec{\varphi}_j. \tag{16.86}$$

This serves as a defining expression for $\widehat{P}_{\vec{\mathcal{S}}}$. A projector is *n*-dimensional if the subspace onto which it projects is *n*-dimensional. Here *n* can be infinite.[48]

If $\vec{\mathcal{S}}$ is spanned by a single unit vector $\vec{\varphi}$, then $\vec{\mathcal{S}}$ is one-dimensional. The associated projector is also called *one-dimensional*. In keeping with earlier notation we often express the projector as $\widehat{P}_{\vec{\varphi}}$, or $|\vec{\varphi}\rangle\langle\vec{\varphi}|$, and call $\widehat{P}_{\vec{\varphi}}$ the projector *generated* by the unit vector $\vec{\varphi}$. Equation (16.86) can be written as

$$\widehat{P}_{\vec{\mathcal{S}}} = \sum_j \widehat{P}_{\vec{\varphi}_j} = \sum_j |\vec{\varphi}_j\rangle\langle\vec{\varphi}_j|. \tag{16.87}$$

Definition 13.2.2(1) *on projections, projectors and their order relation and complete orthogonal families of projectors and subspaces applies*. For example, for a given complete orthonormal set of vectors $\vec{\varphi}_\ell$ in $\vec{\mathcal{H}}$ we have a corresponding **complete orthogonal family of projectors** $\widehat{P}_{\vec{\varphi}_\ell} = |\vec{\varphi}_\ell\rangle\langle\vec{\varphi}_\ell|$. Projectors on a Hilbert space are idempotent and selfadjoint in the sense of Eq. (13.24), i.e.,

$$\widehat{P}^2_{\vec{\mathcal{S}}} = \widehat{P}_{\vec{\mathcal{S}}} \quad \text{and} \quad \langle \vec{\phi} \mid \widehat{P}_{\vec{\mathcal{S}}}\,\vec{\psi} \rangle = \langle \widehat{P}_{\vec{\mathcal{S}}}\,\vec{\phi} \mid \vec{\psi} \rangle \ \ \forall \vec{\phi},\ \vec{\psi} \in \vec{\mathcal{H}}. \tag{16.88}$$

[48] *See* Definition 13.2.2(1).

Theorem 13.2.2(1) also remains valid, e.g., an operator is a projector if and only if it is selfadjoint and idempotent, and we also have

$$\widehat{P}_{\vec{\mathcal{S}}} + \widehat{P}_{\vec{\mathcal{S}}^{\perp}} = \widehat{I\!I}. \tag{16.89}$$

Exercises and Problems

Q16(1) Prove the inequalities in Eq. (16.4).

Q16(2) Verify Eqs. (16.21), (16.22) and (16.23).

Q16(3) Prove the equality

$$||\vec{\varphi} + \vec{\phi}\,||^2 + ||\vec{\varphi} - \vec{\phi}\,||^2 = 2||\vec{\varphi}\,||^2 + 2||\vec{\phi}\,||^2. \tag{16.90}$$

Q16(4) Prove the Schwarz inequality in Eq. (16.75).

Q16(5) Prove the triangle inequalities in Eqs. (16.76) and (16.77).

Q16(6) Show that $\phi(x),\ \psi(x) \in \mathcal{L}^2(I\!R) \Rightarrow \phi(x) + \psi(x) \in \mathcal{L}^2(I\!R)$ and that the integral on the right hand side of Eq. (16.28) is finite for functions in $\mathcal{L}^2(I\!R)$.

Q16(7) Show that the sequence of vectors $\vec{f}_n$ defined by functions $f_n(x)$ in Eq. (16.60) is a Cauchy sequence in the space $\vec{\mathcal{C}}(\Lambda)$.[49]

Q16(8) Express the spherical harmonics in Eqs. (16.64) to (16.67) in the Cartesian coordinates x, y, z which are related to the spherical coordinates r, θ, φ by Eq. (16.41).

Q16(9) Show that Eq. (16.81) defines a continuous linear functional on $\vec{\mathcal{H}}$.

Q16(10) In the Dirac notation a scalar product is denoted by $\langle \vec{\varphi} \mid \vec{\phi}\,\rangle$. Dirac formally consider the notation as the product of two quantities: (1) $\langle \vec{\varphi}\,|$ called a *bra* and (2) $|\vec{\phi}\,\rangle$ called a *ket*. Their product forms a *bracket* $\langle \vec{\varphi} \mid \vec{\phi}\,\rangle$ which is the scalar product. Explain how we can interpret *bras* and *kets* in terms of vectors and linear functionals.[50]

Q16(11) Explain the concept of separability in a Hilbert space.

[49] Fano p. 251.
[50] Jauch p. 32.

Chapter 17

Operators in a Hilbert space $\vec{\mathcal{H}}$

17.1 Boundedness, Continuity and Closedness

Operators have already been introduced in finite-dimensional spaces. Many of the definitions and properties of operators introduced in Chapters 8, 9 and 13 remain applicable in a Hilbert space. However, the infinite-dimensional nature of a general Hilbert space gives rise to some serious complications. An example is the non-existence of eigenvalues and eigenvectors for some operators. Another is the emergence of *unbounded operators*. The idea of unbounded operators are similar to unbounded functions. As examples let us consider three different types of functions:

E17.1(1) The function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = \sin x$ on the real line $\mathbb{R}$ is *bounded* in that the values of the function are less than or equal to a certain fixed number, i.e., $\sin x \leq 1$ for all $x \in \mathbb{R}$.

E17.1(2) The function $f(x) = \exp x^2$ defined on $\mathbb{R}$ has a well-defined value for all $x \in \mathbb{R}$ but these values are not bounded by a fixed number. The function is then said to be *unbounded*.

E17.1(3) The function $f(x) = 1/x$ is also unbounded. The function becomes arbitrarily large as $x \rightarrow 0$. An additional problem

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

is that the function is not defined on the entire real line, i.e., it is undefined for $x = 0$ since it gives a formal value of infinity for $x = 0$.[1]

Similar cases arise for operators in a Hilbert space. There are operators which are not definable on the entire Hilbert space, like the function $f(x) = 1/x$ is not defined on the entire real line $I\!R$. We have to consider operators $\widehat{A}$ defined on a domain $\vec{\mathcal{D}}(\widehat{A})$ which is smaller than $\vec{\mathcal{H}}$. Its range $\vec{\mathcal{R}}(\widehat{A})$ can also be smaller than $\vec{\mathcal{H}}$, i.e.,

$$\widehat{A} : \vec{\mathcal{D}}(\widehat{A}) \rightarrow \vec{\mathcal{R}}(\widehat{A}) \tag{17.1}$$

where

$$\vec{\mathcal{D}}(\widehat{A}) \subset \vec{\mathcal{H}}, \quad \vec{\mathcal{R}}(\widehat{A}) \subset \vec{\mathcal{H}}. \tag{17.2}$$

From now on we shall confine ourselves to operators defined on a domain which is a linear subset of $\mathcal{H}$ with the property

$$\widehat{A}\left(c_1\vec{\phi}_1 + c_2\vec{\phi}_2\right) = c_1\widehat{A}\vec{\phi}_1 + c_2\widehat{A}\vec{\phi}_2 \tag{17.3}$$

for all $\vec{\phi}_1, \vec{\phi}_2 \in \vec{\mathcal{D}}(\widehat{A})$ and $c_1, c_2 \in \mathbb{C}$. These operators are said to be *linear*. The **quadratic form** $\mathcal{Q}(\widehat{A}, \vec{\phi})$ generated by $\widehat{A}$ on $\vec{\mathcal{H}}$ is defined to be

$$\mathcal{Q}(\widehat{A}, \vec{\phi}) := \langle \vec{\phi} \mid \widehat{A}\,\vec{\phi} \rangle, \quad \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.4}$$

An operator cannot act on vectors lying outside its domain. For example, Eq. (17.3) is undefined if $\vec{\phi}_1, \vec{\phi}_2 \notin \vec{\mathcal{D}}(\widehat{A})$ and Eq. (17.4) is also undefined if $\vec{\phi} \notin \vec{\mathcal{D}}(\widehat{A})$.

For physical applications we require the domain $\vec{\mathcal{D}}(\widehat{A})$ to be sufficiently large, i.e., $\vec{\mathcal{D}}(\widehat{A})$ should be almost as large as $\vec{\mathcal{H}}$ itself in that there are sequences of vectors in $\mathcal{D}(\widehat{A})$ converging to every $\vec{\phi} \in \vec{\mathcal{H}}$. This requirement is satisfied if $\mathcal{D}(\widehat{A})$ contains an orthonormal basis $\{\vec{\varphi}_\ell\}$ of the Hilbert space $\vec{\mathcal{H}}$ so that for each $\vec{\phi} \in \vec{\mathcal{H}}$ we can, following Eqs. (16.46) and (16.56), construct a sequence $\vec{\phi}_n$ which would approach $\vec{\phi}$ as a limit. Then finite linear combinations of $\vec{\varphi}_\ell$ can approximate any vector with arbitrary accuracy. Such a domain is said to be *dense*.

[1] Infinity is not a number in $I\!R = (-\infty, \infty)$. An infinite value is not a well-defined value.

Definition 17.1(1) *A subset D of $\vec{\mathcal{H}}$ is said to be* **dense** *if there is a sequence of vectors $\vec{\phi}_n$ in $\vec{\mathcal{D}}$ converging to every vector $\vec{\phi}$ in $\vec{\mathcal{H}}$.*

A Hilbert space admits many different dense subsets, and a dense subset itself may also contain many subsets which are also dense. Formally $\vec{\mathcal{H}}$ is a dense subset of $\vec{\mathcal{H}}$. For the spaces introduced in §16.1.2 we can show that $\vec{C}_c^\infty(\Lambda)$ and $\vec{AC}(\Lambda)$ are both dense in $\vec{L}^2(\Lambda)$, and $\vec{C}_c^\infty(I\!R)$ and $\vec{AC}(I\!R)$ are both dense in $\vec{L}^2(I\!R)$.

From now on all operators are meant to be linear and defined on a dense domain.

Definition 17.1(2) *An operator $\widehat{A}$ is said to be* **bounded** *if the set of values*

$$\left\{ ||\widehat{A}\vec{\phi}|| \;\; \forall\, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}),\;\; ||\vec{\phi}|| = 1 \right\} \tag{17.5}$$

is bounded from above. The supremum of the set of values is defined to be the norm of the operator which is denoted by $||\widehat{A}||$, i.e.,[2]

$$||\widehat{A}|| := \sup \left\{ ||\widehat{A}\vec{\phi}|| \;\; \forall\, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}),\;\; ||\vec{\phi}|| = 1 \right\}. \tag{17.6}$$

The operator is said to be **unbounded** *if it is not bounded.*

For bounded operators Eq. (8.15) applies, i.e., we have

$$||\widehat{A}\vec{\phi}|| \leq ||\widehat{A}||\,||\vec{\phi}||. \tag{17.7}$$

The norm $||\widehat{A}||$ of an operator $\widehat{A}$ can be considered to be a measure of the "size" of the operator. It is the biggest norm of the output vectors $\vec{\phi}' = \widehat{A}\vec{\phi}$ can reach for all unit input vectors $\vec{\phi}$ in the domain of the operator. The identity operator is a trivial example of bounded operator with unit norm. Projectors in §16.2.3 are bounded operators also of unit norm. The set of bounded operators in a Hilbert space $\vec{\mathcal{H}}$ will be denoted by $\widehat{\mathcal{B}}(\vec{\mathcal{H}})$.

The domain of a bounded operator can be uniquely extended to the entire Hilbert space.[3] So, we shall assume from now on that bounded operators are defined on the entire Hilbert space. All

[2] *See* Eq. (8.13) for the norm of operators on $\vec{I\!E}^3$.
[3] Roman Vol. 2 p. 507.

operators defined on a finite-dimensional scalar product space such as $\vec{I\!E}^3$ and $\vec{I\!R}^N$ are bounded.[4]

Unbounded operators are generally defined on a domain smaller than the entire Hilbert space.[5] This is because for an arbitrary input vector the operator may produce an output vector with an infinite norm. Such an output vector is not a member of the Hilbert space. An operator can only act on input vectors such that the output vectors have a finite norm, i.e., we require

$$\vec{\mathcal{D}}(\widehat{A}) := \left\{ \vec{\phi} \in \mathcal{H} : \widehat{A}\vec{\phi} \in \vec{\mathcal{H}} \right\}. \tag{17.8}$$

Definition 17.1(3) *Let $\widehat{A}$ be an operator in a Hilbert space $\mathcal{H}$ and let $\{\vec{\varphi}_\ell\}$ be an arbitrary sequence of vectors in the domain $\vec{\mathcal{D}}(\widehat{A})$ of $\widehat{A}$. Then*[6]

(1) *The operator $\widehat{A}$ is said to be* **continuous** *if whenever $\{\vec{\varphi}_\ell\}$ converges to the vector $\vec{\varphi}$ in $\vec{\mathcal{D}}(\widehat{A})$ the sequence $\{\widehat{A}\vec{\varphi}_\ell\}$ converges to the vector $\widehat{A}\vec{\varphi}$ in $\vec{\mathcal{H}}$, i.e.,*

$$\vec{\varphi}_\ell \in \vec{\mathcal{D}}(\widehat{A}) \;\to\; \vec{\varphi} \in \vec{\mathcal{D}}(\widehat{A}) \;\Rightarrow\; \widehat{A}\vec{\varphi}_\ell \;\to\; \widehat{A}\vec{\varphi} \in \vec{\mathcal{H}}. \tag{17.9}$$

(2) *$\widehat{A}$ is said to be* **closed** *if when the sequence $\{\vec{\varphi}_\ell\}$ converges to a vector $\vec{\varphi} \in \vec{\mathcal{H}}$ and the sequence $\{\widehat{A}\vec{\varphi}_\ell\}$ converges to a vector $\vec{\phi}$ in $\vec{\mathcal{H}}$ then the vector $\vec{\varphi}$ is in the domain of $\widehat{A}$ and $\widehat{A}\vec{\varphi} = \vec{\phi}$, i.e.,*

$$\vec{\varphi}_\ell \;\to\; \vec{\varphi} \in \vec{\mathcal{H}}, \quad \widehat{A}\vec{\varphi}_\ell \;\to\; \vec{\phi} \;\Rightarrow\; \vec{\varphi} \in \vec{\mathcal{D}}(\widehat{A}), \quad \widehat{A}\vec{\varphi} = \vec{\phi}.$$

We know that a real-valued function $f(\tau)$ on $I\!R$ is said to be continuous if whenever a sequence of numbers $\tau_\ell \in I\!R$ converges to a number $\tau \in I\!R$ the values $f(\tau_\ell)$ of the function converges to the value $f(\tau)$. The concept of continuity of operators is really the same. Continuous operators are related to boundedness of the operators as seen in the following theorem:

---

[4] Halmos p. 177.

[5] There are exceptions. An unbounded function, e.g., an unbounded exponential function e^x is defined on the entire real line $I\!R$. *See* Roman Vol. 2 p. 526.

[6] Generally an unbounded operator is said to be defined *in* $\vec{\mathcal{H}}$ while a bounded operator is defined *on* $\vec{\mathcal{H}}$.

Theorem 17.1(1) *An operator is continuous if and only if it is bounded.*[7]

The notion of closedness is different. It deduces that the vector $\vec{\varphi}$ should lie in the domain $\vec{\mathcal{D}}(\widehat{A})$ from the convergence of the sequences $\{\vec{\varphi}_\ell\}$ and $\{\widehat{A}\vec{\varphi}_\ell\}$. The above theorem tells us that unbounded operators are not continuous. However, many unbounded operators are closed. Closedness is required to establish many useful results, e.g., Theorem 19.1 (1).

In the next two sections we shall examine some explicit examples of multiplication and differential operators in Hilbert spaces of square-integrable functions before proceeding with further general definitions.

17.2 Multiplication Operators

Consider following examples of multiplication operators in $\vec{L}^2(I\!R)$:

E17.2(1) Bounded multiplication operators Let $V(x)$ be a real-valued bounded function of $x \in I\!R$, not necessarily continuous. Then $V(x)$ defines an operator $\widehat{V}$ on $\vec{L}^2(I\!R)$ by

$$\widehat{V}\vec{\phi} := V(x)\phi(x) \quad \forall \phi(x) \in L^2(I\!R). \tag{17.10}$$

The function $V(x)\phi(x)$ is square-integrable whenever $\phi(x)$ is square-integrable and is hence in $L^2(I\!R)$. It follows that the operator $\widehat{V}$ is bounded.[8] Two important examples are

E17.2(1)(a) The characteristic function $\chi_\Lambda(x)$ in Eq. (4.3) has the value 1 for x in the interval Λ and the value 0 for x outside Λ. As a multiplication operator, denoted by $\widehat{\chi}_\Lambda$, it can act on every vector $\vec{\phi}$ in $\vec{L}^2(I\!R)$ by

$$\widehat{\chi}_\Lambda\vec{\phi} := \chi_\Lambda(x)\phi(x). \tag{17.11}$$

The function $\chi_\Lambda(x)\phi(x)$ is square-integrable with a norm less than that of $\vec{\phi}$. Hence the operator is bounded.

[7] Jordan p. 17.
[8] Roman Vol. 2 p. 533

E17.2(1)(b) The step function introduced in Eq. (4.9) also defines a bounded operator on $\vec{L}^2(I\!R)$.

E17.2(2) Unbounded multiplication operators If the function $V(x)$ is unbounded the resulting operator $\widehat{V}(x)$ is unbounded. The most familiar example is $V(x) = x$ on $I\!R$. The corresponding operator $\widehat{x}(I\!R)$ acts on $\vec{\phi}$ in $\vec{L}^2(I\!R)$ by

$$\widehat{x}(I\!R)\vec{\phi} := x\phi(x). \tag{17.12}$$

The square-integrability of $\phi(x)$ does not imply that $x\phi(x)$ is square-integrable. It follows that $\widehat{x}(I\!R)$ cannot act on every vector in $\vec{L}^2(I\!R)$. To define the operator $\widehat{x}(I\!R)$ properly we must specify its domain. In accordance with Eq. (17.8) the multiplication operator $\widehat{x}(I\!R)$ is defined on the domain[9]

$$\vec{\mathcal{D}}\big(\widehat{x}(I\!R)\big) := \left\{\vec{\phi} \in \vec{L}^2(I\!R) : \widehat{x}(I\!R)\vec{\phi} \in \vec{L}^2(I\!R)\right\}. \tag{17.13}$$

The operator is unbounded. It is the **position operator** in quantum mechanics for a particle in one-dimensional motion along the x-axis. The domain can be more explicitly written in terms of functions in $L^2(I\!R)$ as

$$\vec{\mathcal{D}}\big(\widehat{x}(I\!R)\big) := \left\{\phi(x) \in L^2(I\!R) : x\phi(x) \in L^2(I\!R)\right\}. \tag{17.14}$$

We shall adopt such an explicit expression from now on.

The square of $\widehat{x}(I\!R)$ is defined on a smaller domain, i.e.,

$$\widehat{x}^{\,2}(I\!R)\vec{\phi} := x^2\phi(x). \tag{17.15}$$

acting on

$$\vec{\mathcal{D}}\big(\widehat{x}^{\,2}(I\!R)\big) := \left\{\phi(x) \in L^2(I\!R) : x^2\phi(x) \in L^2(I\!R)\right\}. \tag{17.16}$$

These domains, i.e., $\mathcal{D}(\widehat{x}(I\!R))$ and $\mathcal{D}(\widehat{x}^{\,2}(I\!R))$, can be shown to be dense in in $L^2(I\!R)$.

[9] The condition $\widehat{x}(I\!R)\vec{\phi} \in \vec{L}^2(I\!R)$ requires the functions $\phi(x)$ to be such that $x\phi(x)$ is square-integrable over $I\!R$.

We can extend our discussion to the Hilbert space $\vec{L}^2(I\!R^2)$ where we have two obvious position operators, i.e.,

$$\widehat{x}(I\!R^2)\vec{\phi} := x\phi(x, y), \quad \widehat{y}(I\!R^2)\vec{\phi} := y\phi(x, y) \tag{17.17}$$

acting, respectively on

$$\vec{\mathcal{D}}\big(\widehat{x}(I\!R^2)\big) := \big\{\phi(x, y) \in L^2(I\!R^2) : x\phi(x, y) \in L^2(I\!R^2)\big\}, \tag{17.18}$$

$$\vec{\mathcal{D}}\big(\widehat{y}(I\!R^2)\big) := \big\{\phi(x, y) \in L^2(I\!R^2) : y\phi(x, y) \in L^2(I\!R^2)\big\}. \tag{17.19}$$

In $\vec{L}^2(I\!R^3)$ we have three obvious position operators, e.g.,

$$\widehat{x}(I\!R^3)\vec{\phi} := x\phi(x, y, z), \tag{17.20}$$

acting on

$$\vec{\mathcal{D}}\big(\widehat{x}(I\!R^3)\big) := \big\{\phi(x, y, z) \in L^2(I\!R^3) : x\phi(x, y, z) \in L^2(I\!R^3)\big\}, \tag{17.21}$$

The other two position operators $\widehat{y}(I\!R^3)$ and $\widehat{z}(I\!R^3)$ are similarly defined.

It must be emphasised that an operator expression alone does not define an operator. The operator expression does not even determine the boundedness or otherwise of an operator. One must specify the Hilbert space and the domain the operator expression acts on. For example, the operator expression x acting on the Hilbert space $\vec{L}^2(\Lambda)$ produces a bounded operator with the entire $\vec{L}^2(\Lambda)$ as its domain, i.e., we have

$$\widehat{x}(\Lambda)\vec{\phi} := x\phi(x), \quad \vec{\mathcal{D}}\big(\widehat{x}(\Lambda)\big) = \vec{L}^2(\Lambda) \tag{17.22}$$

since $x\phi(x)$ is square integrable over the finite interval Λ. Similarly in $\vec{L}^2(\mathcal{C}_a)$ the multiplication operator $\widehat{\theta}(\mathcal{C}_a)$ is defined on the entire Hilbert space, i.e., we have

$$\widehat{\theta}(\mathcal{C}_a)\vec{\phi} := \theta\phi(\theta), \quad \mathcal{D}\big(\widehat{\theta}(\mathcal{C}_a)\big) = \vec{L}^2(\mathcal{C}_a). \tag{17.23}$$

17.3 Differential Operators and Boundary Conditions

17.3.1 Introduction

A differential operator is defined by a differential expression together with a domain of differentiable functions in a given Hilbert space. As an example let us try to define an operator with the differential expression d/dx in the Hilbert space $\vec{L}^2(\Lambda)$, $\Lambda = [0, L]$. Clearly a differential expression cannot operate on every vector in $\vec{L}^2(\Lambda)$. When acting on an input vector $\vec{\phi}$ defined by a function $\phi(x) \in L^2(\Lambda)$ we get an out vector defined by the derivative $d\phi(x)/dx$. If $\phi(x)$ is discontinuous its derivative has an infinite value, i.e., it is undefined, at the discontinuities, and hence the derivative is not square-integrable on $I\!R$. In other words, the derivative $d\phi(x)/dx$ cannot define a vector in $\vec{L}^2(\Lambda)$, violating the condition in Eq. (17.8). Hence the differential expression d/dx cannot act on discontinuous functions. Generally an operator resulting from a differential expression would also be unbounded. To define an operator with a differential expression in $\vec{L}^2(\Lambda)$ we must specify its domain explicitly. There may be other conditions imposed in addition to differentiability of the functions in the domain. In the case of operators in $\vec{L}^2(\Lambda)$ a domain is often specified by additional conditions, known as *boundary conditions*, imposed on the functions at the boundaries of the interval Λ.

A function $\phi(x)$ in $L^2(\Lambda)$ is said to satisfy[10]:

(1) The **Dirichlet boundary condition**[11] if

$$\phi(0) = \phi(L) = 0. \tag{17.24}$$

(2) The **quasi-periodic boundary condition** if[12]

$$\phi(0) = e^{i\lambda}\phi(L), \quad \lambda \in (-\pi, \pi]. \tag{17.25}$$

[10] Wan p. 284. There are other possible boundary conditions.
[11] Dirichlet (1805–1859) was a German mathematician.
[12] Any extension of the values of λ outside the interval $(-\pi, \pi]$ would lead to a repetition of the boundary condition.

(3) The **periodic boundary condition** if

$$\phi(0) = \phi(L). \tag{17.26}$$

Boundary conditions are motivated by the desire to obtain some specific type of operators. We shall list some of these operators in the next subsection. In §19.3 we shall discuss how boundary conditions help to obtain selfadjoint differential operators.

17.3.2 Specific operators

We shall introduce a number of operators here. Their properties will be studied later in §19.2 and §19.3. Many of these operators will be directly relevant to the formulation of quantum mechanics.

17.3.2.1 Differential operators in $\vec{L}^2(\Lambda),\ \Lambda = [0, L]$

We wish to define an operator in $\vec{L}^2(\Lambda)$ with the differential expression $-i\hbar d/dx$.[13] We need to specify a domain for $-i\hbar d/dx$ to act on. Different domains will result in different operators. Generally we would start with a small but dense subset to serve as a domain. Different operators are produced by enlarging or modifying the domain in different ways as seen in the following examples:

E17.3.2.1(1) Smooth functions Consider the set $C_D^\infty(\Lambda)$ of smooth functions on Λ introduced in E16.1.2.5(1) which vanish at the boundaries. We can define an operator $\widehat{p}_{\vec{C}_D^\infty}(\Lambda)$ on the domain

$$\vec{\mathcal{D}}\big(\widehat{p}_{\vec{C}_D^\infty}(\Lambda)\big) := \vec{C}_D^\infty(\Lambda) \tag{17.27}$$

by

$$\widehat{p}_{\vec{C}_D^\infty}(\Lambda)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx} \quad \forall\, \phi(x) \in C_D^\infty(\Lambda). \tag{17.28}$$

The output vector $\widehat{p}_{\vec{C}_D^\infty}(\Lambda)\vec{\phi}$ is also differentiable and is hence square-integrable over the interval Λ, i.e., $\widehat{p}_{\vec{C}_D^\infty}(\Lambda)\vec{\phi} \in \vec{L}^2(\Lambda)$.

[13] The numerical factor $-i\hbar$ does not affect the domain. The imaginary unit i is included in order to produce selfadjoint operators with real eigenvalues, as illustrated by E19.3.1(1) and E19.3.1(2) .

E17.3.2.1(2) Absolutely continuous function Since the operator involves only a first order differentiation we can enlarge the domain to once-differentiable functions. The biggest set of such functions is the set $AC(\Lambda)$ of absolutely continuous functions on Λ introduced in E16.1.2.2(1). These functions are differentiable almost everywhere in Λ. However, the derivative of an arbitrary function in $AC(\Lambda)$ may not be continuous and bounded, and hence it may not be square-integrable over the interval Λ. It follows that we have to settle for a subset of $AC(\Lambda)$ consisting of functions such that their derivatives are square integrable, i.e., we have an operator $\widehat{p}_{\vec{AC}}(\Lambda)$ acting on the domain[14]

$$\vec{\mathcal{D}}\big(\widehat{p}_{\vec{AC}}(\Lambda)\big) := \Big\{\phi(x) \in AC(\Lambda):\ d\phi/dx \in L^2(\Lambda)\Big\} \tag{17.29}$$

by

$$\widehat{p}_{\vec{AC}}(\Lambda)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx}. \tag{17.30}$$

This domain turns out to be too big for many applications. Further conditions need to be imposed in order to produce useful operators. For example, we can impose various boundary conditions. Each boundary condition would lead to a distinct operator as shown in the following examples:

E17.3.2.1(2)(a) The Dirichlet boundary condition[15] We can define an operator $\widehat{p}_D(\Lambda)$ acting on the domain $\vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$ consisting of vectors corresponding to functions in $\phi(x) \in AC(\Lambda)$ which also satisfy

$$\phi(0) = \phi(L) = 0 \quad \text{and} \quad \frac{d\phi(x)}{dx} \in L^2(\Lambda) \tag{17.31}$$

by

$$\widehat{p}_D(\Lambda)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx}. \tag{17.32}$$

[14] Fano pp. 279–280. Akhiezer and Glazmann Vol. 1 p. 106. Wan p. 115.

[15] Akhiezer and Glazmann Vol. 1 p. 106. Reed and Simon Vol. 1 p. 254, pp. 257–259. The subscript D indicates the imposition of the Dirichlet boundary condition in Eq. (17.24).

E17.3.2.1(2)(b) Quasi-periodic boundary condition[16] We can define a family of differential operators $\widehat{p}_\lambda(\Lambda)$, one for each value of $\lambda \in (-\pi, \pi]$, with an operator expression $-i\hbar d/dx$ acting on the domain $\vec{\mathcal{D}}\big(\widehat{p}_\lambda(\Lambda)\big)$ consisting of vectors corresponding to functions in $AC(\Lambda)$ which also satisfy

$$\phi(0) = e^{i\lambda}\phi(L) \quad \text{and} \quad \frac{d\phi(x)}{dx} \in L^2(\Lambda) \tag{17.33}$$

by

$$\widehat{p}_\lambda(\Lambda)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx}. \tag{17.34}$$

E17.3.2.1(2)(c) Periodic boundary condition[17] This is a special case of the quasi-periodic boundary condition when we have $\lambda = 0$ in Eq. (17.33). Explicitly we have an operator $\widehat{p}_{\lambda=0}(\Lambda)$ acting on the domain $\vec{\mathcal{D}}\big(\widehat{p}_{\lambda=0}(\Lambda)\big)$ consisting of vectors corresponding to functions in $AC(\Lambda)$ which also satisfy

$$\phi(0) = \phi(L) \quad \text{and} \quad \frac{d\phi(x)}{dx} \in L^2(\Lambda) \tag{17.35}$$

by

$$\widehat{p}_{\lambda=0}(\Lambda)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx}. \tag{17.36}$$

The operators $\widehat{p}_\lambda(\Lambda)$ and $\widehat{p}_{\lambda=0}(\Lambda)$ are different from $\widehat{p}_D(\Lambda)$. Note that the notation for the operators, e.g., $\widehat{p}_{\lambda=0}(\Lambda)$, incorporates the domain automatically. The notation also shows that different operators are generated by the same differential expression acting on different domains. The fact that these operators possess different properties will be demonstrated in §19.2 and §19.3.

17.3.2.2 Differential operators in $\vec{L}^2(\mathcal{C}_a)$

E17.3.2.2(1) Functions on $\mathcal{C}_a$ such as the set of absolutely continuous functions $AC(\mathcal{C}_a)$ are functions of the angular position variable θ introduced in §16.1.2.8. We can define a differential

[16] Akhiezer and Glazmann Vol. 1 p. 109. The subscript λ indicates the imposition of the quasi-periodic boundary condition in Eq. (17.25).
[17] Fano p. 282.

operator with an operator expression $-i\hbar d/ds = -i\hbar d/ad\theta$ acting on functions on the circle satisfying the periodic boundary condition shown in Eq. (16.39), i.e., we have an operator $\widehat{p}(\mathcal{C}_a)$ defined on the domain $\vec{\mathcal{D}}\big(\widehat{p}(\mathcal{C}_a)\big)$ consisting of vectors corresponding to functions $\phi(\theta)$ in $AC(\mathcal{C}_a)$ which also satisfy[18]

$$\phi(0) = \phi(2\pi) \quad \text{and} \quad \frac{d\phi(\theta)}{d\theta} \in L^2(\mathcal{C}_a) \tag{17.37}$$

by

$$\widehat{p}(\mathcal{C}_a)\vec{\phi} := -\frac{i\hbar}{a}\frac{d\phi(\theta)}{d\theta}. \tag{17.38}$$

Since $\theta = 0, 2\pi$ refer to the same point in the circle the requirement for all the functions to be single-valued on the circle leads to the periodic boundary conditions.

E17.3.2.2(2) If we ignore the single-valuedness requirement we can adopt the quasi-periodic boundary conditions to introduce a family of differential operators $\widehat{p}_\lambda(\mathcal{C}_a)$, one for each value of $\lambda \in (-\pi, \pi]$, acting on the domain $\vec{\mathcal{D}}\big(\widehat{p}_\lambda(\mathcal{C}_a)\big)$ consisting vectors corresponding to functions satisfying[19]

$$\phi(0) = e^{i\lambda}\phi(2\pi) \quad \text{and} \quad \frac{d\phi(\theta)}{d\theta} \in L^2(\mathcal{C}_a) \tag{17.39}$$

by

$$\widehat{p}_\lambda(\mathcal{C}_a)\vec{\phi} := -i\hbar\frac{1}{a}\frac{d\phi(\theta)}{d\theta}. \tag{17.40}$$

Functions in Eq. (17.39) are not single-valued on the circle, i.e., $\phi(\theta)$ has two different values at the point on $\mathcal{C}_a$ specified by $\theta = 0$ since this coincides with the point on $\mathcal{C}_a$ specified by $\theta = 2\pi$.

---

[18] Martin pp. 46–47. Wan p. 481.

[19] These operators are useful for certain physical systems. *See* Wan p. 490. Without the single-valuedness requirement functions on $\mathcal{C}_a$ are like functions on an interval of $\theta \in [0, 2\pi]$ rather than on a circle, i.e., these functions are similar to the functions in Eq. (17.33). Multi-valued functions are related to a *path space formulation* of quantum mechanics (*see* Wan pp. 637–657).

17.3.2.3 Differential operators in $\vec{L}^2(\mathcal{S}_u)$

The situation is similar to $\vec{L}^2(\mathcal{C}_a)$. Here we have differential expressions like $\partial/\partial\theta$ and $\partial/\partial\varphi$. Let $AC(\partial/\partial\varphi, \mathcal{S}_u)$ denotes the set of functions $\phi(\theta, \varphi)$ on the unit sphere $\mathcal{S}_u$ which are absolutely continuous in φ. We can define an operator $\widehat{L}_z(\mathcal{S}_u)$ with the operator expression $-i\hbar\partial/\partial\varphi$ acting on the domain $\vec{\mathcal{D}}\big(\widehat{L}_z(\mathcal{S}_u)\big)$ consisting of vectors corresponding to functions in $AC(\partial/\partial\varphi, \mathcal{S}_u)$ which also satisfy

$$\phi(\theta, 0) = \phi(\theta, 2\pi) \quad \text{and} \quad \frac{\partial\phi(\theta, \varphi)}{\partial\varphi} \in L^2(\mathcal{S}_u) \tag{17.41}$$

by

$$\widehat{L}_z(\mathcal{S}_u)\vec{\phi} := -i\hbar\,\frac{\partial\phi(\theta, \varphi)}{\partial\varphi}. \tag{17.42}$$

This operator can be identified with the z-component orbital angular momentum operator, hence the subscript z.[20]

17.3.2.4 Differential operators in $\vec{L}^2(I\!R^+)$

E17.3.2.4(1) We can define an operator $\widehat{p}_{\vec{C}_c^\infty}(I\!R^+)$ on the domain $\vec{C}_c^\infty(I\!R^+)$ of vectors corresponding to the set $C_c^\infty(I\!R^+)$ of smooth functions $\psi(x)$ of compact support on $I\!R^+$ by[21]

$$\widehat{p}_{\vec{C}_c^\infty}(I\!R^+)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \tag{17.43}$$

E17.3.2.4(2) We can define a new operator $\widehat{p}_D(I\!R^+)$ on a bigger domain, i.e., a domain $\mathcal{D}\big(\widehat{p}_D(I\!R^+)\big)$ consisting of vectors corresponding to absolutely continuous functions $\psi(x)$ in $L^2(I\!R^+)$ which also satisfy the Dirichlet boundary condition at $x = 0$ and

$$\frac{d\psi(x)}{dx} \in L^2(I\!R^+) \tag{17.44}$$

---

[20] *See* §27.4.

[21] Smooth functions of compact support on $I\!R^+$ are introduced in E16.1.2.5(2). The subscript $\vec{C}_c^\infty$ of the operator tells us that the domain consists of vectors corresponding to functions of compact support in $C_c^\infty(I\!R^+)$.

by

$$\widehat{p}_D(I\!R^+)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \tag{17.45}$$

The Dirichlet boundary condition at $x = \infty$ is automatically satisfied since all functions in $L^2(I\!R^+)$ vanish at infinity.

E17.3.2.4(3) We can also go one step further to drop the Dirichlet condition at $x = 0$ to obtain a still bigger domain to define a new operator $\widehat{p}(I\!R^+)$ on the domain $\vec{\mathcal{D}}\big(\widehat{p}(I\!R^+)\big)$ consisting vectors corresponding to absolutely functions $\psi(x)$ in $L^2(I\!R^+)$, i.e., functions in $L^2(I\!R^+) \cap AC(I\!R^+)$,[22] which also satisfy the condition in Eq. (17.44) by

$$\widehat{p}(I\!R^+)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \tag{17.46}$$

17.3.2.5 Differential operators in $\vec{L}^2(I\!R)$

E17.3.2.5(1) We can define an operator $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ acting on the domain $\vec{C}_c^\infty(I\!R)$ corresponding to the set $C_c^\infty(I\!R)$ of smooth functions $\phi(x)$ of compact support on $I\!R$ by[23]

$$\widehat{p}_{\vec{C}_c^\infty}(I\!R)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx}. \tag{17.47}$$

E17.3.2.5(2) It is desirable to go beyond $\vec{C}_c^\infty(I\!R)$ since most functions we will encounter in physical applications are not of compact support. A natural extension is to include the set $S_s(I\!R)$ of Schwartz functions on $I\!R$.[24] The Schwartz space $\vec{S}_s(I\!R)$ is a subset of $\vec{L}^2(I\!R)$ corresponding to all the functions in $S_s(I\!R)$. We can define an operator $\widehat{p}_{\vec{S}_s}(I\!R)$ on the domain $\vec{S}_s(I\!R)$[25]

$$\widehat{p}_{\vec{S}_s}(I\!R)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx} \quad \forall\, \phi(x) \in S_s(I\!R) \tag{17.48}$$

[22] $AC(I\!R^+)$ is the set of absolutely continuous functions on $I\!R^+$ introduced in E16.1.2.2(2).
[23] Smooth functions of compact support on $I\!R$ are introduced in E16.1.2.5(3).
[24] *See* E16.1.2.3(5).
[25] The subscripts $\vec{S}_s$ indicate the restriction to the Schwartz space $\vec{S}_s(I\!R)$.

E17.3.2.5(3) We can have a new operator $\widehat{p}(I\!R)$ acting on a still bigger domain $\vec{\mathcal{D}}\big(\widehat{p}(I\!R)\big)$ defined by functions in the set[26]

$$\mathcal{D}(\widehat{p}(I\!R)) = \left\{\psi(x) \in L^2(I\!R) \cap AC(I\!R) : \frac{d\psi(x)}{dx} \in L^2(I\!R)\right\} \quad (17.49)$$

by

$$\widehat{p}(I\!R)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \quad (17.50)$$

The Dirichlet boundary condition at infinities are automatically satisfied.

17.3.2.6 Differential operators in $\vec{L}^2(I\!R^2)$ and $\vec{L}^2(I\!R^3)$

E17.3.2.6(1)[27] We can define an operator $\widehat{p}_{x\vec{C}_c^\infty}$ on $\vec{C}_c^\infty(I\!R^2)$ by

$$\widehat{p}_{x\vec{C}_c^\infty}(I\!R^2)\vec{\phi} := -i\hbar\frac{\partial\phi(x,y)}{\partial x}. \quad (17.51)$$

We can similarly define another operator $\widehat{p}_{y\vec{C}_c^\infty}(I\!R^2)$ on $\vec{C}_c^\infty(I\!R^2)$.

E17.3.2.6(2) We can have a new operator $\widehat{p}_x(I\!R^2)$ acting on a bigger domain $\vec{\mathcal{D}}\big(\widehat{p}_x(I\!R^2)\big)$ defined by functions $\psi(x,y)$ in $\vec{L}^2(I\!R^2)$ which are absolutely continuous in x which also satisfy the condition $\partial\psi(x,y)/\partial x \in L^2(I\!R^2)$ by[28]

$$\widehat{p}_x(I\!R^2)\vec{\psi} := -i\hbar\frac{\partial\psi(x,y)}{\partial x}. \quad (17.52)$$

Similarly we can define another operator $\widehat{p}_y(I\!R^2)$.

E17.3.2.6(3) We can define an operator $\widehat{p}_{x\vec{C}_c^\infty}(I\!R^3)$ on the domain $\vec{C}_c^\infty(I\!R^3)$ by

$$\widehat{p}_{x\vec{C}_c^\infty}(I\!R^3)\vec{\phi} := -i\hbar\frac{\partial\phi(x,y,z)}{\partial x}, \quad (17.53)$$

Similarly we have operators $\widehat{p}_{y\vec{C}_c^\infty}(I\!R^3)$ and $\widehat{p}_{z\vec{C}_c^\infty}(I\!R^3)$ on $\vec{C}_c^\infty(I\!R^3)$.

---

[26] $AC(I\!R)$ is the set of absolutely continuous functions on $I\!R$ in E16.1.2.2(3).
[27] *See* Eqs. (24.65).
[28] Wan p. 117.

E17.3.2.6(4) We can have a new operator $\widehat{p}_x(I\!R^3)$ acting on a bigger domain $\vec{\mathcal{D}}\big(\widehat{p}_x(I\!R^3)\big)$ defined by functions $\psi(x, y, z)$ in $\vec{L}^2(I\!R^3)$ which are absolutely continuous in x which also satisfy the condition $\partial\psi(x, y, z)/\partial x \in L^2(I\!R^3)$ by

$$\widehat{p}_x(I\!R^3)\vec{\psi} := -i\hbar\frac{\partial\psi(x, y, z)}{\partial x}. \tag{17.54}$$

Similarly we can define $\widehat{p}_y(I\!R^3)$ and $\widehat{p}_z(I\!R^3)$.

17.4 Algebraic Operations of Operators

The algebraic operations introduced for operators on $\vec{I\!E}^3$ in §8.2.2 apply equally to bounded operators on $\vec{\mathcal{H}}$. For unbounded operators the situation becomes complicated because different operators act on different domains.

Let $\widehat{A}$ and $\widehat{B}$ be two operators with domains $\vec{\mathcal{D}}(\widehat{A})$ and $\vec{\mathcal{D}}(\widehat{B})$, respectively. Then we have the following algebraic operations:

1. Scalar Multiplication

$$\widehat{A}' = c\widehat{A},\ \ c \in \mathbb{C} \quad \Leftrightarrow \quad \widehat{A}'\vec{\phi} := c(\widehat{A}\vec{\phi})\ \ \forall\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.55}$$

2. Addition $\widehat{A} + \widehat{B}$ is defined on the domain

$$\vec{\mathcal{D}}(\widehat{A} + \widehat{B}) := \vec{\mathcal{D}}(\widehat{A}) \cap \vec{\mathcal{D}}(\widehat{B}) \tag{17.56}$$

by

$$(\widehat{A} + \widehat{B})\vec{\phi} := \widehat{A}\vec{\phi} + \widehat{B}\vec{\phi} \quad \forall\,\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}) \cap \vec{\mathcal{D}}(\widehat{B}). \tag{17.57}$$

3. Multiplication $\widehat{A}\widehat{B}$ is defined on the domain

$$\vec{\mathcal{D}}(\widehat{A}\widehat{B}) := \left\{\vec{\phi} : \vec{\phi} \in \vec{\mathcal{D}}(\widehat{B}),\ \ \widehat{A}\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})\right\} \tag{17.58}$$

by

$$(\widehat{A}\widehat{B})\vec{\phi} := \widehat{A}(\widehat{B}\vec{\phi}). \tag{17.59}$$

An example is the operator $\widehat{A}^2$ which would act on a smaller domain than that of $\widehat{A}$, as demonstrated in Eq. (17.16).

4. Equality $\widehat{A} = \widehat{B}$ if

$$\vec{\mathcal{D}}(\widehat{A}) = \vec{\mathcal{D}}(\widehat{B}) = \vec{\mathcal{D}} \quad \text{and} \quad \widehat{A}\vec{\phi} = \widehat{B}\vec{\phi} \quad \forall \vec{\phi} \in \vec{\mathcal{D}}. \tag{17.60}$$

The following comments help to clarify the above definitions:

C17.4(1) The operator $\widehat{A} + \widehat{B}$ can only act on vectors belonging to both $\vec{\mathcal{D}}(\widehat{A})$ and $\vec{\mathcal{D}}(\widehat{B})$, i.e., it cannot act on vectors lying outside the intersection of $\vec{\mathcal{D}}(\widehat{A})$ and $\vec{\mathcal{D}}(\widehat{B})$.

C17.4(2) In the product $\widehat{A}\widehat{B}$ the operator $\widehat{A}$ does not act directly on $\widehat{B}$; $\widehat{A}$ acts on the vector $\widehat{B}\vec{\phi}$. This is why $\vec{\phi}$ must first be in the domain of $\widehat{B}$ and then $\widehat{B}\vec{\phi}$ must be in the domain of $\widehat{A}$. As an example consider the operators $\widehat{x}(I\!R)$ in Eq. (17.12) and $\widehat{p}(I\!R)$ in Eq. (17.50) in the Hilbert space $\vec{L}^2(I\!R)$. The product operator $\widehat{C} = \widehat{p}(I\!R)\widehat{x}(I\!R)$ has an operator expression $(d/dx)(x)$. However, $\widehat{p}(I\!R)$ does not operate directly on $\widehat{x}(I\!R)$ since $\widehat{x}(I\!R)$ is an operator in its own right, i.e., (d/dx) does not act on x in the above operator expression. We must let $\widehat{C}$ act in accordance with the multiplication rule in Eq. (17.59):

(1) Let $\widehat{x}(I\!R)$ operate on $\vec{\phi}$ to arrive at a vector $\vec{\psi}$ defined by the new function $\psi(x) = x\phi(x)$. In the expression $\psi(x) = x\phi(x)$ the variable x is longer an operator.

(2) Next let $\widehat{p}(I\!R)$ act on $\vec{\psi}$ by $\widehat{p}(I\!R)\vec{\psi} := -i\hbar d/dx\big(x\phi(x)\big)$.
In other words, we have

$$\begin{aligned}
\widehat{C}\vec{\phi} = \widehat{p}(I\!R)\left(\widehat{x}(I\!R)\vec{\phi}\right) &:= -i\hbar\frac{d}{dx}\Big(x\phi(x)\Big) \\
&= -i\hbar\left(\left(\frac{d}{dx}x\right)\phi(x) + x\left(\frac{d}{dx}\right)\phi(x)\right) \\
&= -i\hbar\left(\phi(x) + x\frac{d\phi(x)}{dx}\right) \\
&= -i\hbar\left(1 + x\frac{d}{dx}\right)\phi(x).
\end{aligned} \tag{17.61}$$

When restricted to an appropriate domain, e.g., the Schwartz space $\vec{S}_s(I\!R)$, the result can be conveniently written as[29]

[29] *See* Eqs. (17.80) to (17.82).

$$\widehat{p}(I\!R)\widehat{x}(I\!R) = -i\hbar + \widehat{x}(I\!R)\widehat{p}(I\!R), \tag{17.62}$$

or

$$\left(\frac{d}{dx}\right)\left(x\right) = 1 + x\frac{d}{dx}. \tag{17.63}$$

C17.4(3) While the domain becomes smaller after a number of algebraic operations it can remain dense. We shall assume this to be true in all our discussions.

17.5 Invertible Operators

This is an extension to our previous study into inverses and invertible operators on $\vec{I\!E}^3$ in §8.2.2. We shall give a definition of inverse operators applicable to both bounded and unbounded operators in a Hilbert space.

Definition 17.5(1) *An operator $\widehat{A}$ is said to be invertible if it maps its domain $\vec{\mathcal{D}}(\widehat{A})$ onto its range $\vec{\mathcal{R}}(\widehat{A})$ in a one-to-one manner, i.e.,*[30]

$$\vec{\phi}_1 \neq \vec{\phi}_2 \quad \Rightarrow \quad \widehat{A}\vec{\phi}_1 \neq \widehat{A}\vec{\phi}_2 \qquad \forall\, \vec{\phi}_1, \vec{\phi}_2 \in \mathcal{D}(\widehat{A}). \tag{17.64}$$

We can then define the inverse $\widehat{A}^{-1}$ of $\widehat{A}$ to be the operator acting on the domain $\vec{\mathcal{D}}(\widehat{A}^{-1}) = \vec{\mathcal{R}}(\widehat{A})$ by

$$\widehat{A}^{-1}\vec{\psi} := \vec{\phi} \quad \forall\, \vec{\psi} \in \vec{\mathcal{R}}(\widehat{A}) \text{ such that } \vec{\psi} = \widehat{A}\vec{\phi}, \tag{17.65}$$

or more directly by[31]

$$\widehat{A}^{-1}\widehat{A}\,\vec{\phi} := \vec{\phi} \quad \forall\, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.66}$$

The following theorem is similar to Theorem 8.2.2(1) for finite-dimensional spaces.

Theorem 17.5(1) *An operator $\widehat{A}$ with domain $\vec{\mathcal{D}}(\widehat{A})$ and range $\vec{\mathcal{R}}(\widehat{A})$ is invertible if and only if the zero vector $\vec{0}$ is the only vector*

[30] Note that very vector $\vec{\psi} \in \vec{\mathcal{R}}(\widehat{A})$ has a vector $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$ such that $\widehat{A}\vec{\phi} = \vec{\psi}$.
[31] Naimark Part 1 pp. 27–28.

mapped onto the zero vector by $\widehat{A}$, i.e.,

$$\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}), \quad \widehat{A}\vec{\phi} = \vec{0} \quad \Rightarrow \quad \vec{\phi} = \vec{0}. \tag{17.67}$$

This theorem can be established in the same way as the corresponding discussion in §8.2.2. The followings are some useful comments.

C17.5(1) If $\widehat{A}$ is invertible then the inverse operator is unique and that $\widehat{A}$ itself is the inverse of $\widehat{A}^{-1}$, i.e., we have

$$\left(\widehat{A}^{-1}\right)^{-1} = \widehat{A}. \tag{17.68}$$

C17.5(2) The inverse of a bounded operator is not necessarily bounded.[32] As an example consider the multiplication operator $\widehat{x}(\Lambda)$ acting on $\vec{L}^2(\Lambda)$ defined by Eq. (17.22) for an $\Lambda = [0, L]$. This operator is invertible since

$$\widehat{x}(\Lambda)\vec{\phi} := x\phi(x) = 0 \ \Rightarrow \ \vec{\phi} = \vec{0}. \tag{17.69}$$

The inverse is given by

$$\widehat{x}(\Lambda)^{-1}\vec{\phi} := \frac{1}{x}\phi(x), \tag{17.70}$$

which is unbounded. An example of a bounded multiplication operator and its unbounded inverse in $\vec{L}^2(I\!R)$ is the pair

$$e^{-\hat{x}^2/2} \quad \text{and} \quad e^{\hat{x}^2/2}. \tag{17.71}$$

C17.5(3) An operator expression alone cannot determine whether an operator is invertible or not as illustrated by examples below.

(1) Consider the operator $\widehat{p}_{\vec{C}^\infty_{\vec{D}}}(\Lambda)$ in $\vec{L}^2(\Lambda)$ defined by Eq. (17.28). We have

$$\widehat{p}_{\vec{C}^\infty_{\vec{D}}}(\Lambda)\vec{\phi} := -i\hbar\frac{d\phi(x)}{dx} = 0 \ \Rightarrow \ \phi(x) = a, \tag{17.72}$$

which is a constant not necessarily zero. Since a constant function $\phi(x) = a \neq 0 \quad \forall x \in \Lambda$ corresponds to a non-zero vector in $\vec{L}^2(\Lambda)$ the operator $\widehat{p}_{\vec{C}^\infty_{\vec{D}}}(\Lambda)$ in $\vec{L}^2(\Lambda)$ is not invertible.

[32] Roman Vol. 2 p. 500.

Intuitively we can see that since different constant functions are mapped to the zero vector by $\widehat{p}_{\vec{C}_D^\infty}(\Lambda)$ the inverse mapping would not be single-valued, i.e., the inverse mapping would map the zero vector to all these different constant functions making it impossible to define an inverse operator.

(2) The operator $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ in $\vec{L}^2(I\!R)$ defined by Eq. (17.47) also satisfies Eq. (17.72). However, a constant function $\phi(x) = a \neq 0 \quad \forall x \in I\!R$ is not square-integrable over the range $(-\infty, \infty)$. So it does not correspond to a vector in $\vec{L}^2(I\!R)$. It follows that $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ in $\vec{L}^2(I\!R)$ is invertible by Theorem 17.5(1).

C17.5(4) The equality shown in Eq. (8.23) holds only in special cases or in a restricted domain. Generally we may even have $\widehat{A}^{-1}\widehat{A}$ and $\widehat{A}\widehat{A}^{-1}$ operating on different domains.[33] It follows that we may have

$$\widehat{A}^{-1}\widehat{A} \neq \widehat{A}\widehat{A}^{-1}. \tag{17.73}$$

17.6 Extensions and Restrictions of Operators

Definition 17.6(1) *Let $\widehat{A}$ be an operator defined on a dense domain $\vec{\mathcal{D}}(\widehat{A})$ in a Hilbert space $\vec{\mathcal{H}}$. Let $\vec{\mathcal{D}}_{\text{ext}}(\widehat{A})$ and $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$ be two dense subsets of $\vec{\mathcal{H}}$ such that $\vec{\mathcal{D}}_{\text{res}}(\widehat{A}) \subset \vec{\mathcal{D}}(\widehat{A}) \subset \mathcal{D}_{\text{ext}}(\widehat{A})$.*[34]

(1) *The operator $\widehat{A}_{\text{ext}}$ defined on the domain $\vec{\mathcal{D}}_{\text{ext}}(\widehat{A})$ satisfying the condition*

$$\widehat{A}_{ext}\vec{\phi} = \widehat{A}\vec{\phi} \quad \forall \phi \in \vec{\mathcal{D}}(\widehat{A}) \tag{17.74}$$

is called an **extension** *of $\widehat{A}$ to $\vec{\mathcal{D}}_{ext}(\widehat{A})$, a relation is denoted by*

$$\widehat{A} \subset \widehat{A}_{\text{ext}} \quad \text{or} \quad \widehat{A}_{\text{ext}} \supset \widehat{A}. \tag{17.75}$$

---

[33] Roman Vol. 2 p. 493. Jauch p. 34.

[34] $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$ is a subset of $\vec{\mathcal{D}}(\widehat{A})$ which is a subset of $\vec{\mathcal{D}}_{\text{ext}}(\widehat{A})$. All these are assumed to be linear subsets.

(2) *An operator* $\widehat{A}_{\text{res}}$ *defined on the domain* $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$

$$\widehat{A}_{\text{res}}\vec{\phi} = \widehat{A}\vec{\phi} \quad \forall \vec{\phi} \in \vec{\mathcal{D}}_{\text{res}}(\widehat{A}) \tag{17.76}$$

is called the **restriction** of $\widehat{A}$ to $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$. *This relation is denoted by* by $\widehat{A}_{\text{res}} \subset \widehat{A}$.[35]

(3) *A subset* $\vec{\mathcal{D}}$ *of* $\vec{\mathcal{D}}(\widehat{A})$ *is said to be* **invariant under the operator** $\widehat{A}$ *if*

$$\widehat{A}\vec{\phi} \in \vec{\mathcal{D}} \quad \forall\, \vec{\phi} \in \vec{\mathcal{D}}. \tag{17.77}$$

While an extension $\widehat{A}_{\text{ext}}$ acts in the same way as the operator $\widehat{A}$ on the domain $\vec{\mathcal{D}}(\widehat{A})$ of $\widehat{A}$ a restrction $\widehat{A}_{\text{res}}$ acts in the same way as the operator $\widehat{A}$ only on the restricted domain $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$. Many examples of restrictions and extensions of operators have already been presented in §17.3.2. For example, $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ and $\widehat{p}_{\vec{S}_s}(I\!R)$ acting, respectively on $\vec{C}_c^\infty(I\!R)$ and $\vec{S}_s(I\!R)$ in Eqs. (17.47) and (17.48) are restrictions of the operator $\widehat{p}(I\!R)$ in Eq. (17.50). Conversely $\widehat{p}(I\!R)$ is an extension of $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ and $\widehat{p}_{\vec{S}_s}(I\!R)$.

It is often easier to choose a smaller domain to carry out some operations. For example, $\widehat{A}$ may have a restriction $\widehat{A}_{\text{res}}$ to a domain $\mathcal{D}_{\text{res}}(\widehat{A})$ which is invariant under its operations. Then we can carry out any algebraic operations of $\widehat{A}_{\text{res}}$ on $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$, e.g., $\widehat{A}^2_{\text{res}}$ can meaningfully operate on $\vec{\mathcal{D}}_{\text{res}}(\widehat{A})$.[36] The following examples serve to illustrate this situation.

E17.6(1) $\vec{C}_c^\infty(I\!R)$ and $\vec{S}_s(I\!R)$ are invariant under $\widehat{p}(I\!R)$ since the derivative of a smooth function of compact support or rapid decrease on $I\!R$ is again a smooth function of compact support or rapid decrease on $I\!R$. In contrast the domain $\vec{\mathcal{D}}\big(\widehat{p}(I\!R)\big)$ in Eq. (17.49) is not invariant under the operator $\widehat{p}(I\!R)$.

E17.6(2) $\vec{C}_c^\infty(I\!R)$ and $\vec{S}_s(I\!R)$ are invariant under the position operator $\widehat{x}(I\!R)$ in Eq. (17.12). We can define two restrictions, $\widehat{x}_{\vec{C}_c^\infty}(I\!R)$

[35] The relationship between an operator and its restriction or extension should not be confused with the order relationship between projectors given by Definition 13.2.2(2).

[36] This may not be true on the original domain $\mathcal{D}(\widehat{A})$, e.g., the domain of $\widehat{A}^2$ is generally not equal to $\vec{\mathcal{D}}(\widehat{A})$, as seen in Eq. (17.16).

and $\widehat{x}_{\vec{S}_s}(I\!R)$, of $\widehat{x}(I\!R)$ acting on the domains $\vec{C}_c^{\infty}(I\!R)$ and $\vec{S}_s(I\!R)$, respectively.

Technically and intuitively it is easier to introduce an operator on a smaller domain, and then extend it to act on a bigger domain. This is why quantisation schemes in quantum mechanics often start by defining an operator, e.g., $\widehat{p}_{\vec{C}_c^{\infty}}(I\!R)$, on a small domain, and then extend the domain to obtain the desired operator, e.g., $\widehat{p}(I\!R)$.[37]

17.7 Commutation Relations

Let us follow the construction of the commutator of two bounded operators in Eq. (8.19) to define the commutator of two possibly unbounded operators $\widehat{A}$ and $\widehat{B}$ in a Hilbert space to be $\widehat{A}\widehat{B} - \widehat{B}\widehat{A}$ with the notation $[\,\widehat{A},\,\widehat{B}\,]$, i.e., we have

$$[\,\widehat{A},\,\widehat{B}] := \widehat{A}\widehat{B} - \widehat{B}\widehat{A}. \tag{17.78}$$

The domain of $\widehat{A}\widehat{B}$ may well differ from that of $\widehat{B}\widehat{A}$.[38] This causes complications in manipulating commutators. Generally it would not be correct to write down equalities like

$$\widehat{A}\widehat{B} = \widehat{B}\widehat{A} \quad \text{or} \quad [\,\widehat{A},\,\widehat{B}\,] = \widehat{0}, \tag{17.79}$$

since the domain of both sides of the equations may not agree.

There are many cases where two operators $\widehat{A}$ and $\widehat{B}$ may possess a common dense domain $\vec{\mathcal{D}}$ which is invariant under the operation of both $\widehat{A}$ and $\widehat{B}$. We can then perform any algebraic operations of $\widehat{A}$ and $\widehat{B}$ on $\vec{\mathcal{D}}$. For example, the operators $\widehat{p}_{\vec{S}_s}(I\!R)$ and $\widehat{x}_{\vec{S}_s}(I\!R)$ in $\vec{L}^2(I\!R)$ have $\vec{S}_s(I\!R)$ as a common dense domain which is invariant under both $\widehat{p}_{\vec{S}_s}(I\!R)$ and $\widehat{x}_{\vec{S}_s}(I\!R)$. The commutator $[\,\widehat{x}_{\vec{S}_s}(I\!R),\,\widehat{p}_{\vec{S}_s}(I\!R)\,]$ is well-defined on $\vec{S}_s(I\!R)$, and we have, following Eq. (17.63),

$$\begin{aligned}[\,\widehat{x}_{\vec{S}_s}(I\!R),\,\widehat{p}_{\vec{S}_s}(I\!R)\,]\vec{\phi} &=: -i\hbar\left(x\frac{d}{dx} - \frac{d}{dx}x\right)\phi(x) \\ &= -i\hbar\,\phi(x) \qquad \forall\phi(x) \in S_s(I\!R).\end{aligned} \tag{17.80}$$

[37] Wan pp. 252–255.
[38] Jauch pp. 42–43. Akhiezer and Glasmann Vol. 1 pp. 31–32. Wan p. 102.

In terms of operators we have[39]

$$[\,\widehat{x}_{\vec{S}_s}(I\!R),\ \widehat{p}_{\vec{S}_s}(I\!R)\,] \subset i\hbar\,\widehat{I\!I}\,(I\!R), \tag{17.81}$$

namely $[\,\widehat{x}_{\vec{S}_s}(I\!R),\ \widehat{p}_{\vec{S}_s}(I\!R)\,]$ is a restriction of the identity operator which is defined on the entire $\vec{L}^2(I\!R)$. This relation is can be rewritten as[40]

$$[\,\widehat{x}_{\vec{S}_s}(I\!R),\ \widehat{p}_{\vec{S}_s}(I\!R)\,] = i\hbar\,\widehat{I\!I}_{\vec{S}_s}(I\!R). \tag{17.82}$$

It is common to write[41]

$$[\,\widehat{x}(I\!R),\ \widehat{p}(I\!R)\,] = i\hbar\,\widehat{I\!I}\,(I\!R). \tag{17.83}$$

Since this is so widely used we shall also adopt this usage in this book when dealing with commutation relations such as those of the orbital angular momentum operators in §27.4, with the understanding that the equality holds in an appropriate dense subset of the Hilbert space such as $\vec{C}_c^{\infty}(I\!R^3)$ or $\vec{S}_s(I\!R^3)$. In some specific cases we have to specify the domain of operation of the commutator explicitly. Formal calculations can lead to erroneous results. An example relating to the uncertainty relation in §28.3.3 will illustrate this point.

The problems with domains make it difficult to give a general definition of the commutativity of two arbitrary unbounded operators in a Hilbert space $\vec{\mathcal{H}}$. Fortunately such a general definition is not necessary for our purposes. All we need is to have a definition of the commutativity in three cases:

Case (1) *commutativity of two bounded operators* $\widehat{B}_1,\ \widehat{B}_2$ Our previous definition given by Eq. (8.20) applies. There is no problem defining their products $\widehat{B}_1\widehat{B}_2$ and $\widehat{B}_2\widehat{B}_1$ since all operators are defined on the entire Hilbert space. We have

$$\widehat{B}_1\widehat{B}_2 - \widehat{B}_2\widehat{B}_1 = \widehat{0} \quad \text{or} \quad \widehat{B}_1\widehat{B}_2 = \widehat{B}_2\widehat{B}_1. \tag{17.84}$$

In other words, if $\widehat{B}_1$ and $\widehat{B}_2$ commute then $\widehat{B}_1\widehat{B}_2 = \widehat{B}_2\widehat{B}_1$.

[39] Here $\widehat{I\!I}\,(I\!R)$ is the identity operator on $\vec{L}^2(I\!R)$.
[40] Here $\widehat{I\!I}_{\vec{S}_s}(I\!R)$ is the restriction of $\widehat{I\!I}\,(I\!R)$ to the Schwartz space $\vec{S}_s(I\!R)$.
[41] This is incorrect, as pointed out by Jauch p. 43. *See* also Eq. (27.154).

Case (2) *commutativity of a bounded operator* $\widehat{B}$ *and an unbounded oeprators* $\widehat{A}$ This is defined below.

Definition 17.7(1)[42] *A bounded operator* $\widehat{B}$ *is said to commute with a possibly unbounded operator* $\widehat{A}$ *if* $\widehat{A}\widehat{B}$ *is an extension of* $\widehat{B}\widehat{A}$*, i.e.,*

$$\widehat{B}\widehat{A} \subset \widehat{A}\widehat{B}. \tag{17.85}$$

To appreciate this definition we observe that

(1) The domain $\vec{\mathcal{D}}(\widehat{B}\widehat{A})$ of the product operator $\widehat{B}\widehat{A}$ is the same as $\vec{\mathcal{D}}(\widehat{A})$ of $\widehat{A}$.

(2) The domain $\vec{\mathcal{D}}(\widehat{A}\widehat{B})$ of the product operator $\widehat{A}\widehat{B}$ is generally not the same as $\vec{\mathcal{D}}(\widehat{A})$, i.e.,

$$\vec{\mathcal{D}}(\widehat{A}\widehat{B}) := \left\{\vec{\phi} \in \vec{\mathcal{H}} : \ \widehat{B}\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})\right\}, \tag{17.86}$$

where $\vec{\phi}$ is not neccessarily in $\vec{\mathcal{D}}(\widehat{A})$.

(3) Equation (17.85) means that $\widehat{B}\widehat{A}$ is a restriction of $\widehat{A}\widehat{B}$:

(a) The domain $\vec{\mathcal{D}}(\widehat{B}\widehat{A})$ is smaller than or equal to $\vec{\mathcal{D}}(\widehat{A}\widehat{B})$. This requires

$$\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}) \ \Rightarrow \ \widehat{B}\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.87}$$

(b) On $\vec{\mathcal{D}}(\widehat{B}\widehat{A}) = \vec{\mathcal{D}}(\widehat{A})$ the two operators, i.e., $\widehat{A}\widehat{B}$ and $\widehat{B}\widehat{A}$, are the same, i.e.,

$$\left(\widehat{B}\widehat{A} - \widehat{A}\widehat{B}\right)\vec{\phi} = \vec{0} \quad \text{or} \quad \widehat{B}\widehat{A}\vec{\phi} = \widehat{A}\widehat{B}\,\vec{\phi} \tag{17.88}$$

for all $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$ for which $\widehat{B}\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$.

(c) Applications of this type of commutativity can be found in Theorems 17.9(1) and 20.7(1), and in §27.9.

Case (3) *commutativity of unbounded selfadjoint operators* This will be discussed in §20.6, after a detailed study of unbounded selfadjoint operators.

[42] Akhiezer and Galzman Vol. 1 p. 31. Jauch p. 42.

17.8 Adjoints of Operators

For bounded operators in a Hilbert space the simple condition for the definition of their adjoints in Eq. (8.33) applies. The properties listed in Eqs. (8.38) to (8.41) also remain true. For unbounded operators the condition in Eq. (8.33) cannot be applied. To establish a general definition of selfadjoint operators we have to define the adjoint of an operator first.

Let $\widehat{A}$ be a possibly unbounded operator defined on a dense domain $\vec{\mathcal{D}}(\widehat{A})$ in a Hilbert space $\vec{\mathcal{H}}$. Let $\vec{\psi} \in \vec{\mathcal{H}}$ such that

(1) There exists another vector $\vec{\psi}' \in \vec{\mathcal{H}}$ satisfying the following condition:

$$\langle \vec{\psi} \mid \widehat{A}\vec{\phi} \rangle = \langle \vec{\psi}' \mid \vec{\phi} \rangle \quad \forall\, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.89}$$

Note that the above equality has to hold for all the vectors $\vec{\phi}$ in the domain of $\widehat{A}$.

(2) The vector $\vec{\psi}'$ is unique to $\vec{\psi}$, i.e., there are no other vectors which can satisfy the above condition for the given $\vec{\psi}$.

It can be shown that for an operator $\widehat{A}$ defined on a dense domain there are vectors $\vec{\psi}$ having the two properties listed above. Let $\{\vec{\psi}\}$ be the set of all vectors in $\vec{\mathcal{H}}$ satisfying the above two conditions. The unigue relation between $\vec{\psi}$ and $\vec{\psi}'$ enables us to define an operator to act on the set $\{\vec{\psi}\}$, i.e., we can define a new operator, denoted by $\widehat{A}^\dagger$, acting on the domain $\vec{\mathcal{D}}(\widehat{A}^\dagger) := \{\vec{\psi}\}$ by[43]

$$\widehat{A}^\dagger \vec{\psi} := \vec{\psi}' \quad \forall \vec{\psi} \in \vec{\mathcal{D}}(\widehat{A}^\dagger). \tag{17.90}$$

We can rewrite Eq. (17.89) in terms of this new operator as

$$\langle \vec{\psi} \mid \widehat{A}\vec{\phi} \rangle = \langle \widehat{A}^\dagger \vec{\psi} \mid \vec{\phi} \rangle \quad \forall\, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}) \text{ and } \forall \vec{\psi} \in \vec{\mathcal{D}}(\widehat{A}^\dagger). \tag{17.91}$$

In $\vec{L}^2(I\!R)$ an explicit expression for Eq. (17.91) is

$$\int_{-\infty}^{\infty} \psi^*(x)\phi'(x)\,dx = \int_{-\infty}^{\infty} \psi'^*(x)\phi(x)\,dx, \tag{17.92}$$

where $\widehat{A}\vec{\phi} := \phi'(x)$ and $\widehat{A}^\dagger \vec{\psi} := \psi'(x)$.

[43] Prugovečki p. 187.

Definition 17.8(1)

(1) *The adjoint of an operator* $\widehat{A}$ *defined on a dense domain is the operator* $\widehat{A}^\dagger$ *defined by* Eq. (17.90).

(2) *An operator is* **selfadjoint** *if it is equal to its adjoint.*[44]

Generally the domain of the adjoint operator is not necessarily dense. If $\widehat{A}$ is bounded then the condition in Eq. (8.33) applies, i.e., Eq. (17.91) becomes[45]

$$\langle \vec{\psi} \mid \widehat{A}\vec{\phi} \rangle = \langle \widehat{A}^\dagger \vec{\psi} \mid \vec{\phi} \rangle \quad \forall \vec{\phi},\ \vec{\psi} \in \vec{\mathcal{H}}. \tag{17.93}$$

The adjoint operation possesses the following properties:

P17.8(1) The adjoint $\widehat{A}^\dagger$ exists if $\vec{\mathcal{D}}(\widehat{A})$ is dense.[46] From now on we will confine ourselves to operators whose adjoints are also defined on a dense domain.

P17.8(2) Equations (8.39) and (8.33) apply, i.e., we have[47]

$$(c\widehat{A})^\dagger = c^* \widehat{A}^\dagger, \quad (\widehat{A}^{-1})^\dagger = (\widehat{A}^\dagger)^{-1}. \tag{17.94}$$

P17.8(3) If $\widehat{A}$ is bounded then:

(1) The adjoint $\widehat{A}^\dagger$ is bounded and $\widehat{A}^{\dagger\dagger} = \widehat{A}$.[48]

(2) Equations (13.3) and (13.4) remain valid, i.e.,[49]

$$(a\widehat{A} + b\widehat{B})^\dagger = a^* \widehat{A}^\dagger + b^* \widehat{B}^\dagger. \tag{17.95}$$

$$(\widehat{A}\widehat{B})^\dagger = \widehat{B}^\dagger \widehat{A}^\dagger. \tag{17.96}$$

These equations are not satisfied for unbounded operators as seen in some of the properties listed below.

---

[44] This agrees with Definition 9.4.1(1).

[45] Roman Vol. 2 pp. 518–519. The adjoint of a bounded operator is bounded.

[46] All the operators in this book are assumed to be defined on a dense domain. We do sometimes state this assumption explicitly to highlight its importance.

[47] Akhiezer and Glasman Vol. 1 p. 80. Here $c \in \mathbb{C}$.

[48] Weidman p. 68. Roman Vol. 2 p. 519.

[49] Weidman p. 73. Riesz and Nagy p. 301. Here $a, b \in \mathbb{C}$. For unbounded operators these results may not hold.

P17.8(4) If $\widehat{A}$ is a restriction of $\widehat{B}$ then $\widehat{A}^\dagger$ is an extension of $\widehat{B}^\dagger$, i.e.,

$$\widehat{A} \subset \widehat{B} \quad \Rightarrow \quad \widehat{A}^\dagger \supset \widehat{B}^\dagger \tag{17.97}$$

P17.8(5) The adjoint of $\widehat{A}^\dagger$, i.e., $\widehat{A}^{\dagger\dagger} = (\widehat{A}^\dagger)^\dagger$, is an extension of $\widehat{A}$, i.e.,[50]

$$\widehat{A} \subset \widehat{A}^{\dagger\dagger}. \tag{17.98}$$

P17.8(6) The adjoint $\widehat{A}^\dagger$ of an operator $\widehat{A}$ is closed and[51]

$$\widehat{A}^{\dagger\dagger} = \widehat{A} \quad \Leftrightarrow \quad \widehat{A} \text{ is closed.} \tag{17.99}$$

P17.8(7) When $\widehat{A}$ and $\widehat{B}$ are unbounded we have[52]

$$(a\widehat{A} + b\widehat{B})^\dagger \supset a^*\widehat{A}^\dagger + b^*\widehat{B}^\dagger. \tag{17.100}$$

$$(\widehat{A}\widehat{B})^\dagger \supset \widehat{B}^\dagger\widehat{A}^\dagger. \tag{17.101}$$

Generally the sum of two densely defined operators may not even be densely defined. In other words, the adjoint of the sum of two unbounded operators may not exist. In physical applications we may encounter operators with more favourable properties such that Eq. (17.101) becomes an equality.[53]

P17.8(8) A selfadjoint operator $\widehat{A}$ must satisfy

$$\langle \vec{\psi} \mid \widehat{A}\vec{\phi} \rangle = \langle \widehat{A}\vec{\psi} \mid \vec{\phi} \rangle \quad \forall \vec{\phi}, \ \vec{\psi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.102}$$

on account of Eq. (17.91). This is a necessary but not sufficient condition for selfadjointness for unbounded operators.[54] As before selfadjoint operators possess many desirable properties. The situation can be complicated when dealing with unbounded selfadjoint operators.[55] For example, the sum of two unbounded selfadjoint

[50] Akhiezer and Glasman Vol. 1 pp. 80, 96.
[51] Fano p. 278.
[52] *See* Eqs. (35.23) and (35.24) for an example.
[53] *See* examples in Eqs. (27.82) and (35.9).
[54] *See* also Eq. (19.1). This condition is related to *symmetric operators* in Definition 19.1(1), and is hence known as is known as the symmetry condition.
[55] We shall return to examine unbounded selfadjoint operators in more details in Chapters 18 and 20.

operators is not necessarily selfadjoint because of Eq. (17.101). For unbounded operators this poses severe challenges when it comes to the problem of quantisation in quantum mechanics.[56] There are cases where we can still establish a selfadjoint sum of two unbounded selfadjoint operators. The sum $\widehat{A}+\widehat{B}$ is defined on the domain $\mathcal{D}(\widehat{A})\cap\mathcal{D}(\widehat{B})$ by definition.[57] It may be possible to find a **selfadjoint extension** to $\widehat{A}+\widehat{B}$.[58] In physical applications an appropriate selfadjoint extension can be taken to be the selfadjoint sum of $\widehat{A}$ and $\widehat{B}$. An example is given in §36.1.1 on the quantisation of orbital angular momentum operators.

17.9 Reduction of Operators

In addition to extending and restricting an operator we may also be able to reduce an operator. We start by extending the concept of invariant subsets of an operator in Definition 17.6(1) to subspaces.

Definition 17.9(1) *A subspace $\vec{\mathcal{S}}$ of $\vec{\mathcal{H}}$ is said to be invariant under an operator $\widehat{A}$ if*

$$\vec{\phi}\in\vec{\mathcal{S}}\cap\vec{\mathcal{D}}(\widehat{A})\quad\Rightarrow\quad\widehat{A}\vec{\phi}\in\vec{\mathcal{S}}. \tag{17.103}$$

The subspace is referred to as an **invariant subspace** *of $\widehat{A}$.*

Every vector $\vec{\phi}$ in $\vec{\mathcal{S}}$ which is also in $\vec{\mathcal{D}}(\widehat{A})$ can be acted on by $\widehat{A}$. If the output vector $\widehat{A}\vec{\phi}$ for every input vector $\vec{\phi}\in\vec{\mathcal{S}}\cap\vec{\mathcal{D}}(\widehat{A})$ is again in $\vec{\mathcal{S}}$ then the subspace is said to be invariant under $\widehat{A}$. Note that $\vec{\mathcal{S}}$ being an invariant subspace of $\widehat{A}$ does not imply that its orthogonal complement $\vec{\mathcal{S}}^{\perp}$ is also an invariant subspace of $\widehat{A}$.

Definition 17.9(2)[59] *A subspace $\vec{\mathcal{S}}$ of $\vec{\mathcal{H}}$ is said to be a* **reducing subspace** *of $\widehat{A}$ if*

(1) *the subspace $\vec{\mathcal{S}}$ and its orthogonal complement $\vec{\mathcal{S}}^{\perp}$ are both invariant under $\widehat{A}$, and*

[56] *See* Chapter 27 on quantisation and Wan §3.3.
[57] *See* Eq. (17.56).
[58] A selfadjoint extension is an extension which is selfadjoint (*see* §19.3).
[59] Akhiezer and Glasman Vol. 1 p. 82. Roman Vol. 2 p. 572. Weidman pp. 127–128. Smirnov p. 559–562.

(2) *the domain of* $\widehat{A}$ *is invariant under the projector* $\widehat{P}_{\vec{\mathcal{S}}}$*, onto the subspace* $\vec{\mathcal{S}}$*, i.e.,*[60]

$$\widehat{P}_{\vec{\mathcal{S}}}\,\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}) \quad \forall\, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.104}$$

The operator $\widehat{A}$ *is then said to possesses a reducing subspace.*

The following comments aim to clarify the above definition:

C17.9(1) For a bounded operator $\widehat{A}$ the condition in Eq. (17.104) is redundant since $\vec{\mathcal{D}}(\widehat{A}) = \vec{\mathcal{H}}$. An obvious example is that a subspace $\vec{\mathcal{S}}$ is a reducing subspace of its associated projector $\widehat{P}_{\vec{\mathcal{S}}}$.

C17.9(2) For an unbounded operator $\widehat{A}$ the condition in Eq. (17.104) ensures that the projection of every vector in $\vec{\mathcal{D}}(\widehat{A})$ onto $\vec{\mathcal{S}}$ can be acted on by $\widehat{A}$.

C17.9(3) If $\vec{\mathcal{S}}$ is a reducing subspace of $\widehat{A}$ then its orthogonal complement $\vec{\mathcal{S}}^{\perp}$ is also a reducing subspace of $\widehat{A}$, i.e.,[61]

$$\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}), \quad \widehat{P}_{\vec{\mathcal{S}}}\,\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}) \tag{17.105}$$

$$\Rightarrow \quad \vec{\phi} - \widehat{P}_{\vec{\mathcal{S}}}\,\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}) \tag{17.106}$$

$$\Rightarrow \quad \widehat{P}_{\vec{\mathcal{S}}^{\perp}}\vec{\phi} = \big(\widehat{\mathbb{I}} - \widehat{P}_{\vec{\mathcal{S}}}\big)\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{17.107}$$

C17.9(4) The properties stated in C17.9(2) and C17.9(3) enable us to reduce $\widehat{A}$ into the sum of two operators which also act on the domain $\vec{\mathcal{D}}(\widehat{A})$, i.e., we have[62]

$$\widehat{A} = \widehat{A}_{\vec{\mathcal{S}}} + \widehat{A}_{\vec{\mathcal{S}}^{\perp}}, \quad \text{where } \widehat{A}_{\vec{\mathcal{S}}} := \widehat{A}\widehat{P}_{\vec{\mathcal{S}}}, \quad \widehat{A}_{\vec{\mathcal{S}}^{\perp}} := \widehat{A}\widehat{P}_{\vec{\mathcal{S}}^{\perp}}, \tag{17.108}$$

because

$$\widehat{A} = \widehat{A}\Big(\widehat{P}_{\vec{\mathcal{S}}} + \widehat{A}_{\vec{\mathcal{S}}^{\perp}}\Big) = \widehat{A}\widehat{P}_{\vec{\mathcal{S}}} + \widehat{A}\widehat{P}_{\vec{\mathcal{S}}^{\perp}}, \tag{17.109}$$

Here $\widehat{A}_{\vec{\mathcal{S}}}$ and $\widehat{A}_{\vec{\mathcal{S}}^{\perp}}$ can act on $\vec{\mathcal{D}}(\widehat{A})$ on account of Eqs. (17.105) and (17.107).

---

[60] Projectors in a Hilbert space are defined by Definition 16.2.3(2).

[61] Note that $\mathcal{D}(\widehat{A})$ is a linear subset.

[62] The concept of reduction can be appreciated when we discuss a direct sum decomposition of Hilbert spaces and operators in §24.1.2.

Definition 17.9(3)

(1) *An operator* $\widehat{A}$ *is said to be* **reducible** *if it possesses a reducing subspace* $\vec{\mathcal{S}}$*. The two operators defined by*

$$\widehat{A}_{\vec{\mathcal{S}}} := \widehat{A}\widehat{P}_{\vec{\mathcal{S}}} \quad \text{and} \quad \widehat{A}_{\vec{\mathcal{S}}^\perp} := \widehat{A}\widehat{P}_{\vec{\mathcal{S}}^\perp} \tag{17.110}$$

are called the parts of $\widehat{A}$ *in* $\vec{\mathcal{S}}$ *and in* $\vec{\mathcal{S}}^\perp$*, respectively. The subspace is also said to reduce the operator.*

(2) *An operator* $\widehat{A}$ *is said to be* **irreducible** *if it does not possess a reducing subspace.*[63]

The following theorem help to determine if an operator is reducible or not.[64]

Theorem 17.9(1)

(1) *A subspace* $\vec{\mathcal{S}}$ *reduces an operator* $\widehat{A}$ *if and only if* $\widehat{A}$ *commutes with the projector* $\widehat{P}_{\vec{\mathcal{S}}}$ *onto* $\vec{\mathcal{S}}$*.*

(2) *A subspace* $\vec{\mathcal{S}}$ *which reduces an operator* $\widehat{A}$ *also reduces its adjoint* $\widehat{A}^\dagger$ *if* $\widehat{A}$ *and* $\widehat{A}^\dagger$ *have the same domain, i.e., if* $\vec{\mathcal{D}}(\widehat{A}) = \vec{\mathcal{D}}(\widehat{A}^\dagger)$*.*

(3) *A subspace* $\vec{\mathcal{S}}$ *reduces a bounded selfadjoint operator* $\widehat{A}$ *if and only if* $\vec{\mathcal{S}}$ *is invariant under* $\widehat{A}$*.*

(4) *If subspace* $\vec{\mathcal{S}}$ *reduces a bounded selfadjoint operator* $\widehat{A}$ *then its parts* $\widehat{A}_{\vec{\mathcal{S}}}$ *in* $\vec{\mathcal{S}}$ *and* $\widehat{A}_{\vec{\mathcal{S}}^\perp}$ *in* $\vec{\mathcal{S}}^\perp$ *are also selfadjoint.*

If $\vec{\mathcal{S}}$ reduces $\widehat{A}$ its orthogonal complement $\vec{\mathcal{S}}^\perp$ also reduces $\widehat{A}$. It follows that $\widehat{A}$ would also commute with $\widehat{P}_{\vec{\mathcal{S}}^\perp}$. Since $\widehat{A}$ commutes with $\widehat{P}_{\vec{\mathcal{S}}}$ and $\widehat{P}_{\vec{\mathcal{S}}^\perp}$ we can rewrite $\widehat{A}_{\vec{\mathcal{S}}}$ and $\widehat{A}_{\vec{\mathcal{S}}^\perp}$ as

$$\widehat{A}_{\vec{\mathcal{S}}} := \widehat{P}_{\vec{\mathcal{S}}}\widehat{A}\widehat{P}_{\vec{\mathcal{S}}} \quad \text{and} \quad \widehat{A}_{\vec{\mathcal{S}}^\perp} := \widehat{P}_{\vec{\mathcal{S}}^\perp}\widehat{A}\widehat{P}_{\vec{\mathcal{S}}^\perp}. \tag{17.111}$$

It is useful to extend the concept of reducibility or otherwise to a set of two or more operators.[65]

[63] We exclude trivial cases, i.e., the subspace containing only the zero vector and the whole Hilbert space.
[64] Roman Vol. 2 pp. 572–577. Weidman pp. 127–128.
[65] Jordan pp. 67–69.

Definition 17.9(4) *A subspace is said to reduce a set of operators if it reduces every operator in the set. A set of operators is said to be irreducible if there does not exist a subspace which reduces every operator of the set.*

Applications of the notion of reducibility and irreducibily of operators can be found in §17.10, §27.9 and §35.2.

17.10 Annihilation and Creation Operators

Generally the adjoint $\widehat{A}^\dagger$ of an operator $\widehat{A}$ may be very different from $\widehat{A}$. They may have different domains, i.e., $\vec{\mathcal{D}}(\widehat{A}) \neq \vec{\mathcal{D}}(\widehat{A}^\dagger)$, and they may act differently even on the intersection $\vec{\mathcal{D}}(\widehat{A}) \cap \vec{\mathcal{D}}(\widehat{A}^\dagger)$ of their domains. The following two cases are of particular interest:

(1) The domain $\vec{\mathcal{D}}(\widehat{A})$ is a subset of $\vec{\mathcal{D}}(\widehat{A}^\dagger)$, and on the intersection of their domains, i.e., on $\vec{\mathcal{D}}(\widehat{A}) \cap \vec{\mathcal{D}}(\widehat{A}^\dagger)$ which is equal to $\vec{\mathcal{D}}(\widehat{A})$, they act in the same way. In other words, the adjoint $\widehat{A}^\dagger$ is an extension of $\widehat{A}$, i.e., $\widehat{A} \subset \widehat{A}^\dagger$. Operators with these properties are called *symmetric*. We shall return to study these operators in Chapter 19.

(2) They have the same domain, i.e., $\vec{\mathcal{D}}(\widehat{A}) = \vec{\mathcal{D}}(\widehat{A}^\dagger)$ but they act differently in their common domain, i.e., $\widehat{A}^\dagger \neq \widehat{A}$ on $\vec{\mathcal{D}}(\widehat{A})$. The best known examples in quantum mechanics are the *annihilation* and *creation operators*.

Definition 17.10(1) *Let* $\{\vec{\varphi}_n, n = 0, 1, 2, \ldots\}$ *be an orthonormal basis in an infinite-dimensional Hilbert space* $\vec{\mathcal{H}}$.[66] *The* **annihilation operator** *associated with the basis, denoted by* $\widehat{a}$*, is the operator defined on the domain*[67]

[66] Roman Vol. 2 p. 528. We number the basis vectors from $n = 0$ rather than the usual numbering 1, 2, 3 . . .. Such a numbering is more intuitive for many applications.

[67] It is a standard convention that annihilation creation operators are denoted by the lower case letter a. Note that the index n in $\vec{\varphi}_n$ starts with 0.

$$\vec{\mathcal{D}}(\hat{a}) = \left\{ \vec{\phi} \in \vec{\mathcal{H}} : \sum_{n=1}^{\infty} |\langle \vec{\varphi}_n \mid \vec{\phi} \rangle|^2 \, n < \infty \right\} \tag{17.112}$$

$$\text{by} \quad \hat{a}\,\vec{\phi} = \sum_{n=1}^{\infty} \sqrt{n}\, \langle \vec{\varphi}_n \mid \vec{\phi} \rangle\, \vec{\varphi}_{n-1}. \tag{17.113}$$

We can appreciate the definition with a more intuitive approach[68]:

(1) Define the action of the operator on the basis vectors by

$$\hat{a}\,\vec{\varphi}_0 = \vec{0}, \tag{17.114}$$

$$\hat{a}\,\vec{\varphi}_n = \sqrt{n}\,\vec{\varphi}_{n-1}, \qquad \forall n \geq 1. \tag{17.115}$$

The operator is said to *annihilate* $\vec{\varphi}_0$.

(2) Define the action of the operator on a general vector $\vec{\phi}$ by

(a) expressing $\vec{\phi}$ in terms of the basis vectors as

$$\vec{\phi} = \sum_{n=0}^{\infty} c_n\, \vec{\varphi}_n, \quad c_n = \langle \vec{\varphi}_n \mid \vec{\phi} \rangle, \tag{17.116}$$

(b) and then extending the action of $\hat{a}$ linearly to $\vec{\phi}$, i.e.,

$$\hat{a}\,\vec{\phi} = \sum_{n=0}^{\infty} c_n\, \hat{a}\,\vec{\varphi}_n = \sum_{n=1}^{\infty} \sqrt{n}\, c_n\, \vec{\varphi}_{n-1}. \tag{17.117}$$

This agrees with the action given in Eq. (17.113).

(3) For the extension in Eq. (17.117) to be meaningful the output vector $\hat{a}\,\vec{\phi}$ must have a finite norm, i.e., we must have

$$\begin{aligned} ||\,\hat{a}\,\vec{\phi}\,||^2 &= \langle\, \hat{a}\,\vec{\phi} \mid \hat{a}\,\vec{\phi}\,\rangle \\ &= \langle \sum_{n=1}^{\infty} \sqrt{n}\, c_n\, \vec{\varphi}_{n-1} \mid \sum_{m=1}^{\infty} \sqrt{m}\, c_m\, \vec{\varphi}_{m-1} \rangle \\ &= \sum_{n=1}^{\infty} |c_n|^2\, n < \infty, \end{aligned} \tag{17.118}$$

which is the condition imposed on vectors in the domain $\vec{\mathcal{D}}(\hat{a})$ in Eq. (17.112). This shows that $\hat{a}$ is an unbounded operator not capable of acting on every vector in the Hilbert space.

[68] Jauch p. 44.

Definition 17.10(2) *Let* $\{\vec{\varphi}_n,\, n = 0, 1, 2, \ldots\}$ *be an orthonormal basis in an infinite-dimensional Hilbert space* $\vec{\mathcal{H}}$. *The* **creation operator** *associated with the basis, denoted by* $\widehat{a}^\dagger$, *is the operator defined on the domain*

$$\vec{\mathcal{D}}(\widehat{a}^\dagger) = \Big\{\vec{\phi} \in \vec{\mathcal{H}} : \sum_{n=1}^{\infty} |\langle \vec{\varphi}_n \mid \vec{\phi}\,\rangle|^2\, n < \infty\Big\}. \quad (17.119)$$

$$\text{by} \quad \widehat{a}^\dagger \vec{\phi} = \sum_{n=0}^{\infty} \sqrt{n+1}\, \langle \vec{\varphi}_n \mid \vec{\phi}\,\rangle\, \vec{\varphi}_{n+1}. \quad (17.120)$$

The definition can again be introduced by first letting the operator acts on the basis vectors $\vec{\varphi}_n$, i.e.,

$$\widehat{a}^\dagger \vec{\varphi}_n = \sqrt{n+1}\, \vec{\varphi}_{n+1}, \quad n = 0, 1, 2, \ldots, \quad (17.121)$$

and then extending the action linearly to a general vector $\vec{\phi}$. The same condition has to be imposed on $\vec{\phi}$ resulting in the operator $\widehat{a}^\dagger$ having the same domain as $\widehat{a}$.

We shall show that $\widehat{a}^\dagger$ is the adjoint of $\widehat{a}$.[69] To verify this we have, for all $\vec{\phi}, \vec{\phi}' \in \vec{\mathcal{D}}(\widehat{a})$,

$$\begin{aligned}\langle \vec{\phi} \mid \widehat{a}\vec{\phi}'\,\rangle &= \langle \sum_{n=0}^{\infty} c_n\, \vec{\varphi}_n \mid \sum_{n'=0}^{\infty} c'_{n'}\, \widehat{a}\vec{\varphi}_{n'}\rangle = \sum_{n=0,n'=0}^{\infty} c_n^*\, c'_{n'} \langle \vec{\varphi}_n \mid \widehat{a}\vec{\varphi}_{n'}\rangle \\ &= \sum_{n=0}^{\infty} c_n^*\, c'_{n+1}\, \sqrt{n+1}. \quad (17.122)\end{aligned}$$

$$\begin{aligned}\langle\, \widehat{a}^\dagger\vec{\phi} \mid \vec{\phi}'\,\rangle &= \langle \sum_{n=0}^{\infty} c_n\, \sqrt{n+1}\, \vec{\varphi}_{n+1} \mid \sum_{n'=0}^{\infty} c'_{n'}\, \vec{\varphi}_{n'}\rangle \\ &= \sum_{n=0}^{\infty} c_n^*\, c'_{n+1}\, \sqrt{n+1}. \quad (17.123)\end{aligned}$$

It follows that

$$\langle\, \widehat{a}^\dagger\vec{\phi} \mid \vec{\phi}'\,\rangle = \langle \vec{\phi} \mid \widehat{a}\,\vec{\phi}'\,\rangle, \quad (17.124)$$

[69] The notation $\widehat{a}^\dagger$ in Definition 17.10(2) anticipates this result.

showing that $\widehat{a}^\dagger$ is the adjoint of $\widehat{a}$. The operators $\widehat{a}^\dagger$ and $\widehat{a}$ share the same domain but they are not equal.[70] It can be shown that:

(1) Both $\widehat{a}$ and $\widehat{a}^\dagger$ are *closed and irreducible*.[71]

(2) The adjoint $\widehat{a}^{\dagger\dagger}$ of $\widehat{a}^\dagger$ is equal to $\widehat{a}$, i.e., $\widehat{a}^{\dagger\dagger} = \widehat{a}$ on account of Eq. (17.99). In other words $\widehat{a}$ and $\widehat{a}^\dagger$ are the adjoint of each other.

The following properties can be readily verified:

P17.10(1) The operators $\widehat{a}$ and $\widehat{a}^\dagger$ are not selfadjoint.

P17.10(2) The operator $\widehat{N} := \widehat{a}^\dagger\widehat{a}$ admits $\vec{\varphi}_n$ as its eigenvectors corresponding to eigenvalues n.[72]

P17.10(3) The operators $\widehat{a}$, $\widehat{a}^\dagger$ and $\widehat{N}$ satisfy the following commutation relations:[73]

$$[\,\widehat{a},\,\widehat{a}^\dagger\,] = \widehat{I\!I}, \qquad [\,\widehat{a}^\dagger,\,\widehat{a}\,] = -\widehat{I\!I}. \tag{17.125}$$

$$[\,\widehat{a},\,\widehat{N}\,] = \widehat{a}, \qquad [\,\widehat{a}^\dagger,\,\widehat{N}\,] = -\widehat{a}^\dagger. \tag{17.126}$$

The physical reasoning for the terminology will become clear later when the operators are applied to physical problems. Annihilation and creation operators play an important role in quantum theory.[74] A number of applications of annihilation and creation operators will be presented in this book.[75]

---

[70] The annihilation operator is not equal to its adjoint. It is easily verified that $\langle\, \widehat{a}^\dagger\vec{\phi} \mid \vec{\phi}\,\rangle \neq \langle\, \vec{\phi} \mid \widehat{a}\vec{\phi}\,\rangle$.

[71] *See* §20.7 for a proof of irreducibility.

[72] This is called a *number operator*. It is formally defined in Definition 19.1(4). It is also selfadjoint by Theorem 19.1(1). For the domain of $\widehat{N}$ see solution to Q17(5) in Exercises and Problem for Chapter 17.

[73] It is sufficient to verify these commutation relations on the basis vectors $\vec{\varphi}_n$ using Eqs. (17.114), (17.115) and (17.121). Strictly speaking these commutation relations should be written as $[\,\widehat{a},\,\widehat{a}^\dagger\,] \subset \widehat{I\!I}$, $[\,\widehat{a}^\dagger,\,\widehat{a}\,] \subset -\widehat{I\!I}$. Here and elsewhere $\widehat{I\!I}$ will denote the identity operator in the Hilbert space involved.

[74] The operator $\widehat{a}$ is also known as a *lowering operator* since it lowers the numbering of the basis vectors. Its adjoint is also known as a *raising operator* since it does the opposite.

[75] The importance of creation and annihilation operators is such that there is book with the title Creation and Annihilation Operators (by Avery).

Exercises and Problems

Q17(1) Show that every linear operator on a finite-dimensional Hilbert space is bounded.[76]

Q17(2) Show that projectors in a Hilbert space are bounded with norm 1.

Q17(3) Are projectors invertible? Are they reducible?

Q17(4) Verify that the spherical harmonics given by Eqs. (16.64) to (16.67) are eigenfunctions of the operator $\widehat{L}_z(\mathcal{S}_u)$ in Eq. (17.42), i.e.,

$$-i\hbar \frac{\partial Y_{\ell,m_\ell}(\theta,\varphi)}{\partial \varphi} = m_\ell \hbar\, Y_{\ell,m_\ell}(\theta,\varphi), \tag{17.127}$$

or

$$\widehat{L}_z(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} = m_\ell \hbar\, \vec{Y}_{\ell,m_\ell}. \tag{17.128}$$

Q17(5) Prove properties P17.10(1), P17.10(2) and P17.10(3) for a pair of operators $\widehat{a}$ and $\widehat{a}^\dagger$ in Definitions 17.10(1) and 17.10(2).

Q17(6) Find the domain of the operator $\widehat{N} := \widehat{a}^\dagger \widehat{a}$.

Q17(7) Let $\{\vec{\varphi}_n,\, n = 0, 1, 2, \ldots\}$ is an orthonormal basis for a given Hilbert space. Show that the vector

$$\vec{\Phi}_z = \exp\big(-\frac{1}{2}\,|z|^2\big) \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}}\, \vec{\varphi}_n, \quad z \in \mathbb{C}, \tag{17.129}$$

is the eigenvector of the annihilation operator $\widehat{a}$ associated with the orthonormal basis defined by Eqs. (17.114) and (17.115) corresponding to the eigenvalue z. Show also that the vector $\vec{\Phi}_z$ is normalised.
Finally show that

$$\langle \vec{\Phi}_z \mid \widehat{N}\vec{\Phi}_z \rangle = |\, z\,|^2, \tag{17.130}$$

where $\widehat{N} = \widehat{a}^\dagger \widehat{a}$.

[76] Halmos p. 177.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 18

Bounded Operators on $\vec{\mathcal{H}}$

Bounded operators on a Hilbert space $\vec{\mathcal{H}}$ have many of the properties of operators defined on $\vec{\mathbb{V}}^N$. For example, Theorem 13.1(1) and Corollary 13.1(1) apply, e.g., for bounded operators we have

$$\langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle = \langle \vec{\phi} \mid \widehat{B}\vec{\phi} \rangle \ \forall \vec{\phi} \in \vec{\mathcal{H}} \quad \Rightarrow \quad \widehat{A} = \widehat{B}. \tag{18.1}$$

An immediate application is that

$$\langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle = \langle \vec{\phi} \mid \vec{\phi} \rangle \ \forall \vec{\phi} \in \vec{\mathcal{H}} \quad \Rightarrow \quad \widehat{A} = \widehat{I\!I}. \tag{18.2}$$

These results tell us that bounded operators on $\vec{\mathcal{H}}$ can also be characterised by the quadratic forms they generate on $\vec{\mathcal{H}}$. There are important bounded operators on $\mathcal{H}$ which we shall consider in this chapter.

18.1 Selfadjoint Operators and Projectors

Bounded selfadjoint operators are defined by the following selfadjointness condition:

$$\langle \vec{\psi} \mid \widehat{A}\vec{\phi} \rangle = \langle \widehat{A}\vec{\psi} \mid \vec{\phi} \rangle \quad \forall \vec{\psi}, \vec{\phi} \in \vec{\mathcal{H}}. \tag{18.3}$$

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

Many properties of selfadjoint operators on $\vec{\mathbb{V}}^N$ remain true here:

P18.1(1) Definition 13.3.1(1) on positive operators and the result that bounded positive operators are selfadjoint apply here.[1]

P18.1(2) For bounded operators adjoint operation satisfies Eqs. (13.3) and (13.4), i.e., Eqs. (17.100) and (17.101) become equalities. It follows that the sum of two bounded selfadjoint operators is selfadjoint, and the product of two bounded commuting operators is selfadjoint.

Projectors $\widehat{P}$ presented in §16.2.3 are bounded with a norm of 1. They are also selfadjoint since they satisfy the selfadjointness condition in Eq. (18.3), i.e., projectors are bounded selfadjoint operators. Definitions 13.2.2(1) and 13.2.2(2) and Theorem 13.2.2(1) on projectors remain valid.

Let $\{\widehat{P}_{\vec{\varphi}_\ell}\}$ be a complete orthogonal family of projectors on $\vec{\mathcal{H}}$.[2] Following Eq. (9.50) we can construct many new operators by finite linear combinations of these projectors, i.e.,

$$\widehat{A} := \sum_{\ell=1}^{n} a_\ell \, \widehat{P}_{\vec{\varphi}_\ell}, \quad a_\ell \in I\!R. \tag{18.4}$$

These are bounded operators defined on the entire $\vec{\mathcal{H}}$. On account of Eq. (17.95) these operators are also selfadjoint.[3] Intuitively one would expect to be able to extend the above construction to infinite linear combinations, i.e.,

$$\widehat{B} := \sum_{\ell=1}^{\infty} b_\ell \, \widehat{P}_{\vec{\varphi}_\ell}, \quad b_\ell \in I\!R. \tag{18.5}$$

Such an infinite sum of operators can be understood in a similar way an infinite sum of vectors in Eq. (16.45) is defined.

Definition 18.1(1) *A sequence of bounded operators $\widehat{B}_n$ is said to converge to an operator $\widehat{B}$ if the sequence of vectors $\widehat{B}_n\vec{\phi}$ converges to*

[1] Gallone pp. 571–572. Not generally true for unbounded positive operators as remarked after Definition 19.1(3).

[2] *See* Definition 9.3.2(3).

[3] We can also check that they satisfy the selfadjointness condition in Eq. (18.3).

the vector $\widehat{B}\vec{\phi}$ *for every vector* $\vec{\phi}$ *in the Hilbert space*, i.e.,[4]

$$\lim_{n\to\infty} \widehat{B}_n\vec{\phi} = \widehat{B}\vec{\phi} \quad \forall\vec{\phi} \in \vec{\mathcal{H}}. \tag{18.6}$$

The operator $\widehat{B}$ *is said to be the limit of the sequence*.[5]

The relationship above is often denoted by

$$\widehat{B}_n \to \widehat{B} \quad \text{or} \quad \widehat{B} = \lim_{n\to\infty} \widehat{B}_n. \tag{18.7}$$

To understand Eq. (18.5) we can first introduce a sequence of operators $\widehat{B}_n$ by

$$\widehat{B}_n := \sum_{\ell=1}^{n} b_\ell \, \widehat{P}_{\vec{\varphi}_\ell}. \tag{18.8}$$

The operator $\widehat{B}$ is the limit of this sequence.

The discussion on the eigenvalue problem in §8.2.4 and Definition 9.4.4(1) on *eigensubspaces and eigenprojectors in remain valid provided they exist.* Spectral theorems 13.3.2(1) and 13.3.2(2) *remain valid for bounded selfadjoint operators possessing a complete orthonormal set of eigenvectors*.[6] In a finite dimensional vector space selfadjoint operators possess eigenvalues corresponding to a complete orthonormal set of eigenvectors. This is not true in an infinite-dimensional Hilbert space, not even for bounded selfadjoint operators. It follows that eigensubspaces and eigenprojectors based on the concept of eigenvalues and eigenvectors may not exist. Let us illustrate this with two familiar examples below.

E18.1(1) The multiplication operator $\widehat{x}(\Lambda)$ on $\vec{L}^2(\Lambda)$ in Eq. (17.22) is clearly bounded and is also selfadjoint since it satisfies Eq. (18.3). Its eigenvalue equation

$$\widehat{x}(\Lambda)\vec{\varphi} = x_0\vec{\varphi} \quad \text{or} \quad x\varphi(x) = x_0\varphi(x), \quad x \in \Lambda \tag{18.9}$$

[4] *See* Eq. (16.54) for the convergence of a sequence of vectors.

[5] Roman Vol. 2 p. 512. Wan p. 93. There are different kinds of convergence of an operator sequence. The definition here is commonly known as *strong convergence*. An analogy of Cauchy convergence criterion also applies (*see* Weidmann p. 75).

[6] Roman Vol. 2 p. 536.

has a formal solution in terms of a Dirac delta function for any constant $x_0 \in \Lambda$, i.e., we have[7]

$$x\,\delta(x - x_0) = x_0\delta(x - x_0). \tag{18.10}$$

But the delta function is not square-integrable and it does not define a vector in the Hilbert space $\vec{L}^2(\Lambda)$. In other words $\widehat{x}(\Lambda)$ does not possess any eigenvalues and eigenvectors. However, it is useful to refer to x_0 and $\varphi(x)$ in Eq. (18.9) as **generalised eigenvalues** and **generalised eigenfunctions** of $\widehat{x}(\Lambda)$.[8]

E18.1(2) A similar situation also exists for unbounded operators. The operator $\widehat{p}(I\!R)$ in $\vec{L}^2(I\!R)$ defined by Eqs. (17.49) and (17.50) is unbounded. The eigenvalue equation

$$-i\hbar\frac{d}{dx}f_p(x) = p\,f_p(x), \qquad p \in I\!R \tag{18.11}$$

admits the following plane wave solutions:[9]

$$f_p(x) = \frac{1}{\sqrt{2\pi\hbar}}\,e^{\dot{+}px}, \qquad \dot{+} = \frac{i}{\hbar}. \tag{18.12}$$

Plane waves are not square-integrable and hence are not members of $L^2(I\!R)$, i.e., they do not define any vectors in $\vec{L}^2(I\!R)$. We call them *generalised eigenfunctions* of $\widehat{p}(I\!R)$ corresponding to *generalised eigenvalues* p.

In a finite-dimensional vector space the spectral theorem tells us that every selfadjoint operator can be expressed as a linear combination of its eigenprojectors with the corresponding eigenvalues as coefficients. This is no longer true in an infinite-dimensional space. Not having any eigenvalues and eigenvectors the operator $\widehat{x}(\Lambda)$ cannot be written as a linear combination of projectors in the form of Eq. (18.5), i.e., Theorems 13.3.2(1) and (2) do not apply.[10] We shall return to consider these problems in Chapter 20.

[7] Wan p. 126.
[8] Wan p. 126.
[9] These agree with the plane waves in Eq. (10.23) at $t = 0$.
[10] Even if we were to take x_0 as an eigenvalue we still have to deal with the fact that x_0 has a continuous range of values. This would cause problems in defining the sum in Eq. (18.5).

18.2 Density Operators

Let $\widehat{B}$ be a bounded positive operator.[11] Let $\{\vec{\varphi}_\ell\}$ be an orthonormal basis in $\vec{\mathcal{H}}$ and let $\boldsymbol{M}_{\widehat{B}}$ be the matrix representation of $\widehat{B}$ in basis $\{\vec{\varphi}_\ell\}$. It can be proved that the trace of $\boldsymbol{M}_{\widehat{B}}$ is independent of the choice of a basis. A different choice of basis, e.g., $\{\vec{\varphi}'_\ell\}$, would result in a different matrix $\boldsymbol{M}'_{\widehat{B}}$ for the representation of $\widehat{B}$ with same the trace, i.e.,

$$\mathrm{tr}\left(\boldsymbol{M}_{\widehat{B}}\right) := \sum_\ell \langle\vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell\rangle = \sum_\ell \langle\vec{\varphi}'_\ell \mid \widehat{B}\vec{\varphi}'_\ell\rangle = \mathrm{tr}\left(\boldsymbol{M}'_{\widehat{B}}\right). \tag{18.13}$$

This sum may diverge.[12] When it has a finite value the sum turns out to be an important property of the operator. It is called the **trace** of the operator and is denoted by $\mathrm{tr}\,(\widehat{B}\,)$, i.e.,

$$\mathrm{tr}\left(\widehat{B}\,\right) := \sum_\ell \langle\vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell\rangle. \tag{18.14}$$

Bounded positive operators whose trace is equal to 1 are directly relevant to quantum mechanics, as will be seen in Chapter 31.

Definition 18.2(1) *A bounded positive operator* $\widehat{D}$ *of unit trace, i.e.,* $\mathrm{tr}\,(\widehat{D}) = 1$*, is called a* **density operator**.

Density operators possess the following properties[13]:

P18.2(1) Density operators are selfadjoint.

P18.2(2) For a bounded operator $\widehat{B}$ the trace of $\widehat{B}\widehat{D}$ is well-defined and it is also equal to that of $\widehat{D}\widehat{B}$, i.e.,

$$\mathrm{tr}\,(\widehat{B}\widehat{D}) = \mathrm{tr}\,(\widehat{D}\widehat{B}). \tag{18.15}$$

P18.2(3) A one-dimensional projector is a density operator.[14] Since one-dimensional projectors are generated by unit vectors we

[11] *See* P18.1(1).

[12] For the identity operator the sum is infinite.

[13] Reed and Simon Vol. 1 pp. 206–207. Prugovečki pp. 374–392. Jordan pp. 73–78. Beltrametti and Cassinelli pp. 291–292. *See* also Blum's book *Density Matrix Theory and Applications*. Density operators are also referred to as *density matrices*. For clarity we shall call the matrix representation of a density operator a density matrix. Many of the properties listed are obvious and easily proved.

[14] It is easy to check that the trace of a one-dimensional projector is 1 and the trace of a two-dimensional projector is 2 and so on.

have $\widehat{D} = \widehat{P}_{\vec{\varphi}} = |\vec{\varphi}\rangle\langle\vec{\varphi}|$ for some unit vector $\vec{\varphi}$. Let $\{\vec{\varphi}_\ell\}$ be an orthonormal basis with $\vec{\varphi}_1 = \vec{\varphi}$. Then

(1) The matrix representation $\boldsymbol{D}$ of $\widehat{D} = \widehat{P}_{\vec{\varphi}}$ in basis $\{\vec{\varphi}_\ell\}$ is diagonal with its first element D_{11} equal to 1 and all other elements equal to zero.[15] A matrix representation of a density operator is known as a **density matrix**, e.g., $\boldsymbol{D}$ is a density matrix.

(2) Let $\widehat{B}$ be a bounded operator. We can write down the trace of $\widehat{B}\widehat{D}$ directly in terms of $\vec{\varphi}$. Using the basis $\{\vec{\varphi}_\ell\}$ introduced in the item above we can see that

$$\mathrm{tr}\,(\widehat{B}\widehat{D}) = \mathrm{tr}\,(\widehat{B}\widehat{P}_{\vec{\varphi}}) = \langle\vec{\varphi} \mid \widehat{B}\vec{\varphi}\rangle. \tag{18.16}$$

P18.2(4) If a density operator $\widehat{D}$ is a projector then it is a one-dimensional projector, i.e., $\widehat{D} := \widehat{P}_{\vec{\varphi}}$ for some unit vector $\vec{\varphi}$.

P18.2(5) For the sum and the product of two density operators $\widehat{D}_1$ and $\widehat{D}_2$ we have

$$\mathrm{tr}\,(\widehat{D}_1 + \widehat{D}_2) = \mathrm{tr}\,(\widehat{D}_1) + \mathrm{tr}\,(\widehat{D}_2), \tag{18.17}$$

$$\mathrm{tr}\,(\widehat{D}_1\widehat{D}_2) = \mathrm{tr}\,(\widehat{D}_2\widehat{D}_1). \tag{18.18}$$

P18.2(6) For any given positive real number a and two bounded operators $\widehat{B}_1$ and $\widehat{B}_2$ we have

$$\mathrm{tr}\,(a\widehat{D}) = a\,\mathrm{tr}\,(\widehat{D}), \tag{18.19}$$

$$\mathrm{tr}\,\big(\widehat{D}(\widehat{B}_1 + \widehat{B}_2)\big) = \mathrm{tr}\,\big(\widehat{D}\widehat{B}_1\big) + \mathrm{tr}\,\big(\widehat{D}\widehat{B}_2\big). \tag{18.20}$$

P18.2(7) A density operator $\widehat{D}$ can be shown to have a purely discrete set of eigenvalues ω_ℓ. When all the eigenvalues are nondegenerate they satisfy the following properties:[16]

$$0 \le \omega_\ell \le 1 \quad \text{and} \quad \sum_\ell \omega_\ell = 1. \tag{18.21}$$

[15] The matrix is of the same form as the matrix representation of the projector $\boldsymbol{P}_{\vec{\alpha}_z}$ in Eq. (14.7). For notational simplicity for application in Chapter 31 we denote the density matrix by $\boldsymbol{D}$ rather than by $\boldsymbol{M}_{\widehat{D}}$.

[16] Jordan p. 73. All the eigenvalues being nondegenerate means that their corresponding normalised eigenvectors form a complete orthonormal set.

Theorem 13.3.2(1) applies, i.e., we can express $\widehat{D}$ as a linear combination of its orthogonal family of eigenprojectors $\widehat{P}_{\vec{\varphi}_\ell}$ generated by the normalised eigenvector $\vec{\varphi}_\ell$ corresponding to the eigenvalue ω_ℓ, i.e.,[17]

$$\widehat{D} = \sum_\ell \omega_\ell \, \widehat{P}_{\vec{\varphi}_\ell} = \sum_\ell \omega_\ell \, |\vec{\varphi}_\ell\rangle\langle\vec{\varphi}_\ell|, \tag{18.22}$$

The matrix representation $\boldsymbol{D}$ of $\widehat{D}$ in basis $\{\vec{\varphi}_\ell\}$ is diagonal with ω_ℓ as its diagonal elements, i.e.,

$$\boldsymbol{D} = \begin{pmatrix} \omega_1 & 0 & 0 & \cdots \\ 0 & \omega_2 & 0 & \cdots \\ 0 & 0 & \omega_3 & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{pmatrix}. \tag{18.23}$$

P18.2(8) Given an orthonormal basis $\{\vec{\varphi}_\ell\}$ in a Hilbert space together with a set of real numbers w_ℓ satisfying Eq. (18.21) we can construct a density operator using Eq. (18.22).[18]

P18.2(9) Let $\widehat{D}_\ell$, $\ell = 1, 2 \ldots, n$, be a finite set of density operators, and let w_ℓ, $\ell = 1, 2 \ldots, n$ be a set of real numbers satisfying Eq.(18.21). We call the following sum

$$\widehat{D} := \sum_{\ell=1}^{n} w_\ell \, \widehat{D}_\ell \tag{18.24}$$

a **convex combination** of density operators $\widehat{D}_\ell$ with weights w_ℓ. Clearly $\widehat{D}$ is positive, bounded and has a unit trace. In other words, a convex combination of density operators is again a density operator. This remains true when n goes to infinity.[19] Every density operator is a convex combination of its eigenprojectors with its eigenvalues as weights.

[17] We can still express $\widehat{D}$ in the form of Eq.(18.22) even if there are degeneracies. All we need to do is to allow some ω_ℓ to be the same. The operator can also be written explicitly in the form given in Theorem 13.3.2(2).

[18] There is no need for all ω_ℓ to be different.

[19] Beltrametti and Cassinelli p. 6.

P18.2(10) Given a density operator of the form of Eq. (18.24) we have

$$\text{tr}\,(\widehat{B}\widehat{D}) = \sum_{\ell} w_\ell \,\text{tr}\,(\widehat{B}\widehat{D}_\ell) \quad \forall \widehat{B} \in \widehat{\mathcal{B}}(\vec{\mathcal{H}}). \tag{18.25}$$

In particular, if $\widehat{D}_\ell$ are one-dimensional projectors, i.e., $\widehat{D}_\ell = |\vec{\phi}_\ell\rangle\langle\vec{\phi}_\ell|$ we have[20]

$$\text{tr}\,(\widehat{B}\widehat{D}) = \sum_{\ell} w_\ell \,\langle\vec{\phi}_\ell \mid \widehat{B}\vec{\phi}_\ell\rangle, \tag{18.26}$$

showing that $\text{tr}(\widehat{B}\widehat{D})$ is equal to a weighted sum of the quadratic form $\langle\vec{\phi}_\ell \mid \widehat{B}\vec{\phi}_\ell\rangle$ for vectors $\vec{\phi}_\ell$.

P18.2(11) A density operator $\widehat{D}$ assigns a unique set of real numbers to the set $\widehat{\mathcal{B}}(\vec{\mathcal{H}})$ of bounded operators $\widehat{B}$ on $\vec{\mathcal{H}}$, i.e.,

$$\widehat{\mathcal{B}}(\vec{\mathcal{H}}) \mapsto I\!R \quad \text{by} \quad \widehat{B} \mapsto \text{tr}\,(\widehat{B}\widehat{D}). \tag{18.27}$$

Conversely such a set of real numbers determines a unique density operator in the sense that[21]

$$\text{tr}\,(\widehat{B}\widehat{D}) = \text{tr}\,(\widehat{B}\widehat{D}') \quad \forall \widehat{B} \in \widehat{\mathcal{B}}(\vec{\mathcal{H}}) \quad \Leftrightarrow \quad \widehat{D} = \widehat{D}'. \tag{18.28}$$

P18.2(12) Definition 18.2(1) can be applied to define density operators on a finite-dimensional space $\vec{\mathbb{V}}^N$. Density operators in a two-dimensional space $\vec{\mathbb{V}}^2$ are of particular interest. The eigenprojectors $\widehat{P}_{\vec{\alpha}_z}$ and $\widehat{P}_{\vec{\beta}_z}$ of the operator $\widehat{S}_z$ in Eq. (14.4) can be used to construct a density operator in accordance with Eq. (18.24), e.g., we can have

$$\widehat{D}_z = \frac{1}{2}\widehat{P}_{\vec{\alpha}_z} + \frac{1}{2}\widehat{P}_{\vec{\beta}_z}. \tag{18.29}$$

Using the eigenprojectors of the operator $\widehat{S}_x$ given by Eq. (14.27) we can similarly construct a density operator, e.g.,

$$\widehat{D}_x = \frac{1}{2}\widehat{P}_{\vec{\alpha}_x} + \frac{1}{2}\widehat{P}_{\vec{\beta}_x}. \tag{18.30}$$

[20] The projectors are not necessarily orthogonal.

[21] Jordan pp. 75–76. These numbers have to satisfy certain properties to define a density operators.

The two projectors $\widehat{P}_{\vec{\alpha}_z}$ and $\widehat{P}_{\vec{\beta}_z}$ form a complete orthogonal family of projectors on $\vec{\mathbb{V}}^2$. Similarly the projectors $\widehat{P}_{\vec{\alpha}_x}$ and $\widehat{P}_{\vec{\beta}_x}$ also form such a family. It follows that we have

$$\widehat{P}_{\vec{\alpha}_z} + \widehat{P}_{\vec{\beta}_z} = \widehat{\mathbb{I}} = \widehat{P}_{\vec{\alpha}_x} + \widehat{P}_{\vec{\beta}_x}. \tag{18.31}$$

where $\widehat{\mathbb{I}}$ is the identity operator on $\vec{\mathbb{V}}^2$. It follows that

$$\widehat{D}_z = \frac{1}{2}\,\widehat{\mathbb{I}} = \widehat{D}_x. \tag{18.32}$$

Here we have two distinct convex combinations, as expressed in Eqs. (18.29) and (18.30), giving the same density operator. Using the matrix representations of the projectors in Eqs. (14.7) and (14.29) we can see that the matrix representation of both $\widehat{D}_z$ and $\widehat{D}_x$ are the 2×2 identity matrix, apart from the multiplicative constant $\frac{1}{2}$. This non-unigueness is related to the non-uniqueness of spectral decomposition of the identity.

P18.2(13) The above example demonstrates an important result:

> *a density operator may be written as a convex combination of different sets of orthogonal projectors.*

The non-uniqueness of the decomposition of a density operator is even more general than this. In fact a density operator can be decomposed into a convex combination of non-orthogonal projectors.[22] The importance of these results in quantum physics will be discussed in Chapter 31.

18.3 Unitary Operators

Definition 18.3(1)[23] *A bounded operator $\widehat{U}$ on a Hilbert space $\vec{\mathcal{H}}$ with its range $\vec{\mathcal{R}}(\widehat{U})$ coinciding with $\vec{\mathcal{H}}$ is called a unitary operator on $\vec{\mathcal{H}}$ if it preserves the norm of all vectors in $\vec{\mathcal{H}}$, i.e.,*

$$\langle \widehat{U}\,\vec{\phi} \mid \widehat{U}\,\vec{\phi}\,\rangle = \langle \vec{\phi} \mid \vec{\phi}\,\rangle \quad \text{or} \quad ||\widehat{U}\,\vec{\phi}\,|| = ||\vec{\phi}\,|| \qquad \forall\, \vec{\phi} \in \vec{\mathcal{H}}. \tag{18.33}$$

[22] Beltrametti and Cassinelli p. 9.

[23] Akhiezer and Glazman Vol. 1 pp. 72–73. Note that an operator *on* $\vec{\mathcal{H}}$ means the domain of the operator coincides with $\vec{\mathcal{H}}$.

This differs from the Definition 13.4.1(1) in a finite-dimensional space $\vec{\mathbb{V}}^N$ by the additional requirement on the range of the operator. This is because Eq. (13.62), which remains valid here, does not imply the range of the operator would coincide with $\vec{\mathcal{H}}$ when $\vec{\mathcal{H}}$ is infinite-dimensional. Theorems 13.4.1(1) to 13.4.1(3) apply here,[24] e.g., we have

Theorem 18.3(1)[25] *A bounded operator $\widehat{U}$ on a Hilbert space $\vec{\mathcal{H}}$ is unitary if it is invertible with its inverse $\widehat{U}^{-1}$ equal to its adjoint $\widehat{U}^\dagger$, i.e., if*

$$\widehat{U}^\dagger = \widehat{U}^{-1} \quad \text{or equivalently} \quad \widehat{U}^\dagger \widehat{U} = \widehat{U}\,\widehat{U}^\dagger = \widehat{\mathbb{I}}. \tag{18.34}$$

Properties P13.4.1(1) to P13.4.1(4) also apply here. In particular the adjoint of a unitary operator is unitary.[26] Definition 13.4.2(1) can be applied to define **unitary transforms** of vectors and operators in a Hilbert space $\vec{\mathcal{H}}$. These transformations also possess properties P13.4.2(1) to P13.4.2(6).

Let $\vec{\phi}\,'$ be the unitary transform of a unit vector $\vec{\phi}$ generated by a unitary operator $\widehat{U}$, i.e., $\vec{\phi}\,' = \widehat{U}\,\vec{\phi}$. Then the two projectors

$$\widehat{P} = |\vec{\phi}\,\rangle\langle\vec{\phi}\,| \quad \text{and} \quad \widehat{P}' = |\vec{\phi}\,'\rangle\langle\vec{\phi}\,'| = |\widehat{U}\,\vec{\phi}\,\rangle\langle\widehat{U}\,\vec{\phi}\,| \tag{18.35}$$

generated by these two vectors are related by

$$\widehat{P}' = \widehat{U}\,\widehat{P}\,\widehat{U}^\dagger \quad \text{or} \quad |\widehat{U}\,\vec{\phi}\,\rangle\langle\widehat{U}\,\vec{\phi}\,| = \widehat{U}\,\widehat{P}\,\widehat{U}^\dagger, \tag{18.36}$$

since

$$\widehat{P}'\vec{\phi} = \langle\vec{\phi}\,' \mid \vec{\phi}\,\rangle\vec{\phi}\,' = \langle\widehat{U}\,\vec{\phi} \mid \vec{\phi}\,\rangle\,\widehat{U}\,\vec{\phi}, \tag{18.37}$$

$$\widehat{U}\,\widehat{P}\,\widehat{U}^\dagger\vec{\phi} = \widehat{U}\,\langle\vec{\phi} \mid \widehat{U}^\dagger\vec{\phi}\,\rangle\,\vec{\phi} = \langle\widehat{U}\,\vec{\phi} \mid \vec{\phi}\,\rangle\,\widehat{U}\,\vec{\phi}. \tag{18.38}$$

[24] The orthonormal bases in Theorem 13.4.1(3) mean countable orthonormal bases for an infinite-dimensional Hilbert space.

[25] Weidman p. 86. Roman Vol. 2, p. 554. Fano pp. 287–288. The adjoint $\widehat{U}^\dagger$ is a bounded operator by P17.8(3). The fact that $\widehat{U}^\dagger = \widehat{U}^{-1}$ implies that the inverse $\widehat{U}^{-1}$ is bounded and defined on the entire $\vec{\mathcal{H}}$. It follows that the range of $\widehat{U}$, which coincides with the domain of $\widehat{U}^{-1}$, is the entire $\vec{\mathcal{H}}$.

[26] Weidmann pp. 85–86. The inverse of a bounded operator is not necessarily bounded (*see* C17.5(2)).

These results are often symbolically stated in Dirac notation as[27]

$$|\widehat{U}\,\vec{\phi}\,\rangle = \widehat{U}\,|\vec{\phi}\,\rangle, \quad \langle\,\widehat{U}\,\vec{\phi}\,| = \langle\vec{\phi}\,|\,\widehat{U}^{\dagger}, \tag{18.39}$$

which then lead naturally to

$$|\widehat{U}\,\vec{\phi}\,\rangle\langle\widehat{U}\,\vec{\phi}\,| = \widehat{U}\;|\vec{\phi}\,\rangle\langle\vec{\phi}\,|\,\widehat{U}^{\dagger} = \widehat{U}\,\widehat{P}\,\widehat{U}^{\dagger}. \tag{18.40}$$

The expression in Eq. (18.40) can be extended so that for a given operator $\widehat{A}$, not necessarily unitary, and a complex numbers c we have[28]

$$\langle\,c\widehat{A}\vec{\phi}\,| = c^{*}\langle\vec{\phi}\,|\,\widehat{A}^{\dagger}, \tag{18.41}$$

$$|c\widehat{A}\vec{\phi}\,\rangle\langle c\widehat{A}\vec{\phi}\,| = c^{*}c\;\widehat{A}\,|\vec{\phi}\,\rangle\langle\vec{\phi}\,|\,\widehat{A}^{\dagger} = c^{*}c\;\widehat{A}\widehat{P}\;\widehat{A}^{\dagger}. \tag{18.42}$$

Theorem 13.4.3(1) of Stone still applies, subject to certain restrictions to be discussed in §21.1. We shall study a well-known example of unitary transformation in $\vec{L}^2(I\!R)$, i.e., the Fourier transformation, in the following section.

18.4 Fourier Transformations in $\vec{L}^2(I\!R)$

18.4.1 Notation and Preliminaries

Let us examine the properties of the generalised eigenfunctions of the momentum operator $\widehat{p}(I\!R)$,[29] i.e., $f_p(x)$ in Eq. (18.12). Despite not being members of $L^2(I\!R)$ we can still meaningfully carry out a number of formal manipulations on these functions based on the following well-known relations between plane waves and delta functions:

(1) Plane waves are orthonormal in the sense that

$$\int_{-\infty}^{\infty} dx\; f^{*}_{p''}(x)\; f_{p'}(x) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx\; e^{\frac{i}{\hbar}(p'-p'')x} = \delta(p'-p'') = \delta(p''-p'). \tag{18.43}$$

[27] *See* Eqs. (13.21) to (13.23).
[28] Zettili pp. 85–91.
[29] *See* §19.3.4. $\widehat{p}(I\!R)$ is selfadjoint and is a momentum operator in quantum mechanics.

(2) The dependence of $f_p(x)$ on x and p are the same so that we also have

$$\int_{-\infty}^{\infty} dp\, f_p^*(x'')\, f_p(x') = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dp\, e^{\frac{i}{\hbar}(x'-x'')p}$$
$$= \delta(x' - x'') = \delta(x'' - x'). \quad (18.44)$$

Given a function $\varphi(x)$ on $I\!R$ we define its **Fourier transform** by an integral, known as its *Fourier integral*, i.e.,

$$\underset{\sim}{\varphi}(p) := \int_{-\infty}^{\infty} dx\, f_p^*(x)\, \varphi(x)$$
$$= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx\, e^{-\frac{i}{\hbar}px}\, \varphi(x). \quad (18.45)$$

The above definition makes sense only for functions for which the above integral exist. There are dense subsets of $L^2(I\!R)$, e.g., $C_c^\infty(I\!R)$, for which the Fourier integrals clearly exist since every function in $C_c^\infty(I\!R)$ vanishes outside a bounded interval. It is possible to define the Fourier transform of an arbitrary function in $L^2(I\!R)$ in two steps: (1) construct integrals of the form given in Eq. (18.45) over a sequence of finite intervals which diverges to the infinity interval $(-\infty, \infty)$, and then (2) take an appropriate limit of the these integrals. However, we shall not delve into the mathematical subtlety in the definition of the Fourier transforms of an arbitrary function in $L^2(I\!R)$.[30] We shall be content with formal manipulations using delta functions to demonstrate various properties of Fourier transform operation using Eq. (18.45), with the knowledge that rigorous proofs do exist.

The Fourier transform of $\varphi(x)$ is a function of p. We can introduce the concept of the *inverse Fourier transform operation* by introducing the **inverse Fourier transform** of $\underset{\sim}{\varphi}(p)$ as a function of x defined by the integral

$$\int_{-\infty}^{\infty} dp\, f_p(x)\, \underset{\sim}{\varphi}(p). \quad (18.46)$$

[30] Roman Vol. 2 pp. 557–559.

Performing the integration we get

$$\begin{aligned}
&\int_{-\infty}^{\infty} dp\ f_p(x)\,\underset{\sim}{\varphi}(p) \\
&= \int_{-\infty}^{\infty} dp\ f_p(x)\ \left(\int_{-\infty}^{\infty} f_p^*(x')\,\varphi(x')\,dx'\right) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} dp\ f_p(x)\ f_p^*(x')\right)\ \varphi(x')\,dx' \\
&= \int_{-\infty}^{\infty} dx'\,\delta(x-x')\varphi(x') = \varphi(x). \qquad (18.47)
\end{aligned}$$

So, the Fourier transform of $\varphi(x)$ is

$$\underset{\sim}{\varphi}(p) := \int_{-\infty}^{\infty} dx\ f_p^*(x)\,\varphi(x), \qquad (18.48)$$

and the inverse Fourier transform of $\underset{\sim}{\varphi}(p)$ is

$$\varphi(x) := \int_{-\infty}^{\infty} dp\ f_p(x)\,\underset{\sim}{\varphi}(p). \qquad (18.49)$$

By considering the integrals in Eqs. (18.43) and (18.45) as formal scalar products with the notation $\langle f_{p''} \mid f_{p'}\rangle$ and $\langle f_p \mid \varphi\rangle$ we get

$$\langle f_{p''} \mid f_{p'}\rangle = \delta(p''-p'). \qquad (18.50)$$

$$\underset{\sim}{\varphi}(p) = \langle f_p \mid \varphi\rangle \qquad (18.51)$$

and

$$\varphi(x) = \int_{-\infty}^{\infty} dp\,\underset{\sim}{\varphi}(p)\ f_p(x), \quad \underset{\sim}{\varphi}(p) = \langle f_p \mid \varphi\rangle. \qquad (18.52)$$

We can gain an intuitive understanding of the Fourier transform and the inverse Fourier transform operation as follows:

(1) When the eigenvectors of an operator form an orthonormal basis we can express an arbitrary vector in terms of such a set of eigenvectors using Eq. (16.68).

(2) Not being a discrete set the generalised eigenfunctions $f_p(x)$ of the momentum operator do not form an orthonormal basis. It follows that Eq. (16.68) cannot be applied to obtain an expansion of an arbitrary vector in terms of $f_p(x)$.

(3) The inverse Fourier transform defined by Eq. (18.52) does resemble Eq. (16.68), if we formally consider

   (a) the plane waves $f_p(x)$ as forming a "continuous orthonormal basis" and,
   (b) Eq. (18.52) as an expansion of $\phi(x)$ in the "continuous orthonormal basis" $\{f_p(x),\ p \in I\!R\}$ in the form of an integral.

It follows that conceptually the Fourier transform $\underset{\sim}{\varphi}(p)$ is just the coefficient of expansion of $\varphi(x)$ in terms of the generalised eigenfunctions $f_p(x)$ of the momentum operator and the inverse Fourier transform is just the expansion of $\phi(x)$ in terms of $f_p(x)$.[31]

18.4.2 Fourier transform as unitary transform

18.4.2.1 Coordinate space and momentum space

The Fourier transform $\underset{\sim}{\varphi}(p)$ of a square-integrable function $\varphi(x)$ is square-integrable with respect to p over the range $(-\infty, \infty)$ since

$$\int_{-\infty}^{\infty} dp\, \underset{\sim}{\varphi}^*(p)\underset{\sim}{\varphi}(p) = \int_{-\infty}^{\infty} dx\, \varphi^*(x)\varphi(x). \tag{18.53}$$

This result can be formally verified using the properties of delta functions. Mathematically we can regard $\varphi(x)$ and $\underset{\sim}{\varphi}(p)$ to be members of the same space $L^2(I\!R)$ since x in $\varphi(x)$ and p in $\underset{\sim}{\varphi}(p)$ are dummy variables so that both $\varphi(x)$ and $\underset{\sim}{\varphi}(p)$ are just square-integrable functions on the real line. In terms of our notation which formally distinguish a function from a vector we shall denote the vectors in $\vec{L}^2(I\!R)$ corresponding to $\varphi(x)$ and $\underset{\sim}{\varphi}(p)$ by $\vec{\varphi}$ and $\underset{\sim}{\vec{\varphi}}$, respectively. The Fourier transform operation can then be regarded as effected by an operator on $\vec{L}^2(I\!R)$, i.e., we can introduce an

[31] Some authors even refer to the coefficients c_ℓ in Eq. (16.68) as Fourier coefficients (*see* Prugovečki p. 38).

operator $\widehat{U}_F$ by[32]

$$\underset{\sim}{\vec{\varphi}} = \widehat{U}_F \vec{\varphi} \quad \forall \vec{\varphi} \in \vec{L}^2(I\!R). \tag{18.54}$$

Every vector $\underset{\sim}{\vec{\varphi}} \in \vec{L}^2(I\!R)$ also has a vector $\vec{\varphi}$ associated with it through the inverse Fourier transform, i.e., the range of the transformation coincides with $\vec{L}^2(I\!R)$. In addition we have the preservation of the norm shown in Eq. (18.53). All this means that $\widehat{U}_F$ is unitary. In other words, the Fourier transform operation is a unitary transformation of $\vec{L}^2(I\!R)$. Being unitary the operator is invertible. The inverse operator $\widehat{U}_F^{-1}$ then defines the inverse transform, i.e.,

$$\vec{\varphi} = \widehat{U}_F^{-1} \underset{\sim}{\vec{\varphi}}. \tag{18.55}$$

In physical applications it is more intuitive to consider $\varphi(x)$ and $\underset{\sim}{\varphi}(p)$ as members of two separate spaces in the following sense:

(1) Consider the set $\{x \in (-\infty, \infty)\}$ of coordinate values as forming a space referred to as a **coordinate space** and denoted simply by $I\!R$.[33] Then $\varphi(x)$ is a function on the coordinate space, and $L^2(I\!R)$ is the set of square-integrable functions on the coordinate space $I\!R$. The vector space corresponding to $L^2(I\!R)$ is denoted by $\vec{L}^2(I\!R)$.

(2) Consider the set $\{p \in (-\infty, \infty)\}$ of values as forming a space referred to as a **momentum space** and denoted by $\underset{\sim}{I\!R}$, since p represents the generalised eigenvalues of the momentum operator $\widehat{p}(I\!R)$.[34] Then $\underset{\sim}{\varphi}(p)$ is a function on the momentum space and $L^2(\underset{\sim}{I\!R})$ is the set of square-integrable functions on the momentum space $\underset{\sim}{I\!R}$, i.e.,

$$L^2(\underset{\sim}{I\!R}) = \left\{ \underset{\sim}{\varphi}(p) : \int_{-\infty}^{\infty} |\underset{\sim}{\varphi}(p)|^2 \, dp < \infty \right\}. \tag{18.56}$$

The vector space corresponding to $L^2(\underset{\sim}{I\!R})$ is denoted by $\vec{L}^2(\underset{\sim}{I\!R})$.

(3) The two spaces $\vec{L}^2(I\!R)$ and $\vec{L}^2(\underset{\sim}{I\!R})$ are isomorphic. They are unitarily related by Eqs. (18.54) and (18.55).

[32] Prugovečki Theorem 4.5 on p. 219.
[33] Here the term "space" means a set.
[34] Again the term "space" means a set.

As an application consider a quantum particle in one-dimensional motion. In the usual description a quantum state φ^s is described by a vector $\vec{\varphi}$ defined by a square-integrable function $\varphi(x)$ of the position variable x. These functions are also known as *wave functions*. Such a description is known as a **coordinate representation**. The position operator takes the form a multiplication operator, i.e., $\widehat{x}(I\!R)$ defined by Eqs. (17.12) and (17.13), while the momentum operator acting on the wave function $\varphi(x)$ takes the form of a differential operator, i.e., $\widehat{p}(I\!R)$ defined by Eqs. (17.49) and (17.50). In view of the one-to-one correspondence between functions $\varphi(x)$ and their Fourier transforms $\underset{\sim}{\varphi}(p)$ we can also describe a quantum state in terms of a function of p, i.e., using the Fourier transform $\underset{\sim}{\varphi}(p)$ instead of $\varphi(x)$. Such a description of quantum states is known as a **momentum representation**. In such a momentum representation a state corresponds to a function of the momentum variable p, namely $\vec{\underset{\sim}{\varphi}} := \underset{\sim}{\varphi}(p)$. We shall call $\vec{L}^2(I\!R)$ a **coordinate representation space** and $\vec{L}^2(\underset{\sim}{I\!R})$ a **momentum representation space**. Operators in the momentum representation space are discussed in the next section.

18.4.2.2 Fourier transforms of operators

Fourier transform of operators are performed in the same way as unitary transform of operators as seen in the following examples:

E18.4.2.2(1) Position operator The Fourier transform of the position operator $\widehat{x}(I\!R)$ in the coordinate representation space $\vec{L}^2(I\!R)$ is the operator in the momentum representation space $\vec{L}^2(\underset{\sim}{I\!R})$ defined by

$$\widehat{\underset{\sim}{x}}(\underset{\sim}{I\!R}) := \widehat{U}_F\, \widehat{x}(I\!R)\, \widehat{U}_F^{-1}, \tag{18.57}$$

acting on the domain $\vec{\mathcal{D}}(\widehat{\underset{\sim}{x}}(\underset{\sim}{I\!R}))$ given by the Fourier transform of the domain $\vec{\mathcal{D}}(\widehat{x}(I\!R))$ of $\widehat{x}(I\!R)$. Acting on $\vec{\underset{\sim}{\varphi}} \in \vec{\mathcal{D}}(\widehat{\underset{\sim}{x}}(\underset{\sim}{I\!R}))$ the operator $\widehat{\underset{\sim}{x}}(\underset{\sim}{I\!R})$ takes the form of a differential operator, i.e.,[35]

$$\widehat{\underset{\sim}{x}}(\underset{\sim}{I\!R})\vec{\underset{\sim}{\varphi}} = \widehat{U}_F\, \widehat{x}(I\!R)\, \widehat{U}_F^{-1}\vec{\underset{\sim}{\varphi}} := i\hbar \frac{d}{dp}\, \underset{\sim}{\varphi}(p). \tag{18.58}$$

[35] There is no minus sign in front of $i\hbar$.

This can be shown as follows:

$$\begin{aligned}
\widehat{\underset{\sim}{x}}(\underset{\sim}{I\!R})\vec{\underset{\sim}{\varphi}} &= \left(\widehat{U}_F\,\hat{x}(I\!R)\,\widehat{U}_F^{-1}\right)\widehat{U}_F\vec{\varphi} \\
&= \widehat{U}_F\,\hat{x}(I\!R)\left(\widehat{U}_F^{-1}\widehat{U}_F\right)\vec{\varphi} = \widehat{U}_F\left(\hat{x}(I\!R)\,\vec{\varphi}\right) \\
&:= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} e^{-\frac{i}{\hbar}xp}\left(x\varphi(x)\right)dx \\
&= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} \frac{\hbar}{-i}\frac{d}{dp}\left(e^{-\frac{i}{\hbar}xp}\right)\varphi(x)dx \\
&= i\hbar\frac{d}{dp}\left\{\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} e^{-\frac{i}{\hbar}xp}\,\varphi(x)dx\right\} \\
&= i\hbar\frac{d}{dp}\,\underset{\sim}{\varphi}(p). \qquad (18.59)
\end{aligned}$$

E18.4.2.2(2) Momentum operator The Fourier transform of the momentum operator $\widehat{p}(I\!R)$ in the coordinate representation space $\vec{L}^2(I\!R)$ is the operator $\widehat{\underset{\sim}{p}}(\underset{\sim}{I\!R})$ in the momentum representation space $\vec{L}^2(\underset{\sim}{I\!R})$ given by

$$\widehat{\underset{\sim}{p}}(\underset{\sim}{I\!R}) := \widehat{U}_F\;\widehat{p}(I\!R)\,\widehat{U}_F^{-1} \qquad (18.60)$$

acting on the domain $\vec{\mathcal{D}}(\widehat{\underset{\sim}{p}}(\underset{\sim}{I\!R}))$ given by the Fourier transform of the domain $\vec{\mathcal{D}}(\widehat{p}(I\!R))$ of $\widehat{p}(I\!R)$. On $\vec{\underset{\sim}{\varphi}} \in \vec{\mathcal{D}}(\widehat{\underset{\sim}{p}}(\underset{\sim}{I\!R}))$ the operator $\widehat{\underset{\sim}{p}}$ takes the form of a multiplication operator, i.e.,

$$\widehat{\underset{\sim}{p}}(\underset{\sim}{I\!R})\vec{\underset{\sim}{\varphi}} = \widehat{U}_F\;\widehat{p}(I\!R)\,\widehat{U}_F^{-1}\vec{\underset{\sim}{\varphi}} := p\,\underset{\sim}{\varphi}(p). \qquad (18.61)$$

We can verify this result by first noting that

$$\widehat{\underset{\sim}{p}}(\underset{\sim}{I\!R})\widetilde{\varphi} = \left(\widehat{U}_F\;\widehat{p}(I\!R)\,\widehat{U}_F^{-1}\right)\widehat{U}_F\vec{\varphi} = \widehat{U}_F\;\widehat{p}(I\!R)\vec{\varphi}. \qquad (18.62)$$

The expression on the right hand side is defined by the function

$$\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx\, e^{-\frac{i}{\hbar}xp}\Big(-i\hbar\frac{d\varphi(x)}{dx}\Big). \qquad (18.63)$$

We can evaluate the integral as follows[36]:

$$
\begin{aligned}
&-i\hbar\,\frac{1}{\sqrt{2\,\pi\,\hbar}}\left\{\left[e^{-\frac{i}{\hbar}xp}\,\varphi(x)\right]_{-\infty}^{\infty}-\int_{-\infty}^{\infty}\frac{de^{-\frac{i}{\hbar}xp}}{dx}\,\varphi(x)dx\right\}\\
&=-i\hbar\,\frac{1}{\sqrt{2\,\pi\,\hbar}}\left\{-\int_{-\infty}^{\infty}\frac{-ip}{\hbar}e^{-\frac{i}{\hbar}xp}\,\varphi(x)dx\right\}\\
&=p\,\frac{1}{\sqrt{2\,\pi\,\hbar}}\int_{-\infty}^{\infty}e^{-\frac{i}{\hbar}xp}\,\varphi(x)dx=p\,\underset{\sim}{\varphi}(p).
\end{aligned}
\tag{18.64}
$$

Being unitary a simultaneous Fourier transformation of operators and vectors preserves quadratic forms and commutation relations:

(1) The quadratic forms generated by $\widehat{x}(I\!R)$ and $\widehat{p}(I\!R)$ are preserved, i.e.,

$$\langle\vec{\varphi}\mid\widehat{x}(I\!R)\vec{\varphi}\,\rangle=\langle\underset{\sim}{\vec{\varphi}}\mid\underset{\sim}{\widehat{x}}(I\!R)\underset{\sim}{\vec{\varphi}}\,\rangle,\tag{18.65}$$

$$\langle\vec{\varphi}\mid\widehat{p}(I\!R)\vec{\varphi}\,\rangle=\langle\underset{\sim}{\vec{\varphi}}\mid\underset{\sim}{\widehat{p}}(I\!R)\underset{\sim}{\vec{\varphi}}\,\rangle.\tag{18.66}$$

Explicitly we have

$$\int_{-\infty}^{\infty}\varphi^{*}(x)x\varphi(x)\,dx=\int_{-\infty}^{\infty}\underset{\sim}{\varphi}^{*}(p)\left(i\hbar\frac{d}{dp}\right)\underset{\sim}{\varphi}(p)\,dp,\tag{18.67}$$

$$\int_{-\infty}^{\infty}\varphi^{*}(x)\left(-i\hbar\frac{d}{dx}\right)\varphi(x)\,dx=\int_{-\infty}^{\infty}\underset{\sim}{\varphi}^{*}(p)\;p\,\underset{\sim}{\varphi}(p)\,dp.\tag{18.68}$$

(2) Commutation relations are preserved, i.e., as in Eq. (13.74) we have

$$[\widehat{x}(I\!R),\,\widehat{p}(I\!R)]=[\underset{\sim}{\widehat{x}}(I\!R),\,\underset{\sim}{\widehat{p}}(I\!R)]=i\hbar.\tag{18.69}$$

Exercises and Problems

Q18(1) The characteristic function $\chi_{\Lambda}(x)$ of the interval Λ in $I\!R$ defines a multiplication operator $\widehat{\chi}_{\Lambda}$ on $\vec{L}^{2}(I\!R)$ by Eq. (17.11). Show that $\widehat{\chi}_{\Lambda}$ is a projector. Find the eigenvalues and their corresponding eigenvectors of $\widehat{\chi}_{\Lambda}$.

[36] Note that $\varphi(\infty)=\varphi(-\infty)=0$.

Q18(2) Show that the multiplication operator $\widehat{x}(\Lambda)$ on $\vec{L}^2(\Lambda)$ in Eq. (17.22) is bounded and selfadjoint.

Q18(3) Show that plane waves $f_p(x)$ in Eq. (18.12) is not square-integrable over $I\!R$.

Q18(4) Show that

(1) The trace of a projector is equal to the dimension of the subspace onto which the projector projects.
(2) If a density operator is a projector then it is a one-dimensional projector.

Q18(5) Show that a convex combination of density operators as shown in Eq. (18.24) defines a density operator.

Q18(6) Show that the density operator $\widehat{D}_z$ in Eq. (18.29) can be decomposed in terms of the eigenprojectors $\widehat{P}_{\vec{\alpha}_y}$ and $\widehat{P}_{\vec{\beta}_y}$ of $\widehat{S}_y$ shown in Eq. (14.38) as

$$\widehat{D}_z = \frac{1}{2}\widehat{P}_{\vec{\alpha}_y} + \frac{1}{2}\widehat{P}_{\vec{\beta}_y}. \tag{18.70}$$

Q18(7) Find the matrix representations (density matrices) in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ of the density operators $\widehat{D}_z$ and $\widehat{D}_x$ in Eqs (18.29) and (18.30).

Q18(8) Show that unitary operators on a Hilbert space are invertible.

Q18(9) Show that a unitary transformation preserves the trace of the product operator in Eq. (18.15), i.e.,

$$\mathrm{tr}\big(\widehat{B}\,\widehat{D}\big) = \mathrm{tr}\big(\widehat{B}'\widehat{D}'\big), \tag{18.71}$$

where $\widehat{B}'$ and $\widehat{D}'$ are the unitary transforms of $\widehat{B}$ and $\widehat{D}$, respectively generated by a unitary operator $\widehat{U}$ in accordance with Definition 13.4.2(1).

Q18(10) Show that the unitary transform of a one-dimensional projector is again a one-dimensional projector.

Q18(11) Verify Eq. (18.53).

Q18(12) Using the expression p and $i\hbar d/dp$ for $\underset{\sim}{\widehat{x}}(I\!R)$ and $\underset{\sim}{\widehat{p}}(I\!R)$ verify the commutation relation in Eq. (18.69).

Q18(13) The Fourier transform $\vec{\varphi}$ of $\vec{\varphi} \in \vec{L}^2(I\!R)$ is given by the following characteristic function on the momentum space:

$$\underset{\sim}{\varphi}(p) = \begin{cases} 0, & p \leq -p_0, \\ 1/\sqrt{2\,p_0}, & p \in (-p_0,\, p_0], \\ 0, & p \geq p_0, \end{cases} \tag{18.72}$$

where p_0 is a real and positive constant. By performing the inverse Fourier transform show that the function $\varphi(x) \in L^2(I\!R)$ corresponding to the vector $\vec{\varphi}$ is given by

$$\varphi(x) = \sqrt{\frac{\hbar}{\pi\, p_0}}\, \frac{\sin(p_0 x/\hbar)}{x}. \tag{18.73}$$

Chapter 19

Symmetric and Selfadjoint Operators in $\vec{\mathcal{H}}$

19.1 Definitions and Examples

Definition 19.1(1)[1] *An operator $\widehat{A}$ defined on a dense domain $\vec{\mathcal{D}}(\widehat{A})$ of a Hilbert space $\vec{\mathcal{H}}$ is said to be* **symmetric** *if it satisfy the following symmetry condition*

$$\langle \vec{\varphi} \mid \widehat{A}\vec{\phi} \rangle = \langle \widehat{A}\vec{\varphi} \mid \vec{\phi} \rangle \quad \forall\, \vec{\varphi},\ \vec{\phi} \in \mathcal{D}(\widehat{A}). \tag{19.1}$$

By comparing Eqs. (17.91) and (19.1) we can conclude that for $\widehat{A}$ to be symmetric we must have

(1) the domain of its adjoint $\widehat{A}^\dagger$ bigger than or at least equal to that of $\widehat{A}$, i.e., $\vec{\mathcal{D}}(\widehat{A}^\dagger) \supset \vec{\mathcal{D}}(\widehat{A})$, and
(2) its adjoint $\widehat{A}^\dagger$ agreeing with $\widehat{A}$ on $\vec{\mathcal{D}}(\widehat{A})$.

In other words, we can say that

> *an operator $\widehat{A}$ is symmetric if it is a restriction of its adjoint $\widehat{A}^\dagger$, i.e., if $\widehat{A} \subset \widehat{A}^\dagger$, or we can say that an operator is symmetric if its adjoint is an extension of itself.*

[1] Akhiezer and Glazman p. 85.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

Many differential operators are symmetric. Some explicit examples will be presented in §19.2. In $\vec{\mathbb{V}}^N$ all the operators and their adjoints are bounded and defined on the entire space. There is not need to bring in a separate concept of symmetric operators since the definition would render the operators selfadjoint.[2]

Symmetric operators have the following properties[3]:

P19.1(1) If $\widehat{A}$ is symmetric then the quadratic form it generates on its domain $\vec{\mathcal{D}}(\widehat{A})$ is real-valued, i.e.,

$$\langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle \in I\!R \quad \forall \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{19.2}$$

P19.1(2) The converse is also true. i.e., an operator $\widehat{A}$ defined on a dense domain $\vec{\mathcal{D}}(\widehat{A})$ is symmetric if the quadratic form it generates on its domain $\vec{\mathcal{D}}(\widehat{A})$ is real-valued.

P19.1(3) The eigenvalues of a symmetric operators may not exist. If they do then they are real and eigenvectors corresponding to different eigenvalues are orthogonal to each other.

The concept of positive operators introduced by Definition 13.3.1(1) on $\vec{\mathbb{V}}^N$ can be generalised to an infinite-dimensional Hilbert space $\vec{\mathcal{H}}$.

Definition 19.1(2) *An operator $\widehat{A}$ in $\vec{\mathcal{H}}$ is said to be* **positive** *if the quadratic form it generates on $\vec{\mathcal{D}}(\widehat{A})$ is real-valued and non-negative, i.e.,*

$$\langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle \in I\!R \quad \text{and} \quad \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle \geq 0 \quad \forall \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{19.3}$$

Using property CSP11.2.2(1) of scalar product we can show that positive operators are symmetric, e.g., we have

$$\langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle = \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle^* = \langle \widehat{A}\vec{\phi} \mid \vec{\phi} \rangle \quad \forall \vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}). \tag{19.4}$$

But positive operators in $\vec{\mathcal{H}}$ are not necessarily selfadjoint.[4] More examples of symmetric operators are given in §19.2.

[2] By comparing Eqs. (19.1) and (13.1).
[3] Riesz and Nagy pp. 229–230. Weidman p. 72. Wan p. 99.
[4] In $\vec{\mathbb{V}}^N$ a positive operator is selfadjoint, as stated in Theorem 13.3.1(1).

Definition 19.1(3)[5] *An operator* $\widehat{A}$ *defined on a dense domain* $\vec{\mathcal{D}}(\widehat{A})$ *of a Hilbert space* $\vec{\mathcal{H}}$ *is said to be* **selfadjoint** *if it is equal to its adjoint.*[6]

For an operator to be selfadjoint we must have $\vec{\mathcal{D}}(\widehat{A}^\dagger) = \vec{\mathcal{D}}(\widehat{A})$ and on this common domain $\widehat{A}$ must agree with its adjoint $\widehat{A}^\dagger$. The position operators $\hat{x}(I\!R^3)$, $\hat{y}(I\!R^3)$ and $\hat{z}(I\!R^3)$ are familiar examples of unbounded selfadjoint operators in $\vec{L}^2(I\!R^3)$. Multiplications operators given by real-valued functions of the position variables are selfadjoint in $\vec{L}^2(I\!R^3)$. Some differential operators in $\vec{L}^2(I\!R^3)$ are selfadjoint, e.g., the momentum operator $\widehat{p}(I\!R)$ in Eq. (17.50) is selfadjoint. More examples will be presented in §19.3.

The following properties are obvious:

P19.1(4) Unbounded selfadjoint operators are also symmetric since they satisfy the symmetric condition in Eq. (19.1).

P19.1(5) Bounded symmetric operators are selfadjoint. It follows that bounded positive operators are selfadjoint.[7]

P19.1(6) Because of Eqs. (17.100) and (17.101) the sum of unbounded selfadjoint operators and the product of two commuting unbounded selfadjoint operators are not necessarily selfadjoint.[8]

Selfadjointness is related to closedness as stated below.[9]

Theorem 19.1(1)[10]

(1) *A selfadjoint operator is closed.*
(2) *If* $\widehat{A}$ *is a closed operator then the product* $\widehat{A}^\dagger \widehat{A}$ *is selfadjoint.*

This theorem tells us that the square of a selfadjoint operator is selfadjoint since a selfadjoint operator is closed. Another important application relates to annihilation and creation operators. From §17.10 we know that a pair of annihilation and creation operators

[5] *See* Definition 9.4.1(1).
[6] Unbounded selfadjoint operators do not satisfy the selfadjointness condition of bounded selfadjoint operators given by Eq. (18.3).
[7] Gallone pp. 571–572.
[8] See §19.5 for some relevant results.
[9] *See* Definition 17.1(3) and P17.8(8) on closed operators.
[10] Akhiezer and Glazman p. 97.

$\widehat{a}$ and $\widehat{a}^\dagger$ can be defined in terms of an orthonormal basis $\{\vec{\varphi}_n, n = 0, 1, 2 \ldots\}$ in a Hilbert space $\vec{\mathcal{H}}$. Both these operators are closed but not selfadjoint. Their product defined below has many applications.

Definition 19.1(4) *Let $\widehat{a}$ and $\widehat{a}^\dagger$ be the annihilation and creation operators associated with a given orthonormal basis $\{\vec{\varphi}_n, n = 0, 1, 2 \ldots\}$. The operator*

$$\widehat{N} = \widehat{a}^\dagger \widehat{a} \tag{19.5}$$

is called the **number operator** *associated with the annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$.*

This number operator is selfadjoint by Theorem 19.1 (1). A striking feature of this operator, as suggested by its name, is that it admits 0 and positive integers as its eigenvalues with the basis vectors as eigenvectors, i.e., we have

$$\widehat{N}\vec{\varphi}_n = n\,\vec{\varphi}_n, \quad n = 0, 1, 2, \ldots. \tag{19.6}$$

There are no other eigenvalues or eigenvectors, since $\{\vec{\varphi}_n\}$ is a complete orthonormal set.[11] There are many useful operators whose eigenvalues are lowered-bounded. We shall give a definition of such operators.

Definition 19.1(5)[12] *A selfadjoint operator $\widehat{A}$ is said to be* **bounded below** *if there exists a finite real number M such that the quadratic form $\mathcal{Q}(\widehat{A}, \vec{\phi}) := \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle \geq M\, ||\vec{\phi}||^2$ for all $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$.*

Many physically important selfadjoint operators are bounded below.[13]

Symmetric operators are not necessarily closed. But they are *closable* in the sense defined below.

Definition 19.1(6)[14] *An operator $\widehat{A}$ is said to be* **closable** *if it has extensions which are closed. The smallest extension is called the* **closure** *of the operator which is denoted by $\bar{A}$.*

---

[11] *See* P20.3(5). A new eigenvalue would correspond to a new eigenvector orthogonal to all $\vec{\varphi}_n$ but no non-zero vector can be orthogonal to all $\vec{\varphi}_n$.

[12] Amrein, Jauch and Sinha p. 319.

[13] These operators have some special properties as illustrated in Theorem 24.2.1(2).

[14] Akhiezer and Glazman p. 78. Wan p. 89.

Theorem 19.1(2)[15] *If $\widehat{A}$ is symmetric then its closure exists and the closure is equal to the adjoint of the adjoint of $\widehat{A}$, i.e.,*

$$\bar{A} = \widehat{A}^{\dagger\dagger}, \tag{19.7}$$

and we also have

$$\widehat{A}^{\dagger} = \widehat{A}^{\dagger\dagger\dagger}. \tag{19.8}$$

19.2 Symmetric Differential Operators

Differential operators have been introduced in §17.3. As seen in Eq. (17.28) the usual procedure is to specify an operator expression, i.e., a differential expression, and an appropriate set of differentiable functions as the domain for the operator expression to act on. Quite often the specification of an intuitively obvious domain would lead to a symmetric operator which is not selfadjoint. The adjoint operator would often have the same differential expression but acting on a bigger domain. We shall illustrate this with several examples in the following subsections.[16]

19.2.1 In $\vec{L}^2(\Lambda),\ \Lambda = [0, L]$

The operator $\widehat{p}_D(\Lambda)$ defined by Eq. (17.32) is symmetric, since for all $\vec{\varphi}, \vec{\phi} \in \vec{\mathcal{D}}(\widehat{p}_D(\Lambda))$ we have

$$\langle \vec{\varphi} \mid \widehat{p}_D(\Lambda)\vec{\phi} \rangle := -i\hbar \int_0^L \varphi(x)^* \frac{d\phi(x)}{dx}\, dx. \tag{19.9}$$

Using integration by parts the integral becomes

$$\begin{aligned} &-i\hbar \left\{ \varphi^*(x)\phi(x)\Big|_0^L - \int_0^L \frac{d\varphi^*(x)}{dx}\phi(x)\, dx \right\} \\ &= i\hbar \int_0^L \frac{d\varphi^*(x)}{dx}\phi(x)\, dx = \int_0^L \Big(-i\hbar \frac{d\varphi(x)}{dx} \Big)^* \phi(x)\, dx \\ &= \langle \widehat{p}_D(\Lambda)\vec{\varphi} \mid \vec{\phi} \rangle. \end{aligned} \tag{19.10}$$

It follows that its adjoint $\widehat{p}_D^{\dagger}(\Lambda)$ would have a domain $\vec{\mathcal{D}}\big(\widehat{p}_D^{\dagger}(\Lambda)\big)$ bigger than or at least equal to $\vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$. First we must find

[15] Prugovečki pp. 355–357. Simon and Reed Vol. 1 p. 253. *See* Eqs. (17.98) and (17.99).
[16] Akhiezer and Glazman Vol. 1 pp. 106–111. Roman Vol. 2 pp. 544–549.

out whether $\vec{\mathcal{D}}\big(\widehat{p}_D^{\,\dagger}(\Lambda)\big)$ is actually bigger than $\vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$. If so we must then find the operator expression of $\widehat{p}_D^{\,\dagger}(\Lambda)$ for acting on $\vec{\mathcal{D}}\big(\widehat{p}_D^{\,\dagger}(\Lambda)\big) - \vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$. We cannot simply assume a differential expression since the part of $\vec{\mathcal{D}}\big(\widehat{p}_D^{\,\dagger}(\Lambda)\big)$ lying outside $\vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$ may contain non-differentiable functions.

Let $\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$ and $\vec{\psi} \in \vec{\mathcal{D}}\big(\widehat{p}_D^{\,\dagger}(\Lambda)\big)$. We have

$$\langle \vec{\psi} \mid \widehat{p}_D(\Lambda)\vec{\phi}\,\rangle = \langle \widehat{p}_D^{\,\dagger}(\Lambda)\vec{\psi} \mid \vec{\phi}\,\rangle \tag{19.11}$$

$$\langle \vec{\psi} \mid \widehat{p}_D(\Lambda)\vec{\phi}\,\rangle := -i\hbar \int_0^L \psi(x)^* \frac{d\phi(x)}{dx}\, dx \tag{19.12}$$

$$\langle \widehat{p}_D^{\,\dagger}(\Lambda)\vec{\psi} \mid \vec{\phi}\,\rangle := \langle \vec{\psi}' \mid \vec{\phi}\,\rangle = \int_0^L \psi'(x)^* \phi(x)\, dx. \tag{19.13}$$

Writing $\psi'(x)$ as an integral, i.e.,[17]

$$\psi'(x) = \frac{d}{dx}\left(\int_0^x \psi'(y)\, dy + c\right), \quad c \in \mathbb{C}, \tag{19.14}$$

and using integration by parts and the Dirichlet boundary condition on $\phi(x)$ the integral in Eq. (19.13) can be written as

$$\begin{aligned} &\int_0^L \psi'(x)^* \phi(x)\, dx \\ &= \int_0^L \left\{\frac{d}{dx}\left(\int_0^x \psi'(y)\, dy + c\right)\right\}^* \phi(x)\, dx \\ &= -\int_0^L \left(\int_0^x \psi'(y)\, dy + c\right)^* \frac{d\phi(x)}{dx}\, dx. \end{aligned} \tag{19.15}$$

Compared with Eq. (19.12) we get, for all $\vec{\phi} \in \mathcal{D}\big(\widehat{p}_D(\Lambda)\big)$,

$$\int_0^L \left(i\hbar\psi(x) + \int_0^x \psi'(y)\, dy + c\right)^* \frac{d\phi(x)}{dx}\, dx = 0. \tag{19.16}$$

This result does not imply that the bracketed term must be zero. The bracketed term being a constant c' would satisfy the above

[17] Prugovečki p. 101. A function which is square-integrable is not necessarily integrable unless the integration is over a finite range.

equation,[18] i.e., we get

$$i\hbar\psi(x) + \int_0^x \psi'(y)\,dy + c = c', \tag{19.17}$$

or

$$-i\hbar\psi(x) = \int_0^x \psi'(y)\,dy + (c - c'). \tag{19.18}$$

It follows that $\psi(x)$ is absolutely continuous. Differentiating $\psi(x)$ with respective to x we get

$$-i\hbar\frac{d\psi}{dx} = \psi'(x) \quad \Rightarrow \quad \vec{\psi}(x) = \widehat{p}_D^{\,\dagger}(\Lambda)\vec{\psi} := -i\hbar\frac{d\psi}{dx}. \tag{19.19}$$

There is no requirement for $\psi(x)$ to vanish at $x = 0$ and L. The conclusion is that the adjoint $\widehat{p}_D^{\,\dagger}(\Lambda)$ is defined on a domain composed of absolutely continuous functions in $L^2(\Lambda)$ not subject to the Dirichlet boundary condition, i.e., the adjoint is defined on

$$\vec{\mathcal{D}}\big(\widehat{p}_D^{\,\dagger}(\Lambda)\big) := \{\psi \in AC(\Lambda) :\ d\psi/dx \in L^2(\Lambda)\} \tag{19.20}$$

$$\text{by} \qquad \widehat{p}_D^{\,\dagger}(\Lambda)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \tag{19.21}$$

This domain is clearly bigger than $\vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$. It follows that $\widehat{p}_D(\Lambda)$ is symmetric and is not selfadjoint. Its adjoint $\widehat{p}_D^{\,\dagger}(\Lambda)$ is the operator $\widehat{p}_{A\vec{C}}(\Lambda)$ given by Eqs. (17.29) and (17.30). The operator $\widehat{p}_{\vec{C}_0^\infty}(\Lambda)$ defined by Eq. (17.28) are similarly symmetric and it is not selfadjoint. This example highlights a general result that

> *the adjoint of a differential operator has the same differential expression acting on a possibly larger set of differentiable functions.*[19]

A similar analysis shows that the adjoint $\widehat{p}_D^{\,\dagger\dagger}(\Lambda)$ of $\widehat{p}_D^{\,\dagger}(\Lambda)$ coincides with $\widehat{p}_D(\Lambda)$. It follows that $\widehat{p}_D(\Lambda)$ is closed.[20]

---

[18] Fano p. 281. The bracketed term being a constant enables us to perform the integration resulting in the vanishing of the expression on account of the Dirichlet boundary condition. Technically we say that the set $\{d\phi(x)/dx : \phi(x) \in \vec{\mathcal{D}}(\hat{p}_D^{\,\dagger}(\Lambda))\}$ of functions is not dense in $\vec{L}^2(\Lambda)$.

[19] Weidman p. 180 Theorem 29. Wan p. 188.

[20] Fano p. 282.

19.2.2 In $\vec{L}^2(I\!R^+)$ and $\vec{L}^2(I\!R)$

The operator $\widehat{p}_{\vec{C}_c^\infty}(I\!R^+)$ defined by Eq. (17.43) is symmetric. The operator $\widehat{p}_D(I\!R^+)$ defined on a bigger domain by Eqs. (17.44) and (17.45) is also symmetric. The Dirichlet boundary condition $\psi(0) = 0$ at the origin is necessary to make this operator is symmetric. The operator is not selfadjoint since its adjoint $\widehat{p}_D^\dagger(I\!R^+)$ acts on a still bigger domain defined by[21]

$$\left\{\psi \in L^2(I\!R^+) \cap AC(I\!R^+) : d\psi/dx \in L^2(I\!R)\right\} \tag{19.22}$$

$$\text{by} \qquad \widehat{p}_D^\dagger(I\!R^+)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \tag{19.23}$$

The operator is closed since one can show that $\widehat{p}_D^{\dagger\dagger}(I\!R^+) = \widehat{p}_D(I\!R^+)$.

Arguments similar to that employed for the operator $\widehat{p}_D(\Lambda)$ show that the operator $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ in $\vec{L}^2(I\!R)$ defined by Eq. (17.47) is symmetric and not selfadjoint. Its adjoint $\widehat{p}^\dagger_{\vec{C}_c^\infty}(I\!R)$ is the momentum operator defined by Eqs. (17.49) and (17.50).

19.3 First Order Selfadjoint Differential Operators

In many cases it is possible to choose a suitable enlargement of the domain of a symmetric operator to obtain a new operator with a bigger domain which is selfadjoint operator. The resulting selfadjoint operator is called a **selfadjoint extension** of the symmetric operator. This procedure may produce many different selfadjoint extensions. There is a standard procedure to obtain selfadjoint extensions of symmetric operators.[22] Here we shall examine possible selfadjoint extensions of symmetric differential operators presented in the preceding section.[23]

[21] Wan p. 119, p. 126. Akihezer and Glazman Vol. 1 pp. 106–111. Fano pp. 279–284.

[22] Wan pp. 113–130 for more details and examples. This is directly relevant to the process of quantisation in quantum mechanics for many complex systems.

[23] Wan p. 119. Akhiezer and Glazman Vol. 1 pp. 106–111. Fano pp. 279–284. Roman Vol. 2 p. 548

19.3.1 In $\vec{L}^2(\Lambda)$

The family of operators $\widehat{p}_\lambda(\Lambda)$ in Eq. (17.34) are selfadjoint. The operator $\widehat{p}_{\lambda=0}(\Lambda)$ in Eq. (17.36) is also selfadjoint. These operators form a one-parameter family of selfadjoint extensions of $\widehat{p}_D(\Lambda)$. To see how the quasi-periodic boundary condition would lead to selfadjointness let us start by examining the following calculations for all $\vec{\varphi}, \vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_\lambda(\Lambda)\big)$:

$$
\begin{aligned}
&\langle \vec{\varphi} \mid \widehat{p}_\lambda(\Lambda)\vec{\phi}\,\rangle \\
&:= -i\hbar \int_0^L \varphi(x)^* \frac{d\phi(x)}{dx}\, dx \\
&= -i\hbar \left\{ \varphi(x)^*\phi(x)|_0^L - \int_0^L \frac{d\varphi(x)^*}{dx}\phi(x)\, dx \right\} \\
&= -i\hbar \left\{ \Big(\varphi(L)^*\phi(L) - \varphi(0)^*\phi(0)\Big) - \int_0^L \frac{d\varphi(x)^*}{dx}\phi(x) dx \right\} \\
&= -i\hbar \left(\varphi(L)^*\phi(L) - \varphi(0)^*\phi(0)\right) + \int_0^L \left(-i\hbar \frac{d\varphi(x)}{dx}\right)^* \phi(x)\, dx \\
&= -i\hbar \Big(\varphi(L)^*\phi(L) - \varphi(0)^*\phi(0)\Big) + \langle \widehat{p}_\lambda(\Lambda)\vec{\varphi} \mid \vec{\phi}\,\rangle. \qquad (19.24)
\end{aligned}
$$

Since both $\varphi(x)$ and $\phi(x)$ satisfy the quasi-periodic boundary condition in Eq. (17.25) the bracketed term vanishes, i.e.,

$$\varphi(L)^*\phi(L) - \varphi(0)^*\phi(0) = 0. \qquad (19.25)$$

The symmetry condition in Eq. (19.1) is satisfied. The operator $\widehat{p}_\lambda(\Lambda)$ is therefore symmetric. Since the quasi-periodic boundary condition is an extension of the Dirichlet boundary condition the operator $\widehat{p}_\lambda(\Lambda)$ is an extension of $\widehat{p}_D(\Lambda)$, i.e., $\widehat{p}_D(\Lambda) \subset \widehat{p}_\lambda(\Lambda)$. Following Eq. (17.97) we have

$$\widehat{p}_D^\dagger(\Lambda) \supset \widehat{p}_\lambda^\dagger(\Lambda) \quad \Rightarrow \quad \vec{\mathcal{D}}\big(\widehat{p}_D^\dagger(\Lambda)\big) \supset \vec{\mathcal{D}}\big(\widehat{p}_\lambda^\dagger(\Lambda)\big). \qquad (19.26)$$

This means that

(1) functions $\psi(x)$ corresponding to vectors $\vec{\psi}$ in $\vec{\mathcal{D}}\big(\widehat{p}_\lambda^\dagger(\Lambda)\big)$ are absolutely continuous so that Eq. (19.24) remains valid for every $\vec{\psi} \in \vec{\mathcal{D}}\big(\widehat{p}_\lambda^\dagger(\Lambda)\big)$,[24] and

[24] Eq. (19.24) with $\varphi(x)$ replaced by $\psi(x)$ remains valid since $\psi(x)$ is differentiable. The domain $\vec{\mathcal{D}}\big(\widehat{p}_D^\dagger(\Lambda)\big)$ is given by Eq. (19.20).

(2) $\widehat{p}_\lambda^\dagger(\Lambda)$ acting on $\vec{\psi}$ agrees with $\widehat{p}_D^\dagger(\Lambda)$ acting on $\vec{\psi}$, i.e., we have

$$\widehat{p}_\lambda^\dagger(\Lambda)\vec{\psi} = \widehat{p}_D^\dagger(\Lambda)\vec{\psi} := -i\hbar\frac{d\psi(x)}{dx}. \tag{19.27}$$

$$\Rightarrow \langle \widehat{p}_\lambda^\dagger(\Lambda)\vec{\psi} \mid \vec{\phi}\,\rangle = \int_0^L \Big(-i\hbar\frac{d\psi(x)}{dx}\Big)^* \phi(x)\,dx \tag{19.28}$$

$$= i\hbar\int_0^L \frac{d\psi(x)^*}{dx}\phi(x)\,dx. \tag{19.29}$$

Performing integration by parts the above integral becomes

$$i\hbar\Big(\psi(L)^*\phi(L) - \psi(0)^*\phi(0)\Big) + \int_0^L \psi(x)^*\Big(-i\hbar\frac{d\phi(x)}{dx}\Big)\,dx.$$

$$= i\hbar\Big(\psi(L)^*\phi(L) - \psi(0)^*\phi(0)\Big) + \langle\vec{\psi} \mid \widehat{p}_\lambda(\Lambda)\vec{\phi}\,\rangle. \tag{19.30}$$

By definition the adjoint $\widehat{p}_\lambda^\dagger(\Lambda)$ satisfies the condition

$$\langle \widehat{p}_\lambda^\dagger(\Lambda)\vec{\psi} \mid \vec{\phi}\,\rangle = \langle\vec{\psi} \mid \widehat{p}_\lambda(\Lambda)\vec{\phi}\,\rangle \tag{19.31}$$

By comparing the above condition with Eq. (19.29) we can conclude that every $\vec{\psi}$ in $\vec{\mathcal{D}}\big(\widehat{p}_\lambda^\dagger(\Lambda)\big)$ satisfies that quasi-periodic boundary condition so that $\vec{\mathcal{D}}\big(\widehat{p}_\lambda^\dagger(\Lambda)\big)$ and $\vec{\mathcal{D}}\big(\widehat{p}_\lambda(\Lambda)\big)$ coincide, i.e., the operator $\widehat{p}_\lambda(\Lambda)$ is selfadjoint.

In a finite-dimensional Hilbert space we know that a selfadjoint operator possesses a complete orthonormal set of eigenvectors with a corresponding set of eigenvalues. This property remains true for some selfadjoint operators in an infinite-dimensional Hilbert space as seen in the following examples:

E19.3.1(1) The operator $\widehat{p}_\lambda(\Lambda)$ possesses the following normalised eigenvectors for $n = 0, \pm1, \pm2, \ldots$, i.e.,

$$\vec{\varphi}_{\lambda,n}(\Lambda) := \frac{1}{\sqrt{L}}\exp\left[i\left(\frac{2n\pi - \lambda}{L}\right)x\right], \quad x \in \Lambda, \tag{19.32}$$

corresponding to the eigenvalues

$$p_{\lambda,n}(\Lambda) = \left(\frac{2n\pi - \lambda}{L}\right)\hbar. \tag{19.33}$$

The fact that the eigenvalues are unbounded shows that these operators are unbounded.

E19.3.1(2) The operator $\widehat{p}_{\lambda=0}(\Lambda)$ possesses the following eigenvectors

$$\vec{\varphi}_{\lambda=0,n}(\Lambda) := \frac{1}{\sqrt{L}} \exp\left[i\left(\frac{2n\pi}{L}\right)x\right], \quad n = 0, \pm 1, \pm 2, \ldots, \tag{19.34}$$

corresponding to eigenvalues

$$p_{\lambda=0,n}(\Lambda) = \frac{2n\pi}{L}\hbar. \tag{19.35}$$

From our knowledge of Fourier series we can deduce that the set of vectors $\vec{\varphi}_{\lambda,n}$, whether λ is zero or not, form a complete orthonormal set of $\vec{L}^2(\Lambda)$. These selfadjoint operators can serve as *momentum operators* for a particle confined in a box.[25]

It should be emphasised that:

(1) A different λ implies a different domain and a different operator with a different set of eigenvalues and eigenvectors.

(2) Neither $\vec{\varphi}_{\lambda,n}(\Lambda)$ nor $\vec{\varphi}_{\lambda=0,n}(\Lambda)$ are the eigenvectors of $\widehat{p}_D(\Lambda)$ since they do not satisfy the Dirichlet boundary condition and they are hence not in the domain $\vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big)$.[26] This illustrates a fundamental difference between selfadjoint and symmetric operators, i.e., a symmetric operator which is not selfadjoint does not possess a complete orthonormal set of eigenvectors. We cannot go on to define a spectral function and establish a spectral theorem for such a symmetric operator. This is why symmetric operators are not used to represent physical observables in quantum mechanics.[27]

---

[25] Schiff pp. 43–50.

[26] Do not confuse $\hat{p}_{\lambda=0}(\Lambda)$ with $\hat{p}_{\vec{C}^\infty_D}(\Lambda)$ defined by Eq. (17.28). The function $\varphi_{\lambda=0,n}(x)$ does not vanish at the boundary, and hence it is not in $C^\infty_D(\Lambda)$.

[27] It is possible to generalise the orthodox formulation of quantum mechanics to incorporate some symmetric operators for the description of physical observables. Wan pp. 395–426.

19.3.2 In $\vec{L}^2(\mathcal{C}_a)$ and $\vec{L}^2(\mathcal{S}_u)$

The operator $\widehat{p}(\mathcal{C}_a)$ in $\vec{L}^2(\mathcal{C}_a)$ defined by Eqs. (17.37) and (17.38) is selfadjoint. It possesses the following eigenvectors[28]

$$\vec{\varphi}_n(\mathcal{C}_a) := \frac{1}{\sqrt{2\pi}} e^{in\theta}, \quad n = 0, \pm 1, \pm 2, \ldots \tag{19.36}$$

corresponding to the eigenvalues

$$p_n(\mathcal{C}_a) = \frac{1}{a} n\hbar. \tag{19.37}$$

In §27.8 we shall see that $\widehat{p}(\mathcal{C}_a)$ can serve as the momentum operator for a particle confined to move in the circle $\mathcal{C}_a$.[29] These eigenvectors form a complete orthonormal set in $\vec{L}^2(\mathcal{C}_a)$.

The operators $\widehat{p}_\lambda(\mathcal{C}_a)$ defined by Eqs. (17.39) and (17.40) are also selfadjoint with eigenvectors

$$\vec{\varphi}_{\lambda,n}(\mathcal{C}_a) := \frac{1}{\sqrt{2\pi}} e^{i(n-\frac{\lambda}{2\pi})\theta}, \quad n = 0, \pm 1, \pm 2, \pm 3 \ldots \tag{19.38}$$

with corresponding eigenvalues

$$p_{\lambda,n}(\mathcal{C}_a) = \frac{1}{a}\left(n - \frac{\lambda}{2\pi}\right)\hbar. \tag{19.39}$$

In $\vec{L}^2(\mathcal{S}_u)$ an example is $\widehat{L}_z(\mathcal{S}_u)$ in Eq. (17.42). This operator acts on vectors $\vec{\phi}$ corresponding to functions $\phi(\theta, \varphi)$ of the angle variables θ and φ. With the periodic boundary condition shown in Eq. (17.41) this is operator is selfadjoint.[30]

19.3.3 In $\vec{L}^2(I\!R^+)$

The symmetric operator $\widehat{p}_D(I\!R^+)$ in $\vec{L}^2(I\!R^+)$ defined be Eq. (19.23) is closed but it admits no selfadjoint extension.[31] This would cause

[28] The normalisation factor will be $1/\sqrt{2\pi a}$ if the scalar product in $\vec{L}^2(\mathcal{C}_a)$ is defined by an integration over the linear position variable $a\theta$ instead of the angular position variable θ in Eq. (16.40).

[29] Wan pp. 481, 490. Martin pp. 46–47.

[30] *See* §36.1.2 for detailed calculation.

[31] Akhiezer and Glazman Vol. 1 pp. 110–111. Wan pp. 125–127, 172.

problems if we try to define a momentum operator for motion on the half line $I\!R^+$.[32] A similar problem arises when we try to define a *radial momentum operator*. Explicit expressions are given in Eqs. (27.146) to (27.148) in §27.10.1.[33]

19.3.4 In $\vec{L}^2(I\!R)$

In $\vec{L}^2(I\!R)$ arguments similar to that employed for the operator $\widehat{p}_\lambda(\Lambda)$ in $\vec{L}^2(\Lambda)$ show that the adjoint $\widehat{p}^{\dagger}_{\vec{C}_c^\infty}(I\!R)$ of $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ in Eq. (17.47) is equal to the operator $\widehat{p}(I\!R)$ defined by Eqs. (17.49) and (17.50), and the operator $\widehat{p}(I\!R)$ is selfadjoint.[34] We call this operator the **momentum operator** in $\vec{L}^2(-\infty, \infty)$ since this is the *momentum operator* in quantum mechanics for a particle in one-dimensional motion along the x-axis. The selfadjointness of $\widehat{p}(I\!R)$ can be proved by calculations similar to those in §19.2.1 and §19.3.1.[35]

19.3.5 In $\vec{L}^2(I\!R^2)$ and $\vec{L}^2(I\!R^3)$

The results for $\vec{L}^2(I\!R)$ can be extended to $\vec{L}^2(I\!R^2)$ and $\vec{L}^2(I\!R^3)$:

(1) In $\vec{L}^2(I\!R^2)$ we have a selfadjoint momentum operators $\widehat{p}_x(I\!R^2)$ defined Eq. (17.52). Another momentum operator $\widehat{p}_y(I\!R^2)$ is similarly defined.

(2) In $\vec{L}^2(I\!R^3)$ we have a selfadjoint momentum operators $\widehat{p}_x(I\!R^3)$ defined by Eq. (17.54). Similarly we can define two more selfadjoint momentum operators $\widehat{p}_y(I\!R^3)$ and $\widehat{p}_z(I\!R^3)$. These momentum operators satisfy the commutation relation in Eq. (17.82) with their corresponding position operators $\widehat{x}(I\!R^3)$, $\widehat{y}(I\!R^3)$ and $\widehat{z}(I\!R^3)$.

---

[32] This operator is related to the momentum operator in some superconducting circuit systems (*see* Wan and Menzies).

[33] Richtmyer Vol. 1 pp. 139–140. Wan p. 174. The similarity is due to the fact that in spherical coordinates (r, θ, φ) the radial momentum operator $\widehat{p}_r$ is a differential operator of first order in terms of $\partial/\partial r$ with radial variable r having a range $[0, \infty)$. As a result $\widehat{p}_r$ is only symmetric and not selfadjoint.

[34] Akhiezer and Glazman Vol. 1 p. 111. Fano p. 283.

[35] The equivalence of Eq. (19.25) can be achieved by choosing $\phi(x) \in C_c^\infty(I\!R)$. The operator $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ can be shown to be essentially selfadjoint (*see* Definition 19.5(1)) with $\widehat{p}(I\!R)$ as its unique selfadjoint extension (*see* Hall p. 184).

19.4 Second Order Selfadjoint Differential Operators

Second order differential operators can be conveniently generated from the first order ones, as seen by the following examples:

E19.4(1) In $\vec{L}^2(\Lambda)$ There are two obvious examples. The first example is the operator $\widehat{p}_D^\dagger(\Lambda)\widehat{p}_D(\Lambda)$ which is selfadjoint by Theorem 19.1(1). In accordance with Eq. (17.59) the operator acts on the domain

$$\left\{\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_D(\Lambda)\big),\ \ \widehat{p}_D(\Lambda)\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_D^\dagger(\Lambda)\big)\right\}. \tag{19.40}$$

The second example is the operator $\widehat{p}_\lambda^2(\Lambda)$, one for each value of λ including $\lambda = 0$. The resulting operators are selfadjoint acting on the domain

$$\vec{\mathcal{D}}\big(\widehat{p}_\lambda^2(\Lambda)\big) := \left\{\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_\lambda(\Lambda)\big),\ \ \widehat{p}_\lambda(\Lambda)\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_\lambda(\Lambda)\big)\right\}. \tag{19.41}$$

These results have direct physical applications:

(1) The Hamiltonian operator for a particle of mass m confined to the interval Λ by an infinite square potential well is given by the selfadjoint operator[36]

$$\widehat{H}_D^\infty(\Lambda) = \frac{1}{2m}\,\widehat{p}_D^\dagger(\Lambda)\widehat{p}_D(\Lambda) = -\frac{\hbar^2}{2m}\,\frac{d^2}{dx^2}, \tag{19.42}$$

acting on the domain $\vec{\mathcal{D}}\big(\widehat{H}_D^\infty(\Lambda)\big) = \mathcal{D}\big(\widehat{p}_D^\dagger(\Lambda)\widehat{p}_D(\Lambda)\big)$. While $\widehat{p}_D(\Lambda)$ does not admit any eigenvectors because of the Dirichlet boundary condition the operator $\widehat{H}_D^\infty(\Lambda)$ possesses a complete orthonormal set of eigenvectors

$$\vec{\varphi}_{D,\ell}^{\,\infty}(\Lambda) := \sqrt{\frac{2}{L}}\,\sin\left(\frac{\ell\pi}{L}\,x\right), \quad \ell = 1, 2, \ldots, \tag{19.43}$$

corresponding to eigenvalues

$$E_{D,\ell}^\infty(\Lambda) = \frac{\pi^2\hbar^2}{2mL^2}\,\ell^2, \quad \ell = 1, 2, \ldots. \tag{19.44}$$

[36] Phillips A.C. p. 66. *See* also §27.5.

These eigenvalues are nondegenerate. Note that[37]

$$\widehat{H}^{\infty}_{D}(\Lambda) \neq \frac{1}{2m}\ \widehat{p}^{\,2}_{D}(\Lambda). \tag{19.45}$$

(2) For a particle in a box we can also use the quasi-periodic boundary condition which will lead to a Hamiltonian[38]

$$\widehat{H}^{\infty}_{\lambda}(\Lambda) = \frac{1}{2m}\,\widehat{p}^{\,2}_{\lambda}(\Lambda) = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}, \tag{19.46}$$

acting on the domain $\mathcal{D}\big(\widehat{H}^{\infty}_{\lambda}(\Lambda)\big) = \vec{\mathcal{D}}\big(\widehat{p}^{\,2}_{\lambda}(\Lambda)\big)$. This Hamiltonian shares the same eigenvectors with $\widehat{p}_{\lambda}(\Lambda)$, i.e., those given by Eq. (19.32), with eigenvalues

$$E^{\infty}_{\lambda,n}(\Lambda) = \frac{p^2_{\lambda,n}(\Lambda)}{2m} = \frac{\hbar^2}{2m}\left(\frac{2n\pi - \lambda}{L}\right)^2. \tag{19.47}$$

For selfadjoint operators possessing a complete orthonormal set of eigenvetors, e.g., $\widehat{p}_{\lambda}(\Lambda)$, $\widehat{H}^{\infty}_{D}(\Lambda)$ and $\widehat{H}^{\infty}_{\lambda}(\Lambda)$, we can define eigenprojectors in the same way as before, e.g., $\widehat{H}^{\infty}_{D}(\Lambda)$ has a complete orthogonal family of eigenprojectors

$$\widehat{P}_{\vec{\varphi}^{\infty}_{D,\ell}} = |\vec{\varphi}^{\infty}_{D,\ell}(\Lambda)\rangle\langle\vec{\varphi}^{\infty}_{D,\ell}(\Lambda)|, \tag{19.48}$$

and a spectral theorem in the form of Theorem 13.3(2) can be established. We can then proceed to use these operators to set up probabilistic theories as done in §14 for spin. A general discussion will be presented in §22.2.

E19.4(2) In $\vec{L}^2(\mathcal{C}_a)$ Here we have the selfadjoint operator $\widehat{p}(\mathcal{C}_a)$ in Eq. (17.38). The physical relevance here is that a particle of mass m confined to move in a circle and is otherwise free may be described by the Hamiltonian

$$\widehat{H}(\mathcal{C}_a) = \frac{1}{2m}\,\widehat{p}^{\,2}(\mathcal{C}_a), \tag{19.49}$$

which shares the same eigenvectors with $\widehat{p}(\mathcal{C}_a)$ with eigenvalues $E_n(\mathcal{C}_a) = p^2_n(\mathcal{C}_a)/2m$, where $p_n(\mathcal{C}_a)$ are given by Eq. (19.37).

---

[37] The square $\widehat{p}^{\,2}_{D}(\Lambda)$ of $\widehat{p}_{D}(\Lambda)$ is not selfadjoint.

[38] Schiff pp. 43–50. We can have $\lambda = 0$, i.e., $\widehat{H}^{\infty}_{\lambda=0}(\Lambda)$, and $\widehat{H}^{\infty}_{\lambda=0}(\Lambda) \neq \widehat{H}^{\infty}_{D}(\Lambda)$.

E19.4(3) In $\vec{L}^2(\mathcal{S}_u)$ The operator expression

$$-\hbar^2 \left(\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\varphi^2} \right) \tag{19.50}$$

in spherical coordinates can be shown to define a selfadjoint operator $\widehat{L}^2(\mathcal{S}_u)$ in $\vec{L}^2(\mathcal{S}_u)$. This operator which is known to admit the spherical harmonics as its eigenfunctions, i.e.,

$$\widehat{L}^2(\mathcal{S}_u)\, \vec{Y}_{\ell,m} = \ell(\ell+1)\hbar^2 \vec{Y}_{\ell,m}, \quad \ell = 0, 1, 2, 3, \ldots. \tag{19.51}$$

is directly related to the operator for the *total orbital angular momentum square* in Eq. (27.88). The z-component orbital angular momentum operator $\widehat{L}_z(\mathcal{S}_u)$ in Eq. (17.42) also admits these spherical harmonics as its eigenfunctions.[39]

E19.4(4) In $\vec{L}^2(I\!R^+)$ The operator $\widehat{p}_D^{\dagger}(I\!R^+)\widehat{p}_D(I\!R^+)$ is selfadjoint by Theorem 19.1(1). It acts on the domain

$$\left\{ \vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_D(I\!R^+)\big) : \ \widehat{p}_D(I\!R^+)\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}_D^{\dagger}(I\!R^+)\big) \right\}. \tag{19.52}$$

A direct application of this result is to formulate a selfadjoint energy operator of a particle confined to move on the half-real line $I\!R^+$.[40] A similar product appears in the radial part of the kinetic energy operator of a particle in three-dimensional motion expressed in spherical coordinates as shown in Eqs. (27.149) and (27.150).[41]

E19.4(5) In $\vec{L}^2(I\!R)$ The operator $\widehat{p}^{\,2}(I\!R)$ is selfadjoint acting on

$$\vec{\mathcal{D}}\big(\widehat{p}^{\,2}(I\!R)\big) = \left\{ \vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}(I\!R)\big) : \ \widehat{p}(I\!R)\vec{\phi} \in \vec{\mathcal{D}}\big(\widehat{p}(I\!R)\big) \right\}. \tag{19.53}$$

This is directly related to the kinetic energy operator of a particle in motion along the x-axis shown in Eq. (27.91).

E19.4(6) Operators can have differential and multiplicative parts. An example is

$$\widehat{H}_{ho} = \frac{1}{2m}\widehat{p}^{\,2}(I\!R) + \frac{1}{2}m\omega^2\, \widehat{x}^{\,2}(I\!R), \quad m, \omega \in I\!R \tag{19.54}$$

---

[39] *See* §36.1.1 and §36.1.2 and Eq.(17.128). Do not confuse $\vec{L}^2(\mathcal{S}_u)$ with $\widehat{L}^2(\mathcal{S}_u)$.

[40] Wan p. 285.

[41] Wan pp. 279–281. The radial part of the kinetic energy contains the term $\hat{p}_r^{\dagger}\hat{p}_r$ which is selfadjoint.

acting on the domain[42]

$$\mathcal{D}(\widehat{H}_{ho}) = \vec{\mathcal{D}}(\widehat{p}^{\,2}(I\!R)) \cap \vec{\mathcal{D}}(\widehat{x}^{\,2}(I\!R)). \tag{19.55}$$

This operator is selfadjoint and it is related to the Hamiltonian operator of a **quantised harmonic oscillator** corresponding to the classical Hamiltonian given in Eq. (27.11). In Chapter 35 we shall examine this operator again in details.

19.5 Essentially Selfadjoint Operators

Definition 19.5(1) *An operator* $\widehat{A}$ *defined on the domain* $\vec{\mathcal{D}}(\widehat{A})$ *in a Hilbert space* $\vec{\mathcal{H}}$ *is said to be* **essentially selfadjoint** *if it is symmetric and it possesses a unique selfadjoint extension.*[43]

It is often difficult and cumbersome to specify the domain of a selfadjoint operator explicitly. In practical applications a selfadjoint operator $\widehat{A}$ is applied to act only on a dense subset of vectors in its domain chosen for physical reasons. Let us denote a dense subset of $\vec{\mathcal{D}}(\widehat{A})$ by $\vec{\mathcal{D}}_d(\widehat{A})$. We have in effect introduce a new operator $\widehat{A}_d$ defined on the domain $\vec{\mathcal{D}}_d(\widehat{A})$ by

$$\widehat{A}_d\vec{\phi} = \widehat{A}\vec{\phi} \quad \forall\, \vec{\phi} \in \vec{\mathcal{D}}_d(\widehat{A}) \subset \vec{\mathcal{D}}(\widehat{A}). \tag{19.56}$$

Moreover $\vec{\mathcal{D}}_d(\widehat{A})$ is often chosen so that $\widehat{A}_d$ is symmetric on the domain $\vec{\mathcal{D}}_d(\widehat{A})$. Clearly $\widehat{A}_d$ is a restriction of the original operator $\widehat{A}$. Conversely $\widehat{A}$ is a selfadjoint extension of $\widehat{A}_d$. Let us examine the following examples.

E19.5(1) In $\vec{L}^2(I\!R)$ the operators $\widehat{p}_{\vec{C}_D^\infty}(I\!R)$ in Eq. (17.47) and $\widehat{p}_{\vec{S}_s}(I\!R)$ in Eq. (17.48) are restrictions of $\widehat{p}(I\!R)$ defined by Eq. (17.50). They are both symmetric and they have the same unique selfadjoint extension $\widehat{p}(I\!R)$. It follows that both $\widehat{p}_{\vec{C}_c^\infty}(I\!R)$ and $\widehat{p}_{\vec{S}_s}(I\!R)$ are essentially selfadjoint.[44]

---

[42] *See* Eq. (17.16) for $\vec{\mathcal{D}}(\widehat{x}^{\,2}(I\!R))$.

[43] Weidmann p. 234.

[44] Hall p. 184 where $\widehat{p}_{\vec{C}_D^\infty}(I\!R)$ is proved to be essentially selfadjoint. It follows that $\widehat{p}_{\vec{S}_s}(I\!R)$ is also essentially selfadjoint.

E19.5(2) In $\vec{L}^2(\Lambda)$ the operator $\widehat{p}_{\vec{C}^\infty_D}(\Lambda)$ in Eq. (17.28) is a restriction of $\widehat{p}_\lambda(\Lambda)$ in Eq. (17.34). It is symmetric and it has a one-perameter family of selfadjoint extensions $\widehat{p}_\lambda(\Lambda)$. It follows that $\widehat{p}_{\vec{C}^\infty_D}(\Lambda)$ in $\vec{L}^2(\Lambda)$ is not essentially selfadjoint.

A selfadjoint operator can be defined by one of its essential selfadjoint restrictions. In other words, we can determine a selfadjoint operator in terms of a well-chosen essential selfadjoint restriction. In many applications we would make use of such an essential selfadjoint restriction since the domain of an essentially selfadjoint operator is easier to specified. Examples can be given in Chapter 27 on quantisation and Chapter 35 on harmonic oscillator.[45]

Theorem 19.5(1)[46] *A symmetric operator $\widehat{A}$ is essentially selfadjoint if its closure $\bar{A}$ is selfadjoint.*

It follows that the closure which is equal to $\widehat{A}^{\dagger\dagger}$ by Eq. (19.7) is the unique selfadjoint extension.

Another application of the concept of essential selfadjointness is to the sum and product of two operators. It can be shown that if two selfadjoint operators commute then their sum and product are essentially selfadjoint.[47]

We need to know how to determine whether a symmetric operator is essentially selfadjoint. An idea based on the concept of *analytic vectors* can lead to a theorem on the criterion for essential selfadjointness. The idea originates in the exponential function of a selfadjoint operator $\widehat{A}$ in Eq. (13.89). When we consider an expansion of an exponential function as shown in Eq. (13.89) we need to consider the domain of $\widehat{A}^n$ for all n. Let the domain of $\widehat{A}^n$ be denoted by $\mathcal{D}(\widehat{A}^n)$. A vector $\vec{\phi}$ must be the intersection of all $\mathcal{D}(\widehat{A}^n)$, i.e., $\vec{\phi} \in \cap_{n=1}^\infty \mathcal{D}(\widehat{A}^n)$, for $\widehat{A}^n\vec{\phi}$ to be defined for all n.

[45] Wan pp. 252–28.
[46] Reed and Simon Vol. 1 p. 256. Weidmann p. 108.
[47] Weidmann p. 268. Their sum and product are selfadjoint if the operators are bounded.

Definition 19.5(2)[48] *A vector* $\vec{\phi} \in \cap_{n=1}^{\infty} \mathcal{D}(\widehat{A}^n)$ *is called an analytic vector of* $\widehat{A}$ *if for some positive real number* $t > 0$ *we have*

$$\sum_{n=0}^{\infty} \frac{||\widehat{A}^n \vec{\phi}\,||}{n!}\, t^n \;<\; \infty, \tag{19.57}$$

where $\widehat{A}^0$ *is taken as the identity operator, i.e.,* $\widehat{A}^0 = \widehat{\mathbb{I}}$.

When acting on an analytic vector $\vec{\phi}$ we have the desired expansion[49]

$$e^{it\hat{A}}\,\vec{\phi} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!}\,\widehat{A}^n \vec{\phi}. \tag{19.58}$$

Theorem 19.5(2)[50] *Let* $\widehat{A}$ *be a symmetric operator in a Hilbert space* $\mathcal{H}$ *and let* $\mathcal{D}(\widehat{A})$ *be its domain. Then* $\widehat{A}$ *is essentially selfadjoint if the analytic vectors of* $\widehat{A}$ *in* $\mathcal{D}(\widehat{A})$ *form a complete set of vectors for* $\mathcal{H}$.

As an example consider a symmetric operator $\widehat{A}$ with a known set of eigenvectors. We first note that a normalised eigenvector $\vec{\varphi}$ of $\widehat{A}$ corresponding to the eigenvalue a is automatically an analytic vector of $\widehat{A}$ since the left-hand side of Eq. (19.57) becomes[51]

$$\sum_{n=0}^{\infty} \frac{||a^n \vec{\varphi}\,||}{n!}\, t^n \;=\; \sum_{n=0}^{\infty} \frac{|a\,|^n}{n!}\, t^n \;= e^{|a|\,t} \tag{19.59}$$

which is finite. Theorem 19.5(2) then tells us that $\widehat{A}$ is essentially selfadjoint if the set of eigenvectors of $\widehat{A}$ is complete.[52] Examples are available in §35.2.2.

Exercises and Problems

Q19(1) Show that symmetric operators generate real quadratic forms, as shown in Eq. (19.2).

---

[48] Reed and Simon Vol. I p. 276. Reed and Simon Vol. II p. 201. Moretti p. 230.

[49] Reed and Simon Vol. I p. 276. Blank, Exner and Havlíček p. 183.

[50] Reed and Simon Vol. II p. 202. Moretti p. 231. This theorem is known as Nelson's analytic vector theorem.

[51] Blank, Exner and Havlíček pp. 183, 198.

[52] *See* Read and Simon Vol. II pp. 204–205 for application to operators with a continuous spectrum.

Q19(2) By evaluating $\langle \vec{\phi} \mid \widehat{N}\vec{\psi} \rangle$ and $\langle \widehat{N}\vec{\phi} \mid \vec{\psi} \rangle$ separately verify that the number operator $\widehat{N}$ in Eq. (19.5) satisfies

$$\langle \vec{\phi} \mid \widehat{N}\vec{\psi} \rangle = \langle \widehat{N}\vec{\phi} \mid \vec{\psi} \rangle \quad \forall \vec{\phi},\ \vec{\psi} \in \vec{\mathcal{D}}(\widehat{N}). \tag{19.60}$$

Q19(3) Show that the number operator $\widehat{N}$ in Eq. (19.5) is bounded below. Show also that the square of a selfadjoint operator is also bounded below.

Q19(4) Explain why $\widehat{H}_D^{\infty}(\Lambda)$ cannot act on $\vec{\varphi}_{\lambda=0,n}(\Lambda)$ in Eq. (19.34) and that $\langle \vec{\varphi}_{\lambda=0,n}(\Lambda) \mid \widehat{H}_D^{\infty}(\Lambda)\vec{\varphi}_{\lambda=0,n}(\Lambda)\rangle$ is undefined.[53]

Q19(5) The eigenvectors $\vec{\varphi}_{D,\ell}^{\infty}(\Lambda)$ of $\widehat{H}_D^{\infty}(\Lambda)$ in Eq. (19.43) form an orthonormal basis for $\vec{L}^2(\Lambda)$. We can expression $\vec{\varphi}_{\lambda=0,n}(\Lambda)$ in Eq. (19.34) as a linear combination of $\vec{\varphi}_{D,\ell}^{\infty}(\Lambda)$. Consider the vector $\vec{\varphi}_{\lambda=0,n=0}(\Lambda)$ which corresponds to a constant function. We have

$$\vec{\varphi}_{\lambda=0,n=0}(\Lambda) = \sum_{\ell=1}^{\infty} c_\ell\, \vec{\varphi}_{D,\ell}^{\infty}(\Lambda). \tag{19.61}$$

Evaluate the coefficients c_ℓ.

Investigate whether any one of the following two procedures would yield a meaningful value for the quadratic form $\mathcal{Q}\big(\widehat{H}_D^{\infty}(\Lambda),\, \vec{\varphi}_{\lambda=0,n=0}(\Lambda)\big)$ given formally by

$$\langle \vec{\varphi}_{\lambda=0,n=0}(\Lambda) \mid \widehat{H}_D^{\infty}(\Lambda)\vec{\varphi}_{\lambda=0,n=0}(\Lambda)\rangle. \tag{19.62}$$

(1) Assume that $\widehat{H}_D^{\infty}(\Lambda)\vec{\varphi}_{\lambda=0,n=0}(\Lambda)$ can be calculated by

$$\widehat{H}_D^{\infty}(\Lambda)\vec{\varphi}_{\lambda=0,n=0}(\Lambda) = \widehat{H}_D^{\infty}(\Lambda)\left(\sum_{\ell=1}^{\infty} c_\ell \vec{\varphi}_{D,\ell}^{\infty}(\Lambda)\right) \tag{19.63}$$

$$= \sum_{\ell=1}^{\infty} c_\ell \widehat{H}_D^{\infty}(\Lambda)\vec{\varphi}_{D,\ell}^{\infty}(\Lambda) \tag{19.64}$$

$$= \sum_{\ell=1}^{\infty} c_\ell E_{D,\ell}^{\infty}(\Lambda)\vec{\varphi}_{D,\ell}^{\infty}(\Lambda), \tag{19.65}$$

where $E_{D,\ell}^{\infty}(\Lambda)$ are the corresponding eigenvalues of $\widehat{H}_D^{\infty}(\Lambda)$ in Eq. (19.44). Then assume that the

[53] $\vec{\varphi}_{\lambda=0,n}(\Lambda)$ are the eigenvectors of operator $\widehat{p}_{\lambda=0}(\Lambda)$ in E19.3(2).

expression in Eq. (19.62) can be calculated by

$$\langle \vec{\varphi}_{\lambda=0,n=0}(\Lambda) \mid \sum_{\ell=1}^{\infty} c_\ell E^{\infty}_{D,\ell}(\Lambda)\vec{\varphi}^{\infty}_{D,\ell}(\Lambda)\rangle \tag{19.66}$$

$$= \sum_{\ell=1}^{\infty} c_\ell E^{\infty}_{D,\ell}(\Lambda)\langle \vec{\varphi}_{\lambda=0,n=0}(\Lambda) \mid \vec{\varphi}^{\infty}_{D,\ell}(\Lambda)\rangle \tag{19.67}$$

$$= \sum_{\ell=1}^{\infty} |c_\ell|^2 E^{\infty}_{D,\ell}(\Lambda). \tag{19.68}$$

Determine whether the above sum converges.

(2) Assume that $\widehat{H}^{\infty}_{D}(\Lambda)\vec{\varphi}_{\lambda=0,n=0}(\Lambda)$ can be calculated by formal differentiation using Eq. (19.42). What vector would be obtained? Is the resulting value of the expression

$$\langle \vec{\varphi}_{\lambda=0,n=0}(\Lambda) \mid \widehat{H}^{\infty}_{D}(\Lambda)\vec{\varphi}_{\lambda=0,n=0}(\Lambda)\rangle$$

meaningful?

What conclusion can be drawn from the investigation?

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 20

Spectral Theory of Selfadjoint Operators in $\vec{\mathcal{H}}$

We shall follow Chapter 15 to develop a spectral theory for selfadjoint operators in a Hilbert Space, starting with spectral functions and spectral measures.

20.1 Spectral Functions and Spectral Measures

Definition 20.1(1) *Spectral functions* $\widehat{F}(\tau)$ *and spectral measures* $\widehat{M}(\Lambda)$ *on a Hilbert space* $\vec{\mathcal{H}}$ *are defined by* Definitions 15.1(1) *and* 15.1(2)*, with the understanding that* $\widehat{F}(\tau)$ *and* $\widehat{M}(\Lambda)$ *are projectors on* $\vec{\mathcal{H}}$ *rather than on* $\vec{\mathbb{W}}^N$.

Spectral functions and spectral measures on $\vec{\mathcal{H}}$ are also related in a one-to-one manner by Theorem 15.1(1). However, spectral functions on an infinite-dimensional Hilbert space may not be piecewise-constant.

As an example consider the space $\vec{L}^2(I\!R)$ where we have a multiplication operator $\widehat{\chi}_\Lambda$ defined by Eq. (17.11) for each interval Λ. Letting $\Lambda = (-\infty, \tau]$ we obtain a one-parameter family of multiplication operators $\widehat{\chi}_{(-\infty,\tau]}$, for each value of τ. This family satisfies Definitions 15.1(1) and 15.1(2):

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

(1) Definition 15.1(1) Clearly $\widehat{\chi}_{(-\infty,\tau]}$ is a projector each τ. For each input vector $\vec{\phi}$ the output vector $\widehat{\chi}_{(-\infty,\tau]}\vec{\phi}$ is defined by the truncated function $\chi_{(-\infty,\tau]}(x)\phi(x)$, i.e.,

$$\widehat{\chi}_{(-\infty,\tau]}\vec{\phi} := \chi_{(-\infty,\tau]}(x)\phi(x) = \begin{cases} \phi(x), & x \le \tau, \\ 0, & x > \tau. \end{cases} \tag{20.1}$$

Functions $\phi(x)$ in $L^2(I\!R)$ truncated for all $x > \tau$ define a subspace $\vec{\mathcal{S}}_\tau$ of $\vec{L}^2(I\!R)$. The operator $\widehat{\chi}_{(-\infty,\tau]}$ projects every vector $\vec{\phi}$ onto $\vec{\mathcal{S}}_\tau$.

(2) Definition 15.1(2) The following discussion shows that properties SF15.1(1), SF15.1(1) and SF15.1(1) are satisfied:

(a) We have $\tau_1 \le \tau_2 \Rightarrow \widehat{\chi}_{(-\infty,\tau_1]} \le \widehat{\chi}_{(-\infty,\tau_2]}$, since[1]

$$||\widehat{\chi}_{(-\infty,\tau_1]}\vec{\phi}\,|| \le ||\widehat{\chi}_{(-\infty,\tau_2]}\vec{\phi}\,||. \tag{20.2}$$

(b) Clearly $\widehat{\chi}_{(-\infty,-\infty)}\vec{\phi} = \vec{0}$ and $\widehat{\chi}_{(-\infty,\infty)}\vec{\phi} = \vec{\phi}$, since $\chi_{(-\infty,-\infty)}(x)$ would obliterate any $\phi(x)$ and $\chi_{(-\infty,\infty)}(x)$ would leave $\phi(x)$ unchanged.

(c) We also have $\widehat{\chi}_{(-\infty,\tau+\delta]} \to \widehat{\chi}_{(-\infty,\tau]}$ as $\delta \to +0$ since

$$||\big(\widehat{\chi}_{(-\infty,\tau+\delta]} - \widehat{\chi}_{(-\infty,\tau]}\big)\vec{\phi}\,|| \to 0. \tag{20.3}$$

It follows that the family $\{\,\widehat{\chi}_{(-\infty,\tau]},\ \tau \in I\!R\,\}$ of projectors $\widehat{\chi}_{(-\infty,\tau]}$ defines a spectral function in $\vec{L}^2(I\!R)$.

The corresponding spectral measure $\widehat{M}$ is given Eqs. (15.19) and (15.20), e.g., we have

$$\widehat{M}\big((\tau_1,\,\tau_2]\big) = \widehat{\chi}_{(-\infty,\tau_2]} - \widehat{\chi}_{(-\infty,\tau_1]} = \widehat{\chi}_{(\tau_1,\tau_2]}. \tag{20.4}$$

or more directly on any interval Λ by

$$\widehat{M}\big(\Lambda\big) = \widehat{\chi}_{\Lambda}. \tag{20.5}$$

This spectral function is not piecewise-constant. Every point is a point of change, i.e., we have $\widehat{\chi}_{(-\infty,\tau+\delta]} - \widehat{\chi}_{(-\infty,\tau-\delta]} \ne \widehat{0}$ for any small

[1] The order relation is defined by Eq. (13.20). All the calculations are done in terms of the corresponding wave functions $\phi(x)$.

positive number δ and every τ in $I\!R$. In other words, the spectral function satisfies Eq. (15.47) so that every point τ is a point of continuous growth of $\widehat{F}(\tau)$.

The expression for the decomposition of the identity in Eq. (15.39) applies here, i.e., we have

$$\int_{-\infty}^{\infty} d_\tau \, \widehat{\chi}_{(-\infty,\tau]} = \widehat{I\!I}. \tag{20.6}$$

To verify this we have, for every $\vec{\phi} \in \vec{L}^2(I\!R)$,[2]

$$\begin{aligned}
&\langle \vec{\phi} \mid \left(\int_{-\infty}^{\infty} d_\tau \, \widehat{\chi}_{(-\infty,\tau]} \right) \vec{\phi} \, \rangle \\
&= \langle \vec{\phi} \mid \int_{-\infty}^{\infty} d_\tau \left(\widehat{\chi}_{(-\infty,\tau]} \vec{\phi} \right) \rangle = \int_{-\infty}^{\infty} d_\tau \left(\langle \vec{\phi} \mid \widehat{\chi}_{(-\infty,\tau]} \vec{\phi} \, \rangle \right) \\
&:= \int_{-\infty}^{\infty} d_\tau \left(\int_{-\infty}^{\infty} \phi^*(x) \chi_{(-\infty,\tau]}(x) \phi(x) \, dx \right) \\
&= \int_{-\infty}^{\infty} d_\tau \int_{-\infty}^{\tau} \phi^*(x) \phi(x) \, dx \\
&= \int_{-\infty}^{\infty} \phi^*(\tau) \phi(\tau) \, d\tau = \langle \vec{\phi} \mid \vec{\phi} \, \rangle \quad \Rightarrow \quad \text{Eq. (20.6).}
\end{aligned} \tag{20.9}$$

20.2 Spectral Theorem and Spectrum

Theorem 20.2(1) The Spectral Theorem[3]

(1) *To every selfadjoint operator $\widehat{A}$ in a Hilbert space $\vec{\mathcal{H}}$ there corresponds a unique spectral function, known as the* **spectral**

[2] For the corresponding $\phi(x) \in L^2(I\!R)$ we have

$$d_\tau \int_{-\infty}^{\tau} \phi^*(x) \phi(x) \, dx = \phi^*(\tau) \phi(\tau) \, d\tau. \tag{20.7}$$

This is based on the following standard results in calculus:

$$\frac{d}{dx} \int_{-\infty}^{x} f(y) dy = f(x) \quad \text{and} \quad d_x \int_{-\infty}^{x} f(y) dy = f(x) dx. \tag{20.8}$$

[3] This is a generalisation of Theorem 13.3.2(1). We shall omit the rather complicated proof. Here $\widehat{A} \neq \widehat{I\!I}$.

function *of* $\widehat{A}$ *and denoted by* $\widehat{F}^{\hat{A}}(\tau)$*, such that the following properties are satisfied*[4]:

(a) *For every* $\vec{\varphi}$ *in the domain* $\vec{\mathcal{D}}(\widehat{A})$ *of* $\widehat{A}$ *we have*

$$||\widehat{A}\vec{\varphi}\,||^2 = \int_{-\infty}^{\infty} \tau^2\, d_\tau \langle \vec{\varphi} \mid \widehat{F}^{\hat{A}}(\tau)\vec{\varphi}\,\rangle. \tag{20.10}$$

(b) *For every* $\vec{\psi}$ *in* $\vec{\mathcal{H}}$ *and for every* $\vec{\varphi}$ *in* $\vec{\mathcal{D}}(\widehat{A})$ *we have*[5]

$$\langle \vec{\psi} \mid \widehat{A}\vec{\varphi}\,\rangle = \int_{-\infty}^{\infty} \tau\, d_\tau \langle \vec{\psi} \mid \widehat{F}^{\hat{A}}(\tau)\vec{\varphi}\,\rangle. \tag{20.11}$$

(2) *Every spectral function* $\widehat{F}(\tau)$ *defines a selfadjoint operator* $\widehat{A}$ *acting on a domain* $\vec{\mathcal{D}}(\widehat{A})$ *consisting of vectors* $\vec{\varphi}$ *for which*

$$\int_{-\infty}^{\infty} \tau^2\, d_\tau \langle \vec{\varphi} \mid \widehat{F}(\tau)\vec{\varphi}\,\rangle < \infty, \tag{20.12}$$

and on $\mathcal{D}(\widehat{A})$ the operator possesses properties (a) and (b) above.

This theorem establishes a one-to-one correspondence between spectral functions and selfadjoint operators. It is convenient to adopt the expression for the spectral decomposition operator $\widehat{A}$ used in Theorem 15.3(1), i.e., we express the operator as

$$\widehat{A} = \int_{-\infty}^{\infty} \tau\, d_\tau \widehat{F}^{\hat{A}}(\tau), \tag{20.13}$$

with the understanding that this expression is just symbolic statement for Eqs. (20.10), (20.11) and (20.12).

Definition 15.3(1) applies here, i.e., the spectral function $\widehat{F}^{\hat{A}}(\tau)$ defines a **spectral measure** $\widehat{M}^{\hat{A}}$ associated with $\widehat{A}$. We call the projector $\widehat{M}^{\hat{A}}(\Lambda)$ associated with any Borel set Λ a **spectral projector** of $\widehat{A}$ and the subspace onto which $\widehat{M}^{\hat{A}}(\Lambda)$ projects a **spectral subspace**. Examples of spectral functions of some familiar operators will be given in §20.3 and §20.4.

In relation to the theorem we also have the following results:

(1) A selfadjoint operator commutes with its spectral projectors.[6]

[4] Prugovečki p. 250. Gallone pp. 475–476.
[5] Romain Vol. 2. pp. 636, 641.
[6] Roman Vol. 2 pp. 642–643.

(2) The decomposition of the identity in Eq. (15.39) remains valid.

(3) Theorem 15.3(2) on the relation between the spectral measures and spectral functions of two unitarily related operators remains valid here. We shall present an example of this relation in §20.4.2.

(4) Theorem 15.5(1) on the generation of probability distribution function by a selfadjoint operator $\widehat{A}$ together with a unit vector $\vec{\varphi}$ applies here, i.e.,

$$\mathcal{F}^{\hat{A}}(\vec{\varphi}, \tau) = \langle \vec{\varphi} \mid \widehat{F}^{\hat{A}}(\tau)\vec{\varphi} \rangle \tag{20.14}$$

is a probability distribution. More detailed discussion together with illustrative examples will be presented in Chapter 21.

Definition 15.3(2) on the spectrum of a selfadjoint operator applies here, i.e., the spectrum $sp(\widehat{A})$ of a selfadjoint operator $\widehat{A}$ consists of all the values of τ at which its spectral function undergoes a change. Generally the spectrum of $\widehat{A}$ can be divided into two parts: the *discrete part* which consists of all points of discontinuous growth and the *continuous part* which consists of the rest. A spectrum can be purely discrete, purely continuous or a combination of the two:

(1) **Purely discrete spectrum** A spectrum is said to be a *purely discrete* and is denoted by $sp_d(\widehat{A})$ if it consists of only of points of discontinuous growth. The corresponding spectral function $\widehat{F}^{\hat{A}}(\tau)$ is also said to be *purely discrete*.

(2) **Purely continuous spectrum** A spectrum is said to be *purely continuous* and is denoted by $sp_c(\widehat{A})$ if the spectrum has no discrete part. The corresponding spectral function $\widehat{F}^{\hat{A}}(\tau)$ is also said to be *purely continuous*.

(3) **Mixed spectrum** A spectrum could also be partly continuous and partly discrete. Such a spectrum is known as a *mixed spectrum* and is denoted by $sp_m(\widehat{A})$, with the discrete and continuous parts denoted, respectively by $sp_d(\widehat{A})$ and $sp_c(\widehat{A})$. We have $sp_m(\widehat{A}) = sp_d(\widehat{A}) \cup sp_c(\widehat{A})$.[7]

[7] Roman Vol. 2 p. 652 for a diagrammatic illustration.

For brevity we shall often omit the adjective *pure* and *purely* and call a spectrum and spectral function discrete or continuous.

The nature of the spectrum of $\widehat{A}$ can be appreciated in terms of the probability distribution functions it generates[8]:

(1) If the spectrum of $\widehat{A}$ is (purely) discrete then its associated probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\varphi}, \tau)$ is piecewise-constant with countable discontinuities.
(2) If the spectrum of $\widehat{A}$ is (purely) continuous then its associated probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\varphi}, \tau)$ is continuous.

Similar to continuous probability distribution functions given in Definition 4.3.2(1) a continuous spectrum can be further divided into an **absolutely continuous part**, i.e., when $\mathcal{F}^{\hat{A}}(\vec{\varphi}, \tau)$ is absolutely continuous in τ, and an **singularly continuous part**, i.e., when $\mathcal{F}^{\hat{A}}(\vec{\varphi}, \tau)$ is singularly continuous in τ.[9] Fortunately we rarely encounter operators with a spectrum containing a singular continuous part in practical applications. Since all the continuous spectra we shall encounter will be absolutely continuous in the variable τ we shall assume that a continuous spectrum meant an absolutely continuous spectrum from now on. and This enables us to associate a probability density function to each continuous spectrum by

$$w^{\hat{A}}(\vec{\varphi}, \tau) = \mathcal{F}^{\hat{A}}(\vec{\varphi}, \tau)/d\tau. \tag{20.15}$$

In the next section we shall consider operators with a purely discrete spectrum. This is followed by a section on operators with a purely continuous spectrum.

20.3 Operators with a Discrete Spectrum

In a finite-dimensional space $\vec{\mathbb{V}}^N$ the spectrum of a selfadjoint operator is always discrete and it coincides with the set of eigenvalues. This remains true in a Hilbert space $\vec{\mathcal{H}}$ for a selfadjoint

[8] Weidmann p. 206 Theorem 7.27.
[9] Weidmann pp. 206–209. Amrein, Jauch and Sinha pp. 204–209.

operator $\widehat{A}$ with a pure discrete spectrum.[10] We have the following properties:

P20.3(1) The spectrum $sp_d(\widehat{A})$ coincides with the set of distinct eigenvalues a_m of $\widehat{A}$. The discussion in §9.4.4 and Definition 9.4.4(1) remain valid. In accordance with Eq. (15.42) the eigenprojector $\widehat{P}^{\hat{A}}(a_m)$ associated with the eigenvalue a_m is given by the spectral projector $\widehat{M}^{\hat{A}}\big(\{a_m\}\big)$. There are two cases:

(1) If an eigenvalue a_ℓ is nondegenerate then it would correspond to only one normalised eigenvector $\vec{\varphi}_\ell$ with a corresponding projectors $\widehat{P}_{\vec{\varphi}_\ell} = |\vec{\varphi}_\ell\rangle\langle\vec{\varphi}_\ell|$. We have

$$\widehat{P}^{\hat{A}}(a_\ell) = \widehat{M}^{\hat{A}}\big(\{a_\ell\}\big) = \widehat{P}_{\vec{\varphi}_\ell}. \tag{20.16}$$

(2) If an eigenvalue a_m is degenerate then the spectral projector $\widehat{M}^{\hat{A}}\big(\{a_m\}\big)$ is not one-dimensional. Let d be the degeneracy of a_m. Then associated with this eigenvalue there are d orthonormal eigenvectors $\vec{\varphi}_{mj}$, $j = 1, 2, \ldots, d$, together with their corresponding projectors $\widehat{P}_{\vec{\varphi}_{mj}} = |\vec{\varphi}_{mj}\rangle\langle\vec{\varphi}_{mj}|$. The corresponding spectral projector is given by

$$\widehat{M}^{\hat{A}}\big(\{a_m\}\big) = \sum_{j=1}^{d} \widehat{P}_{\vec{\varphi}_{mj}}. \tag{20.17}$$

The eigenprojector $\widehat{P}^{\hat{A}}(a_m)$ is the same as the spectral projector $\widehat{M}^{\hat{A}}\big(\{a_m\}\big)$, i.e.,

$$\widehat{P}^{\hat{A}}(a_m) = \widehat{M}^{\hat{A}}\big(\{a_m\}\big) = \sum_{j=1}^{d} \widehat{P}_{\vec{\varphi}_{mj}}. \tag{20.18}$$

(3) These operators are reducible by their eigensubspaces.

P20.3(2) The spectral function $\widehat{F}^{\hat{A}}(\tau)$ is piecewise-constant. If the spectrum is of the form $\{a_1, a_2, \ldots\}$, i.e., it has a lower bound by a_1, then $\widehat{F}^{\hat{A}}(\tau)$ is of the form shown in Figure 15.1, with N possibly

[10] Roman Vol. 2 pp. 651–652. Naimark Part 2 pp. 17–18.

going to infinity. Equation (15.2) becomes

$$\widehat{F}^{\hat{A}}(\tau) = \begin{cases} \widehat{0}, & \tau < a_1 \\ \widehat{P}^{\hat{A}}(a_1), & a_1 \le \tau < a_2 \\ \widehat{P}^{\hat{A}}(a_1) + \widehat{P}^{\hat{A}}(a_2), & a_2 \le \tau < a_3 \\ \cdots & \cdots \\ \sum_{\ell=1}^{n} \widehat{P}^{\hat{A}}(a_\ell), & a_n \le \tau < a_{n+1} \\ \cdots & \cdots \end{cases} . \tag{20.19}$$

P20.3(3) The integral expressions for the spectral decomposition of $\widehat{A}$ in Eq. (20.13) and for spectral decomposition of the identity $\widehat{I\!I}$ in Eq. (15.39) reduce to the form of a sum[11]:

$$\widehat{A} = \sum_m a_m \widehat{P}^{\hat{A}}(a_m), \quad \widehat{I\!I} = \sum_m \widehat{P}^{\hat{A}}(a_m). \tag{20.20}$$

As for finite-dimensional spaces the operator $\widehat{A}$ commutes with all its eigenprojectors. When the eigenvalues a_ℓ diverge to infinity as ℓ tends to infinity the infinite sum for $\widehat{A}$ in Eq. (20.20) should be handled with care. We need to specify the domain of operation in order to render the infinite sum meaningful as an operator. This would be the case if $\widehat{A}$ is unbounded.

P20.3(4) When all the eigenvalues are nondegenerate we get, from the spectral decomposition of the identity,

$$\vec{\phi} = \big(\sum_\ell \widehat{P}_{\vec{\varphi}_\ell}\big)\vec{\phi} = \sum_\ell \widehat{P}_{\vec{\varphi}_\ell}\vec{\phi} = \sum_\ell c_\ell \vec{\varphi}_\ell, \quad c_\ell = \langle \vec{\varphi}_\ell \mid \vec{\phi} \rangle, \tag{20.21}$$

for every $\vec{\phi} \in \vec{\mathcal{H}}$. This demonstrates the completeness of the orthonormal set of eigenvectors of a selfadjoint operator with a pure discrete spectrum. A similar decomposition is also obtained when the eigenvalues are degenerate.

Many selfadjoint operators presented earlier possess a purely discrete spectrum:

(1) In $\vec{L}^2(\Lambda)$ the selfadjoint operator $\widehat{p}_\lambda(\Lambda)$ in Eq. (17.38) admits a purely discrete set of nondegenerate eigenvalues $p_{\lambda,\ell}(\Lambda)$ given by Eq. (19.33).

[11] Using the same notation as in Theorem 13.3.2(2).

(2) In $\vec{L}^2(\mathcal{C}_a)$ the selfadjoint operator $\widehat{p}(\mathcal{C}_a)$ admits a purely discrete set of nondegenerate eigenvalues $p_n(\mathcal{C}_a)$ given by Eq. (19.37). The operator $\widehat{H}(\mathcal{C}_a)$ in Eq. (19.49) has both nondegenerate and degenerate eigenvalues, i.e., the lowest eigenvalue $E_0(\mathcal{C}_a) = 0$ is nondegenerate and all the other eigenvalues $E_n(\mathcal{C}_a)$ are degenerate with degeneracy 2.

(3) In $\vec{L}^2(I\!R)$ the Hamiltonian operator $\widehat{H}_{ho}$ in Eq. (19.54) is well-known to possess a purely discrete set of nondegenerate eigenvalues[12]

$$E_n = \big(n + 1/2\big)\hbar\omega, \quad n = 0, 1, 2, \ldots. \tag{20.22}$$

P20.3(5) Theorem 9.4.4(1) applies, i.e., for a selfadjoint operator in an infinite-dimensional Hilbert space $\vec{\mathcal{H}}$ with a purely discrete spectrum its eigenvalues are real, its eigenvectors corresponding to different eigenvalues are orthogonal and eigenvectors can be chosen to form an orthonormal basis for $\vec{\mathcal{H}}$.

20.4 Operators with a Continuous Spectrum

In $\vec{L}^2(I\!R)$ the operators $\widehat{x}(I\!R)$ and $\widehat{p}(I\!R)$ given respectively by Eqs. (17.12) and (17.49) are examples of selfadjoint operators with a purely continuous spectrum. In view of their importance in quantum theory we shall study them in details. For brevity of notation we shall simplify the notation by denoting $\widehat{x}(I\!R)$ and $\widehat{p}(I\!R)$ in $\vec{L}^2(I\!R)$ by $\widehat{x}$ and $\widehat{p}$ and their Fourier transforms by $\widehat{\underset{\sim}{x}}$ and $\widehat{\underset{\sim}{p}}$.

A characteristic feature of a selfadjoint operator $\widehat{A}$ with a purely continuous spectrum is that its spectral measure for a singleton set $\{\tau_0\}$ is zero, i.e., for any τ_0 in the spectrum of the operator we have[13]

$$\widehat{M}^{\hat{A}}\big(\{\tau_0\}\big) := \widehat{F}^{\hat{A}}(\tau_0) - \widehat{F}^{\hat{A}}(\tau_0 - 0) = \widehat{0}. \tag{20.23}$$

This property can be seen clearly in $\widehat{x}(I\!R)$ and $\widehat{p}(I\!R)$.

[12] *See* §35.2 for a derivation of these eigenvalues.

[13] *See* Eq. (15.20) for the expression for $\widehat{M}^{\hat{A}}\big(\{\tau_0\}\big)$.

20.4.1 The Position Operator $\widehat{x}$ in $\vec{L}^2(I\!R)$

Theorem 20.4.1(1)[14] *The spectral function $\widehat{F}^{\hat{x}}(\tau)$ of $\widehat{x}$ is defined by the one-parameter family of projectors $\widehat{\chi}_{(-\infty,\tau]}$ in* Eq. (20.1)*, i.e.,*

$$\widehat{F}^{\hat{x}}(\tau) := \widehat{\chi}_{(-\infty,\tau]}, \quad \tau \in I\!R. \tag{20.24}$$

Proof We have already shown that the family of projectors $\widehat{\chi}_{(-\infty,\tau]}$ is a spectral function. To relate to the position operator we have to show that Eqs. (20.10) and (20.11) of Theorem 20.1(1) are satisfied. To verify Eq. (20.10) we have, for $\vec{\phi} \in \mathcal{D}(\widehat{x})$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \tau^2 \, d_\tau \left(\langle \vec{\phi} \mid \widehat{F}^{\hat{x}}(\tau) \, \vec{\phi} \, \rangle \right) \\
&:= \int_{-\infty}^{\infty} \tau^2 \, d_\tau \int_{-\infty}^{\tau} \phi^*(x)\phi(x) \, dx = \int_{-\infty}^{\infty} \tau^2 \, \phi^*(\tau)\phi(\tau) \, d\tau \\
&= || \, \widehat{x}\vec{\phi} \, ||^2.
\end{aligned} \tag{20.25}$$

Equation (20.11) can be similarly verified. **QED**

The spectral decomposition in Eq. (20.13) of $\widehat{x}$ becomes

$$\widehat{x} = \int_{-\infty}^{\infty} \tau \, d\widehat{F}^{\hat{x}}(\tau). \tag{20.26}$$

To appreciate this expression we can carry out the following calculation for any given $\vec{\psi} \in \vec{\mathcal{H}}$ and $\vec{\phi} \in \mathcal{D}(\widehat{x})$:

$$\begin{aligned}
&\langle \vec{\psi} \mid \left(\int_{-\infty}^{\infty} \tau \, d\widehat{F}^{\hat{x}}(\tau) \right) \vec{\phi} \, \rangle = \int_{-\infty}^{\infty} \tau \, d_\tau \langle \vec{\psi} \mid \widehat{F}^{\hat{x}}(\tau) \, \vec{\phi} \rangle \\
&:= \int_{-\infty}^{\infty} \tau \, d_\tau \left(\int_{-\infty}^{\infty} \psi^*(x) \chi_{(-\infty,\tau]}(x)\phi(x) \, dx \right) \\
&= \int_{-\infty}^{\infty} \tau \, d_\tau \int_{-\infty}^{\tau} \psi^*(x)\phi(x) \, dx = \int_{-\infty}^{\infty} \tau \psi^*(\tau)\phi(\tau) \, d\tau \\
&= \langle \vec{\psi} \mid \widehat{x} \, \vec{\phi} \, \rangle.
\end{aligned} \tag{20.27}$$

Spectral projectors correspond to characteristic functions on $I\!R$, i.e.,

$$\widehat{M}^{\hat{x}}(\Lambda) := \widehat{\chi}_{\Lambda} \quad \text{for all Borel set } \Lambda \text{ of } I\!R. \tag{20.28}$$

For $\Lambda = \{\tau_0\}$ we have $\widehat{M}^{\hat{x}}(\{\tau_0\}) = \widehat{\chi}_{(-\infty,\tau_0]} - \widehat{\chi}_{(-\infty,\tau_0-0]} = \widehat{0}$.

[14] Jordan pp. 42–43. Roman Vol. 2 pp. 467–468.

20.4.2 The Momentum Operator $\widehat{p}$ in $\vec{L}^2(I\!R)$

The operator $\widehat{p}$ in $\vec{L}^2(I\!R)$ is a differential operator. The momentum operator becomes a multiplication operator $\underset{\sim}{\widehat{p}}$ in the momentum representation space $\vec{L}^2(\underset{\sim}{I\!R})$, in the same way that the position operator is a multiplication operator in $\vec{L}^2(I\!R)$. It follows that we can immediately write down the spectral function $\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)$ of $\underset{\sim}{\widehat{p}}$ in the momentum representation space $\vec{L}^2(\underset{\sim}{I\!R})$.

Theorem 20.4.2(1) *The spectral function $\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)$ of $\underset{\sim}{\widehat{p}}$ in the momentum representation space $\vec{L}^2(\underset{\sim}{I\!R})$ is defined by the characteristic function $\underset{\sim}{\chi}_{(-\infty,\tau]}(p)$ of the interval $(-\infty, \tau]$ on the momentum space $\underset{\sim}{I\!R}$ acting as a multiplication operator, i.e.,*

$$\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)\vec{\underset{\sim}{\varphi}} := \underset{\sim}{\chi}_{(-\infty,\tau]}(p)\underset{\sim}{\varphi}(p). \tag{20.29}$$

The operator $\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)$ has the effect of truncating the function $\tilde{\phi}(p)$ on momentum space beyond the point $p = \tau$, in the same way the operator $\widehat{F}^{\hat{x}}(\tau)$ truncates the function $\phi(x)$ on coordinate space beyond the point $x = \tau$. Spectral projectors corresponds to characteristic functions on the momentum space $\underset{\sim}{I\!R}$, i.e.,

$$\underset{\sim}{\widehat{M}}^{\underset{\sim}{\hat{p}}}(\Lambda) := \underset{\sim}{\widehat{\chi}}_{\underset{\sim}{\Lambda}}. \tag{20.30}$$

for all Borel set $\underset{\sim}{\Lambda}$ of $\underset{\sim}{I\!R}$.

Since $\widehat{p}$ in $\vec{L}^2(I\!R)$ is unitarily related by a Fourier transformation to $\underset{\sim}{\widehat{p}}$ in $\vec{L}^2(\underset{\sim}{I\!R})$ we can obtain the spectral function $\widehat{F}^{\hat{p}}(\tau)$ of $\widehat{p}$ in the coordinate representation space $\vec{L}^2(I\!R)$ as the inverse Fourier transform $\widehat{U}_F^{-1}\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)\widehat{U}_F$ of $\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)$ in accordance with Theorem 15.3(2), i.e., we have

$$\begin{aligned}
\widehat{F}^{\hat{p}}(\tau)\vec{\varphi} &= \left(\widehat{U}_F^{-1}\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)\widehat{U}_F\right)\vec{\varphi} = \widehat{U}_F^{-1}\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)\left(\widehat{U}_F\vec{\varphi}\right) \\
&= \widehat{U}_F^{-1}\left(\underset{\sim}{\widehat{F}}^{\underset{\sim}{\hat{p}}}(\tau)\vec{\underset{\sim}{\varphi}}\right) = \widehat{U}_F^{-1}\left(\underset{\sim}{\widehat{\chi}}_{(-\infty,\tau]}(p)\,\vec{\underset{\sim}{\varphi}}\right) \\
&:= \int_{-\infty}^{\infty} f_p(x)\,\underset{\sim}{\chi}_{(-\infty,\tau]}(p)\,\underset{\sim}{\varphi}(p)\,dp \\
&= \int_{-\infty}^{\tau} f_p(x)\,\underset{\sim}{\varphi}(p)\,dp.
\end{aligned} \tag{20.31}$$

The result is

$$\widehat{F}^{\hat{p}}(\tau)\vec{\varphi} := \int_{-\infty}^{\tau} f_p(x)\, \underset{\sim}{\varphi}(p)\, dp. \tag{20.32}$$

This expression shows that $\widehat{F}^{\hat{p}}(\tau)$ has the effect of

taking the Fourier transform $\underset{\sim}{\varphi}(p)$ *of* $\varphi(x)$ *and then taking a "semi-inverse transform" up to the value* τ.

We can also verify directly that the above expression for $\widehat{F}^{\hat{p}}(\tau)$ satisfies the properties required by the spectral theorem. Again the spectral projector for a singleton set is equal to the zero operator.

20.5 Functions of Selfadjoint Operators

20.5.1 Characteristic Functions

Definitions 15.4(1) and 15.4(2) on functions of selfadjoint operators apply. For operators with a discrete spectrum Definition 15.4(1) reduces to Definition 13.3.3(1).[15] Let us illustrate this with the example of the number operator $\widehat{N}$ defined by Eq. (19.5) in a Hilbert space $\vec{\mathcal{H}}$ which is relevant to the discussion in §27.9. This operator is defined in terms of a pair of annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$ which are defined on a given complete orthonormal basis $\{\vec{\varphi}_n,\ n = 0, 1, 2, \ldots\}$ on a common domain $\vec{\mathcal{D}}(\widehat{a})$ given by Eq. (17.112). The number operator, being the product of $\widehat{a}^\dagger$ and $\widehat{a}$ is defined on a smaller domain, i.e.,

$$\vec{\mathcal{D}}(\widehat{N}) := \Big\{\vec{\phi} \in \vec{\mathcal{H}} : \sum_{n=0}^{\infty} |\langle \vec{\varphi}_n \mid \vec{\phi}\rangle|^2\, n^2 < \infty\Big\}, \tag{20.33}$$

with a spectral decomposition

$$\widehat{N} = \sum_{n=0}^{\infty} n\, \widehat{P}^{\hat{N}}(n) = \sum_{n=0}^{\infty} n\, |\vec{\varphi}_n\rangle\langle\vec{\varphi}_n|. \tag{20.34}$$

[15] Generally the sum will be over an infinite number of terms since there may well be an infinite number of eigenvalues.

This operator is positive in the sense of Definition 19.1(2).[16] We can define its square root by

$$\widehat{N}^{\frac{1}{2}} := \sum_{n=0}^{\infty} \sqrt{n}\, \widehat{P}^{\hat{N}}(n). \tag{20.35}$$

This operator is defined on a domain which is different from $\vec{\mathcal{D}}(\widehat{N})$. One can verify that the domain $\vec{\mathcal{D}}(\widehat{N}^{\frac{1}{2}})$ of $\widehat{N}^{\frac{1}{2}}$ is the same as that of the annihilation operator.[17]

For operators with a continuous spectrum, e.g., the position and momentum operators, the situation is different. The integral expression in Definition 15.4(1) does not reduce to a sum.

Consider an example in $\vec{L}^2(I\!R)$. The characteristic function is a real-valued function $\chi_{(-\infty,\tau]}(x)$ on $I\!R$. We can define the characteristic function $\chi_{(-\infty,\tau]}(\widehat{A})$ of a selfadjoint operator $\widehat{A}$ in accordance with Definition 15.4(1). This function is of particular significance when $\widehat{A} = \widehat{x}$. We have the operator $\chi_{(-\infty,\tau]}(\widehat{x})$ acting on every $\vec{\phi} \in \vec{L}(I\!R)$ with a formal expression given by Eq. (15.48) as

$$\chi_{(-\infty,\tau]}(\widehat{x})\vec{\phi} := \int_{-\infty}^{\infty} \chi_{(-\infty,\tau]}(x)\, d_\tau \widehat{F}^{\hat{x}}(\tau)\vec{\phi}. \tag{20.36}$$

We can evaluate the quadratic form generated by this operator to see more explicitly what the operator is, i.e., we have

$$\begin{aligned}
\langle \vec{\phi} \mid \chi_{(-\infty,\tau]}(\widehat{x})\vec{\phi} \rangle &= \langle \vec{\phi} \mid \left(\int_{-\infty}^{\infty} \chi_{(-\infty,\tau]}(x)\, d_\tau \widehat{F}^{\hat{x}}(\tau) \right) \vec{\phi} \rangle \\
&= \int_{-\infty}^{\infty} \chi_{(-\infty,\tau]}(x)\, d_\tau \left(\langle \vec{\phi} \mid \widehat{F}^{\hat{x}}(\tau)\vec{\phi} \rangle \right) \\
&:= \int_{-\infty}^{\infty} \chi_{(-\infty,\tau]}(x)\, d_\tau \int_{-\infty}^{\tau} \phi^*(x)\phi(x)\, dx \\
&= \int_{-\infty}^{\infty} \chi_{(-\infty,\tau]}(x)\, \phi^*(\tau)\phi(\tau)\, d\tau \\
&= \langle \vec{\phi} \mid \widehat{\chi}_{(-\infty,\tau]}\vec{\phi} \rangle.
\end{aligned} \tag{20.37}$$

[16] *See* also Definition 13.3.1(1). A selfadjoint operator is positive if its spectrum in non-negative.

[17] The domain of the annihilation operator $\vec{\mathcal{D}}(\widehat{a})$ is given by Eq. (17.112).

We can conclude by Eq. (18.1) that

$$\chi_{(-\infty,\tau]}(\widehat{x}) = \widehat{\chi}_{(-\infty,\tau]} = \widehat{F}^{\hat{x}}(\tau). \tag{20.38}$$

In other words, the spectral function $\widehat{F}^{\hat{x}}(\tau)$ of the position operator $\widehat{x}$ equal to the characteristic function $\chi_{(-\infty,\tau]}(\widehat{x})$ of the operator.

It turns out that *the spectral function of a selfadjoint operator is generally expressible in terms of a characteristic function of that operator*, i.e., the spectral function of a selfadjoint operator $\widehat{A}$ is the characteristic function $\chi_{(-\infty,\tau]}(\widehat{A})$ of $\widehat{A}$ with the spectral projector associated with a Borel set Λ is given by $\chi_{\Lambda}(\widehat{A})$, i.e., we have[18]

$$\widehat{F}^{\hat{A}}(\tau) = \chi_{(-\infty,\tau]}(\widehat{A}) \quad \text{and} \quad \widehat{M}^{\hat{A}}(\Lambda) = \chi_{\Lambda}(\widehat{A}), \tag{20.39}$$

since

$$\begin{aligned} \chi_{\Lambda}(\widehat{A}) &= \int_{-\infty}^{\infty} \chi_{\Lambda}(\tau) d\widehat{F}^{\hat{A}}(\tau) \\ &= \int_{\Lambda} d\widehat{F}^{\widehat{A}}(\tau) = \widehat{M}^{\hat{A}}(\Lambda). \end{aligned} \tag{20.40}$$

20.5.2 Complex-Valued Functions

When the function is complex the resulting operator is not selfadjoint. Of particular interest are bounded functions. Let $f(\tau)$ and $g(\tau)$ be two bounded functions and let $f^*(\tau)$ and $g^*(\tau)$ be their complex conjugate functions. We have the following results.[19]

Theorem 20.5(1)

$$[f(\widehat{A}), g(\widehat{A})] = \widehat{0}. \tag{20.41}$$

$$(f+g)(\widehat{A}) = f(\widehat{A}) + g(\widehat{A}). \tag{20.42}$$

$$(fg)(\widehat{A}) = f(\widehat{A})g(\widehat{A}). \tag{20.43}$$

$$f^{\dagger}(\widehat{A}) = f^*(\widehat{A}). \tag{20.44}$$

The following comments clarify the theorem:

C20.5(1) $f(\widehat{A})$ and $g(\widehat{A})$ commute.

[18] Prugovečki Eq. (2.16) on p. 277. Akhiezer and Glazman Vol. 2 (1963) p. 72.
[19] Roman Vol. 2 p. 647. Prugovečki p. 275.

C20.5(2) The sum $f(\tau) + g(\tau)$ defines an operator $(f + g)(\widehat{A})$ which is equal to the sum of operators $f(\widehat{A})$ and $g(\widehat{A})$.

C20.5(3) The product $f(\tau)g(\tau)$ defines an operator $(fg)(\widehat{A})$ which is equal to the product of operators $f(\widehat{A})$ and $g(\widehat{A})$.

C20.5(4) The adjoint of the operator $f(\widehat{A})$ is equal to the operator $f^*(\widehat{A})$ defined by the function $f^*(\tau)$.[20]

An important example is the complex exponential function and its conjugate

$$u(\tau) = e^{ia\tau} \quad \text{and} \quad u^*(\tau) = e^{-ia\tau}, \tag{20.45}$$

where a is a real number. We have

$$u(\tau)u^*(\tau) = u^*(\tau)u(\tau) = f_{uc}(\tau) = 1, \tag{20.46}$$

where $f_{uc}(\tau)$ is the unit constant function introduced in relation to Eq. (13.39). The discussion in §13.3.3 shows that these exponential functions define a unitary operator and its adjoint. In an infinite-dimensional Hilbert space $\vec{\mathcal{H}}$ we have similar results. First we can define two operators $u(\widehat{A})$ and $u^*(\widehat{A})$ by

$$u(\widehat{A}) := \int_{-\infty}^{\infty} e^{ia\tau}\, d\widehat{F}^{\hat{A}}(\tau). \tag{20.47}$$

$$u^*(\widehat{A}) := \int_{-\infty}^{\infty} e^{-ia\tau}\, d\widehat{F}^{\hat{A}}(\tau). \tag{20.48}$$

These operators have the following properties:

(1) The adjoint of $u(\widehat{A})$ is $u^*(\widehat{A})$, i.e., $u^\dagger(\widehat{A}) = u^*(\widehat{A})$ by Eq. (20.44).[21]

(2) As seen in Eq. (13.39) the unit constant function $f_{uc}(\tau) = 1$ defines an operator function $f_{uc}(\widehat{A})$ which is equal to the

[20] Roman Vol. 2 p. 647.

[21] Note that $u^*(\widehat{A})$ is the operator defined by the function $u^*(\tau)$, i.e., by Eq. (20.48), while $u^\dagger(\widehat{A})$ is the adjoint of $u(\widehat{A})$.

identity operator $\widehat{I\!I}$ on $\vec{\mathcal{H}}$. It follows from Eqs. (20.43) and (20.46) that

$$u(\widehat{A})u^*(\widehat{A}) = (u^*u)(\widehat{A}) = f_{uc}(\widehat{A}) = \widehat{I\!I}. \tag{20.49}$$

$$u^*(\widehat{A})u(\widehat{A}) = (uu^*)(\widehat{A}) = f_{uc}(\widehat{A}) = \widehat{I\!I}. \tag{20.50}$$

It follows from Eq. (20.44) that the adjoint $u^\dagger(\widehat{A}) = u^*(\widehat{A})$ is the inverse of $u(\widehat{A})$. Hence $u(\widehat{A})$ is unitary.[22]

(3) We can rewrite as $u(\widehat{A})$ and $u^*(\widehat{A})$ explicitly as

$$e^{ia\hat{A}} := \int_{-\infty}^{\infty} e^{i a\tau}\, d\widehat{F}^{\hat{A}}(\tau), \tag{20.51}$$

$$e^{-ia\hat{A}} := \int_{-\infty}^{\infty} e^{-i a\tau}\, d\widehat{F}^{\hat{A}}(\tau). \tag{20.52}$$

The unit constant function $f_{uc}(\tau)$ can be regarded as an exponential function with a zero exponential.

Explicit examples such exponential functions are further discussed in §21.1.

20.5.3 Spectral Functions and Spectral Measures

The spectral function and spectral measure of a real-valued function $f(\widehat{A})$ of a selfadjoint operator $\widehat{A}$ can be expressed in terms of the corresponding quantities of $\widehat{A}$, i.e., we have[23]

$$\widehat{M}^{f(\hat{A})}(\Lambda) = \widehat{M}^{\hat{A}}\Big(f^{-1}(\Lambda)\Big), \tag{20.53}$$

$$\widehat{F}^{f(\hat{A})}(\tau) = \widehat{M}^{f(\hat{A})}\Big((-\infty, \tau]\Big) = \widehat{M}^{\hat{A}}\left(f^{-1}\Big((-\infty, \tau]\Big)\right), \tag{20.54}$$

where

$$f^{-1}(\Lambda) := \Big\{\tau \in I\!R : f(\tau) \in \Lambda\Big\} \tag{20.55}$$

is the inverse image of Λ under $f(x)$. An example is when $f = x^2$. We have

$$f^{-1}\Big((\tau_1, \tau_2]\Big) := \Big\{\tau \in I\!R : \tau^2 \in (\tau_1, \tau_2]\Big\}. \tag{20.56}$$

[22] Roman Vol. 2 p. 656.
[23] Weidmann (1980) p. 197. Wan p. 156.

When $\tau_1 = -\infty$ we have

$$f^{-1}\Big((-\infty, \tau_2]\Big) = \begin{cases} \emptyset & \text{if } \tau_2 < 0 \\ [\,-\sqrt{\tau_2}, \sqrt{\tau_2}\,] & \text{if } \tau_2 \geq 0 \end{cases}. \tag{20.57}$$

These enable us to work out $\widehat{M}^{\hat{A}^2}$ and $\widehat{F}^{\hat{A}^2}$ in terms of $\widehat{M}^{\hat{A}}$ and $\widehat{F}^{\hat{A}^2}$ from Eqs. (20.53) and (20.54).

20.6 Complete Set of Commuting Selfadjoint Operators

The discussions on commuting selfadjoint operators in §13.3.3 need to be generalised in order to apply to infinite-dimensional Hilbert spaces due to the unbounded nature of some operators. As the discussion in §17.7 shows there are complications arising from the operators involved having different domains. In the case of selfadjoint operators we can avoid domain problems by defining commutativity in terms of Eq. (15.43).

Definition 20.6(1)[24] *Two selfadjoint operators $\widehat{A}$ and $\widehat{B}$ are said to commute if their respective spectral projectors $\widehat{M}^{\hat{A}}(\Lambda_1)$ and $\widehat{M}^{\hat{B}}(\Lambda_2)$ commute, i.e.,*

$$[\,\widehat{M}^{\hat{A}}(\Lambda_1),\ \widehat{M}^{\hat{B}}(\Lambda_2)] = \widehat{0} \tag{20.58}$$

for all Borel sets Λ_1 and Λ_2 of $\mathbb{R}$.

The following comments help to clarify the definition:

C20.6(1) This definition applies irrespective of whether the spectra of the operators are discrete or continuous.

C20.6(2) If $\widehat{A}$ and $\widehat{B}$ commute then their respective spectral functions $\widehat{F}^{\hat{A}}(\tau_1)$ and $\widehat{F}^{\hat{B}}(\tau_2)$ also commute.

C20.6(3) If $\widehat{A}$ and $\widehat{B}$ are both bounded then the definition is equivalent to Eq. (17.84).

C20.6(4) If both $\widehat{A}$ and $\widehat{B}$ have a discrete spectrum then Theorems 13.3.4(1) to 13.3.4(3) also apply.

[24] Fano p. 405. Prugovečki p. 261.

C20.6(5) Theorem 13.3.4(4) applies generally whether the spectra of the operators are discrete or continuous.

C20.6(6) Definition 15.4(2) applies to define functions of commuting selfadjoint operators.

We shall now introduce the important concept of *a complete set of commuting selfadjoint operators* in a Hilbert space step by step:

(1) Let $\widehat{A}$ and $\widehat{B}$ be two selfadjoint operators with a discrete spectrum in a Hilbert space $\vec{\mathcal{H}}$ with their respective spectral decompositions[25]

$$\widehat{A} = \sum_{\ell} a_\ell \, \widehat{P}^{\hat{A}}(a_\ell) \quad \text{and} \quad \widehat{B} = \sum_{m} b_m \widehat{P}^{\hat{B}}(b_m). \tag{20.59}$$

(2) Suppose that $\widehat{A}$ and $\widehat{B}$ commute. Then, as pointed out in the comments after Theorem 13.3.4(1), the product $\widehat{P}^{\hat{A}}(a_\ell)\widehat{P}^{\hat{B}}(b_m)$ is a projector, and that given any vector $\vec{\phi}$ the projection

$$\vec{\phi}_{\ell m} = \big(\widehat{P}^{\hat{A}}(a_\ell)\widehat{P}^{\hat{B}}(b_m)\big)\vec{\phi} \tag{20.60}$$

is an eigenvector of both $\widehat{A}$ and $\widehat{B}$, i.e., we have[26]

$$\widehat{A}\vec{\phi}_{\ell m} = a_\ell \, \vec{\varphi}_{\ell m} \quad \text{and} \quad \widehat{B}\vec{\phi} = b_m \vec{\phi}_{\ell m}. \tag{20.61}$$

Generally a pair of eigenvalues a_ℓ and b_m is unable to determine a single eigenvector. Given a different vector $\vec{\phi}'$ Eq. (20.60) may lead to a different eigenvector $\vec{\phi}'_{\ell m}$, since the projector $\widehat{P}^{\hat{A}}(a_\ell)\widehat{P}^{\hat{B}}(b_m)$ may not be one-dimensional.

(3) It may be possible to find an additional selfadjoint operator $\widehat{C}$ to form a set of three commuting selfadjoint operators so that any set of their eigenvalues a_ℓ, b_m, c_n would determine a single common eigenvector. Such as set is called a *complete set*.[27]

[25] As in Eq. (20.20) the eigenvalues for each operator are meant to be different.
[26] Jordan p. 53.
[27] If the set of eigenvalues a_i, b_j, c_k is unable to determine a single eigenvector, we can go on to add another selfadjoint operator to repeat the process.

Definition 20.6(2)[28] *A set* $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, ... *of selfadjoint operators with a discrete spectrum in a Hilbert space* $\vec{\mathcal{H}}$ *is called a* **complete set of commuting selfadjoint operators** *in* $\vec{\mathcal{H}}$ *if:*

(1) *The operators mutually commute.*

(2) *Every set of eigenvalues* a_ℓ, b_m, c_n, ... *of the operators has only a single common eigenvector* $\vec{\varphi}_{\ell mn\ldots}$ *associated with it, i.e., there is only one unit vector* $\vec{\varphi}_{\ell mn\ldots}$ *such that*[29]

$$\widehat{A}\vec{\varphi}_{\ell mn\cdots} = a_\ell\,\vec{\varphi}_{\ell mn\cdots}, \quad \widehat{B}\vec{\varphi}_{\ell mn\cdots} = b_m\,\vec{\varphi}_{\ell mn\cdots}, \quad \cdots . \tag{20.62}$$

(3) *The eigenvectors* $\vec{\varphi}_{\ell mn\ldots}$ *form an orthonormal basis in* $\vec{\mathcal{H}}$.

The purpose of such a set of operators, often referred to as a *complete set* for short, is to determine a single eigenvector by a set of eigenvalues. A selfadjoint operator with a discrete and nondegenerate spectrum serves this purpose and it constitutes a complete set on its own. A complete set is not unique. The followings are examples:

E20.6(1) The number operator $\widehat{N}$ in a Hilbert space $\vec{\mathcal{H}}$ given by Definition 19.1(5) constitutes a complete set in $\vec{\mathcal{H}}$.

E20.6(2) In $\vec{L}^2(\Lambda)$ the operator $\widehat{P}_{\lambda=0}(\Lambda)$ constitutes a complete set and so does $\widehat{H}^{\infty}_{D}(\Lambda)$.

E20.6(3) In $\vec{L}^2(I\!R)$ the operator $\widehat{H}_{ho}$ forms a complete set.

E20.6(4) The operator $\widehat{P}^2_{\lambda=0}(\Lambda)$ has a degenerate spectrum and hence it does not form a complete set on its own.

E20.6(5) In $\vec{L}^2(\mathcal{S}_u)$ the operator $\widehat{L}_z(\mathcal{S}_u)$ defined by Eq. (17.42) has a degenerate spectrum and hence it does not form a complete set on its own. The same applies to the operator $\widehat{L}^2(\mathcal{S}_u)$ defined by Eq. (19.51). Together the set $\{\,\widehat{L}_z(\mathcal{S}_u),\ \widehat{L}^2(\mathcal{S}_u)\,\}$ does form a complete set in $\vec{L}^2(\mathcal{S}_u)$.

---

[28] Jordan pp. 55–61. Prugovečki pp. 312–317. Beltrametti and Cassinelli pp. 19–22. Isham p. 99.

[29] A vector which differs from $\vec{\varphi}_{\ell mn\ldots}$ by a multiplicative constant is not considered to be a different eigenvector.

Theorem 13.3.4(4) which remains valid in a Hilbert space. The theorem can be extended to a complete set of selfadjoint operators by the following theorem.

Theorem 20.6(1)[30] *A set of mutually commuting selfadjoint operators with a discrete spectrum in a Hilbert space $\vec{\mathcal{H}}$ is a complete commuting set of selfadjoint operators in $\vec{\mathcal{H}}$ if and only if every bounded operator on $\mathcal{H}$ which commutes with every operator in the set is a function of the operators of the set.*

This theorem enable us to have a general definition of a complete set of operators, including operators with a continuous spectrum.[31]

Definition 20.6(3)[32] *A set $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, ... of mutually commuting selfadjoint operators in a Hilbert space $\vec{\mathcal{H}}$ is called a* **complete commuting set of selfadjoint operators** *in $\mathcal{H}$ if every bounded operator commuting with every member of the set is a function of the operators of the set.*

It can be shown that the position operator $\widehat{x}$ in $\vec{L}^2(I\!R)$ satisfies the condition of the above theorem, i.e., it forms a complete set on its own in $\vec{L}^2(I\!R)$.[33] Similarly the momentum operator $\widehat{p}$ also constitutes a complete in in $\vec{L}^2(I\!R)$ on its own. These examples show a link with the definition in terms of eigenfunctions and eigenvalues. For example, the spectrum of the momentum operator may be considered nondegenerate in that for each generalised eigenvalue p there corresponds to a single generalised eigenfunction $f_p(x)$ given by Eq. (18.12). In $\vec{L}^2(I\!R^3)$ the position operators $\widehat{x}$, $\widehat{y}$, $\widehat{z}$ form complete set and so does the momentum operators $\widehat{p}_x$, $\widehat{p}_y$, $\widehat{p}_z$.

For the Hilbert spaces used for the description of quantum mechanical systems we shall assume that a finite complete set of commuting selfadjoint operators exists.[34]

---

[30] Jordan pp. 55–61. Prugovečki pp. 312–317. Riesz and Nagy pp. 356–358.
[31] Definition 20.6(2) does not apply to operators with a continuous spectrum.
[32] Jordan pp. 55–61. Prugovečki pp. 312–317. Beltrametti and Cassinelli pp. 19–22. Isham p. 99.
[33] Jordan p. 58. *See* solution to Q20(7).
[34] Prugovečki p. 315.

20.7 Irreducible Sets of Selfadjoint Operators

A complete set of commuting selfadjoint operators is quite different from a set of irreducible selfadjoint operators introduced in Definition 17.9(4). In $\vec{L}^2(I\!R)$ the position operator $\widehat{x}$ constitutes a complete set. But this operator is not irreducible. Theorem 17.9(1) tells us that the position operator $\widehat{x}$ is reducible, since it commutes with its spectral projectors. Generally a spectral subspace of a selfadjoint operator reduces the operator. For the position operator $\widehat{x}$ in $\vec{L}^2(I\!R)$ the set $\vec{\mathcal{S}}^{\widehat{x}}(\Lambda)$ defined by all the functions in $L^2(I\!R)$ vanishing outside an interval $\Lambda = [0, L]$ is a spectral subspace. This subspace $\vec{\mathcal{S}}^{\widehat{x}}(\Lambda)$ reduces $\widehat{x}$. A similar result applies to the momentum operator $\widehat{p}$ in $\vec{L}^2(I\!R)$.

Another example is the number operator $\widehat{N}$ in Eq. (19.5). This operator is reduced by any of its spectral subspaces, and so is its square root $\widehat{N}^{\frac{1}{2}}$.

There is a useful theorem on the reducibility or otherwise of a set of operators.

Theorem 20.7(1)[35] *A set of selfadjoint operators is irreducible if and only the only bounded operators which commute with every operator of the set are multiples of the identity operator.*

The following are useful examples:

E20.7(1) The position and the momentum operators form an irreducible set in $\vec{L}^2(I\!R)$. There is no operator, apart from multiples of the identity, which commutes with both $\widehat{x}$ and $\widehat{p}$. To appreciate this result we first observe that bounded operators commuting with $\widehat{x}$ must be functions of $\widehat{x}$, since the position operator constitutes a complete set. Functions of $\widehat{x}$ cannot commute with $\widehat{p}$ unless it is a constant function, e.g., the unit constant function which maps $\widehat{x}$ to the identity operator. It follows from Theorem 20.7(1) that the set $\{\widehat{x}, \widehat{p}\}$ in $\vec{L}^2(I\!R)$ is irreducible. There are no subspace of $\vec{L}^2(I\!R)$ which is invariant under both $\widehat{x}$ and $\widehat{p}$.

[35] Jordan p. 68. Multiples of the identity operator mean $a\widehat{I\!I}$, $a \in \mathbb{C}$.

E20.7(2) The annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$ in a Hilbert space $\vec{\mathcal{H}}$ defined by Definitions 17.10(1) and 17.10(2) also form an irreducible set of operators. Let us prove this step by step.

(1) Suppose there is a subspace $\vec{\mathcal{S}}$ of $\vec{\mathcal{H}}$ which reduces both $\widehat{a}$ and $\widehat{a}^\dagger$. Theorem 17.9(1) tells us that the projector $\widehat{P}_{\vec{\mathcal{S}}}$ onto this subspace would commute with $\widehat{a}$ and $\widehat{a}^\dagger$. It follows that $\widehat{P}_{\vec{\mathcal{S}}}$ commutes with the number operator $\widehat{N} = \widehat{a}^\dagger \widehat{a}$.

(2) The number operator has a nondegenerate spectrum and hence it forms a complete set on its own. Its spectral decomposition is given by Eq. (20.34) where the eigenprojectors $\widehat{P}^{\hat{N}}(n)$ are all one-dimensional. Theorem 20.6(1) then tells us that the projector $\widehat{P}_{\vec{\mathcal{S}}}$ must be a function of $\widehat{N}$, i.e.,

$$\widehat{P}_{\vec{\mathcal{S}}} = f(\widehat{N}) = \sum_{n=0}^{\infty} f(n) \widehat{P}^{\hat{N}}(n), \tag{20.63}$$

where f is a real-valued function on $\mathbb{R}$.

(3) Since $\widehat{P}_{\vec{\mathcal{S}}}$ is idempotent we must have

$$\begin{aligned} f^2(\widehat{N}) = f(\widehat{N}) \;\; &\Rightarrow \;\; \sum_{n=0}^{\infty} f^2(n) \widehat{P}^{\hat{N}}(n) = \sum_{n=0}^{\infty} f(n) \widehat{P}^{\hat{N}}(n) \\ &\Rightarrow \;\; f^2(n) = f(n) \;\; \Rightarrow \;\; f(n) = 1 \;\text{ or }\; f(n) = 0. \end{aligned} \tag{20.64}$$

(4) We can conclude that the projector is of the form

$$\widehat{P}_{\vec{\mathcal{S}}} = \sum_{n \in J} \widehat{P}^{\hat{N}}(n), \tag{20.65}$$

for some set J of non-negative integers. In other words, $\widehat{P}_{\vec{\mathcal{S}}}$ is a spectral projector of $\widehat{N}$ and $\vec{\mathcal{S}}$ is a spectral subspace of $\widehat{N}$.

(5) A spectral subspace of $\widehat{N}$ cannot be invariant under both $\widehat{a}$ and $\widehat{a}^\dagger$. For example, the spectral subspace spanned by $\vec{\varphi}_1$ cannot be invariant under $\widehat{a}$ since $\widehat{a}\,\vec{\varphi}_1 = \vec{\varphi}_0$ would lie outside the subspace.

(6) We have arrived at a contradiction which would imply the non-existence of a subspace which can reduce $\widehat{a}$ and $\widehat{a}^\dagger$. In other words, $\widehat{a}$ and $\widehat{a}^\dagger$ form an irreducible set in $\vec{\mathcal{H}}$.

Exercises and Problems

Q20(1) Show that the spectral functions of a projector $\widehat{P}$, of the zero operator $\widehat{0}$ and of the identity operator $\widehat{I\!I}$ are given, respectively by[36]

$$\widehat{F}^{\hat{0}}(\tau) = \begin{cases} \widehat{0} & \tau < 0 \\ \widehat{I\!I} & \tau \geq 0 \end{cases} . \tag{20.66}$$

$$\widehat{F}^{\hat{I\!I}}(\tau) = \begin{cases} \widehat{0} & \tau < 1 \\ \widehat{I\!I} & \tau \geq 1 \end{cases} . \tag{20.67}$$

$$\widehat{F}^{\hat{P}}(\tau) = \begin{cases} 0 & \tau < 0 \\ \widehat{I\!I} - \widehat{P} & 0 \leq \tau < 1 \\ \widehat{I\!I} & \tau \geq 1 \end{cases} . \tag{20.68}$$

Q20(2) Show that on an interval Λ of $I\!R$ the spectral measure $\widehat{M}^{\hat{P}}(\Lambda)$ of a projector $\widehat{P}$ is related to $\widehat{P}$ by[37]

$$\widehat{M}^{\hat{P}}(\Lambda) = \begin{cases} \widehat{I\!I} - \widehat{P} & \text{if } \Lambda = \{0\} \\ \widehat{P} & \text{if } \Lambda = \{1\} \\ \widehat{0} & \text{if } \Lambda \text{ does not contain 0 or 1} \end{cases} \tag{20.69}$$

Q20(3) Prove that Theorem 9.4.4(1) remains valid for a selfadjoint operator with a discrete spectrum in an infinite-dimensional Hilbert space.

Q20(4) In $\vec{L}^2(I\!R)$ show that for every $\phi(x) \in C_c^\infty(I\!R)$ we have

$$[\, f(\hat{x}),\, \hat{p}(I\!R)\,]\vec{\phi} := i\hbar \frac{df(x)}{dx}\, \phi(x). \tag{20.70}$$

Q20(5) Show that selfadjoint operators having a discrete spectrum are reducible by their eigensubspaces.

Q20(6) Show that in $\vec{L}^2(I\!R)$ the position operator $\hat{x}(I\!R)$ is reducible and the momentum operator $\hat{p}(I\!R)$ is also reducible.[38]

Q20(7) Show that the position operator $\hat{x}$ in $\vec{L}^2(I\!R)$ constitutes a complete set of selfadjoint operators.

---

[36] Weidmann p. 195. Wan p. 152.

[37] Wan p. 152. *See* Q28(2) for an application.

[38] This is in contrast to the fact discussed in E20.7(1) that together the position and momentum operators form an irreducible set.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 21

Spectral Theory of Unitary Operators on $\vec{\mathcal{H}}$

21.1 Spectral Decomposition of Unitary Operators

Theorem 13.4.1(2) on the eigenvectors and the spectral decomposition of unitary operators on $\vec{\boldsymbol{V}}^N$ needs to be modified to apply to an infinite-dimensional Hilbert space $\vec{\mathcal{H}}$. This is because unitary operators on $\vec{\mathcal{H}}$ may have a continuous spectrum. This is similar to the corresponding situation for selfadjoint operators. As with selfadjoint operators the problem is solved by the introduction of spectral functions for unitary operators.

Theorem 21.1(1) Spectral Theorem *To every unitary operator $\widehat{U}$ on a Hilbert space $\vec{\mathcal{H}}$ there corresponds a unique spectral function, known as the spectral function of $\widehat{U}$ and denoted by $\widehat{F}^{\hat{U}}(\tau)$, such that*

$$\langle \vec{\psi} \mid \widehat{U}\vec{\phi} \rangle = \int_{-\infty}^{\infty} e^{i\tau}\, d_\tau \langle \vec{\psi} \mid \widehat{F}^{\hat{U}}(\tau)\vec{\phi} \rangle \qquad \forall\, \vec{\psi},\ \vec{\phi} \in \vec{\mathcal{H}}. \tag{21.1}$$

Following Eq. (20.13) we write

$$\widehat{U} = \int_{-\infty}^{\infty} e^{i\tau}\, d_\tau \widehat{F}^{\hat{U}}(\tau). \tag{21.2}$$

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

The integral expression is known as the **spectral decomposition** of the unitary operator $\widehat{U}$.[1]

The relation between a selfadjoint operator $\widehat{A}$ and a unitary operator $\widehat{U}$ in Eq. (13.80) of Theorem 13.4.3(1) remains valid here. To see how this can arise let $\widehat{F}^{\hat{A}}(\tau)$ be the spectral function of a selfadjoint operator $\widehat{A}$. Following Eq. (20.51) we can construct a complex exponential function of $\widehat{A}$ by[2]

$$e^{i\widehat{A}} = \int_{-\infty}^{\infty} e^{i\tau}\, d_\tau \widehat{F}^{\hat{A}}(\tau). \tag{21.3}$$

This defines a unitary operator $\widehat{U}$ with the above integral as its spectral decomposition. The operators $\widehat{A}$ and $\widehat{U}$ share the same spectral function, i.e., $\widehat{F}^{\hat{U}}(\tau) = \widehat{F}^{\hat{A}}(\tau)$.

In the special case when $\widehat{A}$ possesses a discrete set of eigenvalues $\{a_\ell\}$ with corresponding complete orthonormal set of eigenvectors $\vec{\varphi}_\ell$ the integral in Eq. (21.3) reduces to a sum. In other words, the spectral decomposition of the unitary operator $\widehat{U}$ defined by the integral reduces to

$$\widehat{U} = e^{i\widehat{A}} = \sum_{\ell=1}^{\infty} e^{ia_\ell}\, \widehat{P}_{\vec{\varphi}_\ell} = \sum_{\ell=1}^{\infty} e^{ia_\ell}\, |\vec{\varphi}_\ell\rangle\langle\vec{\varphi}_\ell|. \tag{21.4}$$

This is a generalisation of the corresponding decomposition expressed in Eq. (13.69). An eigenvector $\vec{\varphi}_\ell$ of $\widehat{A}$ is also an eigenvector of the unitary operator $\widehat{U}$ corresponding to eigenvalue $\exp(ia_\ell)$, i.e., we have

$$\widehat{U}\,\vec{\varphi}_\ell = \left(\sum_{\ell'=1}^{\infty} e^{ia_{\ell'}}\, \widehat{P}_{\vec{\varphi}_{\ell'}}\right)\vec{\varphi}_\ell = e^{ia_\ell}\,\vec{\varphi}_\ell. \tag{21.5}$$

Due to Eq. (21.13) in Stone's theorem and subsequent applications Eq. (21.4) can be more usefully written in the form

$$\widehat{U} = e^{-\frac{i}{\hbar}\widehat{H}t} = \sum_{\ell=1}^{\infty} e^{-\frac{i}{\hbar}E_\ell t}\, \widehat{P}_{\vec{\varphi}_\ell} = \sum_{\ell=1}^{\infty} e^{-\frac{i}{\hbar}E_\ell t}\, |\vec{\varphi}_\ell\rangle\langle\vec{\varphi}_\ell|. \tag{21.6}$$

[1] Roman Vol. 2 p. 657. Prugovečki pp. 241–242. Gallone pp. 469–470.
[2] Jordan pp. 51–52. Roman Vol. 2 pp. 660–661.

Here $\dot{\imath} = i/\hbar$ and $\widehat{H}$ is a selfadjoint operator and $t \in I\!R$.

Some of the properties of unitary operators discussed in §13.4.3 remain valid, albeit with some restrictions:

P21.1(1) The expansion of an exponential function in Eq. (13.89) remains true if $\widehat{A}$ is bounded. If $\widehat{A}$ is unbounded the expansion becomes problematic since we would encounter problems of the domain of $\widehat{A}^n$ as n goes to infinity.[3]

P21.1(2) Let $\widehat{A}$ and $\widehat{B}$ be two bounded commuting selfadjoint operators in $\vec{\mathcal{H}}$. Then we have[4]

$$e^{i\hat{A}}e^{i\hat{B}} = e^{i(\hat{A}+\hat{B})} = e^{i\hat{B}}e^{i\hat{A}}. \tag{21.7}$$

It follows that[5]

$$e^{i\hat{A}}e^{-i\hat{A}} = e^{-i\hat{A}}e^{i\hat{A}} = e^{\widehat{0}} = \widehat{I\!I}. \tag{21.8}$$

This is a confirmation of Eqs. (20.49) and (20.49), and the unitary nature of $e^{i\hat{A}}$.

P21.1(3) The above result does not apply to two non-commuting operators. Take the examples of the position and momentum operators $\widehat{x}$ and $\widehat{p}$ in $\vec{L}^2(I\!R)$.[6] They generate following unitary operators:[7]

$$e^{-\dot{\imath}a\hat{x}} \quad \text{and} \quad e^{-\dot{\imath}b\hat{p}}, \quad a, b \in I\!R. \tag{21.9}$$

These two unitary operators do not satisfy Eq. (21.7). Instead they satisfy the following Weyl relation[8]

$$e^{\dot{\imath}a\hat{x}}e^{-\dot{\imath}b\hat{p}} = e^{-\dot{\imath}ab}e^{-i\hat{p}}e^{i\hat{x}}. \tag{21.10}$$

---

[3] Kato pp. 478–484. Prugovečki pp. 300–301. Reed and Simon pp. 264–265.

[4] Prugovečki pp. 339–340. This may not be true for unbounded operators (*see* Blank, Exner and Havliček pp. 194, 202. Riesz and Nagy p. 397).

[5] The operator $e^{\widehat{0}}$ can be understood as the exponential function of $\widehat{A}$ with zero exponent. Then Eq. (21.3) will lead to the identity operator. It can also be regarded as the operator function of $\widehat{A}$ defined by the constant function $e^0 = 1 = f_c(\tau)$. This again results in the identity operator.

[6] Their commutation relation is given by Eq. (13.74).

[7] Recall the notation $\dot{\imath} = i/\hbar$.

[8] Prugovečki p. 288, pp. 333–342. The operator $e^{\dot{\imath}a\hat{x}}$ can be compared with a similar operator in Eq. (27.79) in Chapter 28. Roman p. 642. Weyl (1885–1955) was a German mathematician.

P21.1(4) It is interesting to note that when acting on a vector $\vec{\phi} \in \vec{L}^2(I\!R)$ we have[9]

$$e^{-\imath b\hat{p}}\vec{\phi} := \phi(x-b). \tag{21.11}$$

The result shows that the operator has the effect of translating the function $\phi(x)$ to the right along the x-axis by b, and it is hence known as a **translation operator**. The inverse of the unitary operator will translate the function to the left, i.e.,

$$e^{\imath b\hat{p}}\vec{\phi} := \phi(x+b). \tag{21.12}$$

21.2 Stone's Theorem

Theorem 13.4.3(2) of Stone on one-parameter groups of unitary operators applies in Hilbert spaces, subject to addition conditions due to possible unboundedness of the generator of the group.[10]

Theorem 21.2(1) Stone's Theorem *Let $\widehat{U}(t)$ be a continuous one-parameter group of unitary operators.*[11] *Then there is a unique selfadjoint operator $\widehat{H}$, known as the generator of the group, such that*

$$\widehat{U}(t) = e^{-\imath t\hat{H}}, \tag{21.13}$$

and

$$\widehat{H}\,\vec{\phi} = i\hbar \lim_{t\to 0} \frac{\widehat{U}(t) - \widehat{I\!I}}{t}\,\vec{\phi}, \quad \forall \vec{\phi} \in \mathcal{D}(\widehat{H}). \tag{21.14}$$

We can express the above limit formally as

$$\widehat{H} = i\hbar \left(\frac{d\widehat{U}(t)}{dt}\right)_{t=0} \quad \text{or} \quad i\hbar \frac{d\widehat{U}(t)}{dt} = \widehat{H}\,\widehat{U}(t). \tag{21.15}$$

[9] Baym pp. 152–153. Fano pp. 286–287. Prugovečki p. 347. Roman Vol. 2 p. 556 for a diagram of translating a function.

[10] For direct physical applications in Chapter 29 the generator is denoted by $\widehat{H}$ instead of $\widehat{A}$ in Theorem 13.4.3(2).

[11] Prugovečki pp. 288, 335. Roman Vol. 2 pp. 662–665. Blank, Exner and Havlíček pp. 317–320. Gallone p. 504. The definition of a continuous one-parameter group of unitary operators given by Definition 13.4.3(1) remains valid in a Hilbert space (*see* Jordan p. 52). The constant $\hbar$ is inserted for physical applications.

The unitary operator $\widehat{U}(t)$ leaves $\mathcal{D}(\widehat{H})$ invariant,[12] i.e.,

$$\vec{\phi}(0) \in \mathcal{D}(\widehat{H}) \quad \Rightarrow \quad \widehat{U}(t)\vec{\phi}(0) \in \mathcal{D}(\widehat{H}). \tag{21.16}$$

Thus, for $\vec{\phi}(0) \in \mathcal{D}(\widehat{H})$ and $\vec{\phi}(t) = \widehat{U}(t)\vec{\phi}(0)$ we have,

$$\frac{d\vec{\phi}(t)}{dt} = \left(\frac{d\widehat{U}(t)\vec{\phi}(0)}{dt}\right) = \frac{d\widehat{U}(t)}{dt}\vec{\phi}(0) = \frac{1}{i\hbar}\widehat{H}\vec{\phi}(t). \tag{21.17}$$

In a more familiar form we have

$$i\hbar\frac{d\vec{\phi}(t)}{dt} = \widehat{H}\,\vec{\phi}(t). \tag{21.18}$$

Stone's theorem and the above equation are employed in Chapter 29 to formulate the time evolution of quantum systems.

Exercises and Problems

Q21(1) Let $\widehat{U}$ be a unitary operator in the Hilbert space $\vec{\mathcal{H}}$ with a spectral decomposition given by Eq. (21.4). Let $\vec{\phi}$ be an arbitrary vector in $\mathcal{H}$. Show that

$$\widehat{U}\,\vec{\phi} = \sum_{\ell} c_\ell\, e^{ia_\ell}\,\vec{\varphi}_\ell, \quad c_\ell = \langle\vec{\varphi}_\ell \mid \vec{\phi}\rangle. \tag{21.19}$$

Q21(2) Equation (10.27), which is a time dependent Schrödinger equation, can be written in the form of Eq. (21.18) which follows from Stone's theorem as a vector equation in the Hilbert space $\vec{L}^2(I\!R)$, i.e.,

$$i\hbar\frac{d\vec{\phi}(t)}{dt} = \widehat{H}\,\vec{\phi}(t), \tag{21.20}$$

where $\vec{\phi}(t)$ is in the domain of $\widehat{H}$. Explain why the norm of vector $\vec{\phi}(t)$ is preserved in time, i.e., $||\vec{\phi}(t_1)|| = ||\vec{\phi}(t_2)||$ for any t_1 and t_2.

[12] Prugovečki p. 291. Weidmann p. 225. Since $\widehat{U}(t)$ and $\widehat{A}$ commute Eq. (17.87) implies the result.

Q21(3) Consider the following unitary transformations of an operator $\widehat{A}$:

$$\widehat{A}(t) = \widehat{U}^{\dagger}(t)\widehat{A}\widehat{U}(t), \tag{21.21}$$

where $\widehat{U}(t) = e^{-\frac{i}{\hbar}t\widehat{H}}$ is a one-parameter family of unitary operators in Eq. (21.13).[13] Show that[14]

$$i\hbar\frac{d\widehat{A}(t)}{dt} = [\widehat{A}(t),\ \widehat{H}\ ]. \tag{21.22}$$

[13] The operators $\widehat{A}$ and $\widehat{H}$ are independent of t and $\widehat{A}(t=0) = \widehat{A}$.
[14] *See* Eq. (29.19) for the physical relevance of this result.

Chapter 22

Probability, Selfadjoint Operators and Unit Vectors

22.1 Generating Probability Distributions on $\vec{\mathcal{H}}$

Theorem 15.5(1) applies to infinite-dimensional Hilbert spaces. In view of the crucial importance of this we shall iterate it below.

Theorem 22.1(1) *A selfadjoint operator $\widehat{A}$ together with a unit vector $\vec{\phi}$ in a Hilbert space generates a unique probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\phi}, \tau)$ and a probability measure $\mathcal{M}^{\hat{A}}(\vec{\phi}, \Lambda)$ given by the quadratic forms $\mathcal{Q}\big(\widehat{F}^{\hat{A}}(\tau), \vec{\phi}\,\big)$ and $\mathcal{Q}\big(\widehat{M}^{\hat{A}}(\Lambda), \vec{\phi}\,\big)$ generated by its spectral function $\widehat{F}^{\hat{A}}(\tau)$ and spectral measure $\widehat{M}^{\hat{A}}(\Lambda)$, i.e.,*

$$\boxed{\mathcal{F}^{\hat{A}}(\vec{\phi}, \tau) := \mathcal{Q}\big(\widehat{F}^{\hat{A}}(\tau), \vec{\phi}\,\big) = \langle \vec{\phi} \mid \widehat{F}^{\hat{A}}(\tau)\vec{\phi}\,\rangle,} \tag{22.1}$$

and

$$\boxed{\mathcal{M}^{\hat{A}}(\vec{\phi}, \Lambda) := \mathcal{Q}\big(\widehat{M}^{\hat{A}}(\Lambda), \vec{\phi}\,\big) = \langle \vec{\phi} \mid \widehat{M}^{\hat{A}}(\Lambda)\vec{\phi}\,\rangle.} \tag{22.2}$$

Two unit vectors which differ only by a *phase factor* would generate the same probability distribution function.[1]

[1] A phase factor is a multiplicative constant of magnitude 1, i.e., $e^{i\theta}$, $\theta \in \mathbb{R}$.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

Corollary 15.5(1) also applies, albeit subject to a domain restriction. In other words, a probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\phi}, \tau)$ leads to a finite expectation value $\mathcal{E}(\widehat{A}, \vec{\phi}\,)$ only if $\vec{\phi}$ is in the domain of $\widehat{A}$. We have, for $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$,

$$\mathcal{E}(\widehat{A}, \vec{\phi}\,) := \int_{-\infty}^{\infty} \tau \, d_\tau \mathcal{F}^{\hat{A}}(\vec{\phi}, \tau) = \langle \vec{\phi} \mid \widehat{A}\,\vec{\phi}\,\rangle. \tag{22.3}$$

As in Eqs. (10.8) and (10.9) this expectation value is seen to be equal to the quadratic form $\mathcal{Q}(\widehat{A}, \vec{\phi}\,)$.[2] Following Eq. (10.10) the uncertainty arising from the probability distribution function can be similarly expressed, i.e., we have

$$\Delta(\widehat{A}, \vec{\phi}\,) = \sqrt{\langle \vec{\phi} \mid \widehat{A}^2 \vec{\phi}\,\rangle - \langle \vec{\phi} \mid \widehat{A}\vec{\phi}\,\rangle^2}. \tag{22.4}$$

If the probability distribution function is absolutely continuous we can introduce a probability density function by

$$w^{\hat{A}}(\vec{\phi}, \tau) = \frac{d}{d\tau} \mathcal{F}^{\hat{A}}(\vec{\phi}, \tau). \tag{22.5}$$

The expectation value can be written directly in terms of the probability density function as

$$\mathcal{E}(\widehat{A}, \vec{\phi}\,) = \int_{-\infty}^{\infty} \tau \, d_\tau \mathcal{F}^{\hat{A}}(\vec{\phi}, \tau) = \int_{-\infty}^{\infty} \tau \, w^{\hat{A}}(\vec{\phi}, \tau)\, d\tau, \tag{22.6}$$

for all $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$. An arbitrary symmetric operator does not possess a spectral function and the spectral theorem does not apply to such an operator. Hence the relation between selfadjoint operators and probability distributions presented above does not apply to symmetric operators. This is the reason that the probabilistic nature of quantum mechanics is formulated in terms of selfadjoint operators, not symmetric operators.[3]

Note that the expression for the uncertainty in Eq. (22.4) requires the $\vec{\phi}$ to be in the domain of $\widehat{A}^2$. This means that for $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$

[2] Recall that $\mathcal{Q}(\widehat{A}, \vec{\phi}\,) = \langle \vec{\phi} \mid \widehat{A}\,\vec{\phi}\,\rangle$.

[3] It is possible to generalise orthodox quantum theory to include certain symmetric operators. Wan pp. 395–428.

we have a finite expectation value but we may not have a finite uncertainty since $\vec{\phi}$ may not be in the domain of $\widehat{A}^2$. For a selfadjoint operator $\widehat{A}$ and $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A}^2)$ we have

$$\langle \vec{\phi} \mid \widehat{A}^2 \vec{\phi} \rangle = \langle \widehat{A}\vec{\phi} \mid \widehat{A}\vec{\phi} \rangle = ||\widehat{A}\vec{\phi}||^2, \tag{22.7}$$

and

$$\Delta(\widehat{A}, \vec{\phi}) = \sqrt{||\widehat{A}\vec{\phi}||^2 - \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle^2}. \tag{22.8}$$

This expression for $\Delta(\widehat{A}, \vec{\phi})$ can be extended to $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$ so that a finite uncertainty would exist for all $\vec{\phi} \in \vec{\mathcal{D}}(\widehat{A})$ if we choose to define uncertainty by Eq. (22.8).[4]

Applications of Theorem 22.1(1) to operators with discrete and continuous spectra are given in the following sections.

22.2 Operators with a Discrete Spectrum

Many selfadjoint operators important to physical applications, e.g., number operators and some first order and second order differential operators, possess a discrete spectrum. For a selfadjoint operator with a discrete spectrum the spectral function reduces to a sum in the form of Eq. (20.20). The probability distribution arising from Theorem 22.1(1) can be given by a probability mass function as seen in the following theorem.

Theorem 22.2(1) *A selfadjoint operator $\widehat{A}$ with a discrete spectrum $sp_d(\widehat{A}) = \{a_m\}$ together with a unit vector $\vec{\phi}$ in a Hilbert space generates a unique probability mass function $\wp^{\hat{A}}(\vec{\phi}, a_m)$ on $sp_d(\widehat{A})$ by the quadratic form $\mathcal{Q}\big(\widehat{P}^{\hat{A}}(a_m), \vec{\phi}\,\big)$, i.e.,*

$$\boxed{\wp^{\hat{A}}(\vec{\phi}, a_m) := \mathcal{Q}\big(\widehat{P}^{\hat{A}}(a_m), \vec{\phi}\,\big) = \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_m)\vec{\phi} \rangle.} \tag{22.9}$$

For a nondegenerate eigenvalue a_ℓ corresponding to the eigenvector $\vec{\varphi}_\ell$ we have

$$\wp^{\hat{A}}(\vec{\phi}, a_\ell) := \langle \vec{\phi} \mid \widehat{P}_{\vec{\varphi}_\ell}\vec{\phi} \rangle. \tag{22.10}$$

[4] Wan. p. 183.

The corresponding probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\phi}, \tau)$ defined by Eq. (22.1) is piecewise-constant since the spectral function is piecewise-constant. Explicitly $\mathcal{F}^{\hat{A}}(\vec{\phi}, \tau)$ is given by

$$\begin{cases} 0, & \tau < a_1 \\ \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_1)\vec{\phi} \rangle, & a_1 \leq \tau < a_2 \\ \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_1)\vec{\phi} \rangle + \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_2)\vec{\phi} \rangle, & a_2 \leq \tau < a_3 \\ \cdots & \cdots \\ \sum_{\ell=1}^{m} \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_\ell)\vec{\phi} \rangle, & a_m \leq \tau < a_{m+1} \\ \cdots & \cdots \end{cases} . \tag{22.11}$$

This is clearly an extension of Eq. (14.58) in $\vec{\mathbb{V}}^N$. For the expectation value we have, for $\vec{\phi} \in \mathcal{D}(\widehat{A})$,

$$\mathcal{E}(\widehat{A}, \vec{\phi}) := \sum_{\ell} \wp^{\hat{A}}(\vec{\phi}, a_\ell)\, a_\ell = \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle. \tag{22.12}$$

Equation (22.8) for the uncertainty can be applied here.

22.3 Operators with a Continuous Spectrum

22.3.1 Position Operator in $\vec{L}^2(I\!R)$

We can generate a probability distribution function $\mathcal{F}^{\hat{x}}(\vec{\phi}, \tau)$ in terms of the spectral function $\widehat{F}^{\hat{x}}(\tau)$ of the position operator $\widehat{x}$ for every vector $\vec{\phi} \in \vec{L}^2(I\!R)$ in accordance with Eq. (22.1) by

$$\mathcal{F}^{\hat{x}}(\vec{\phi}, \tau) := \langle \vec{\phi} \mid \widehat{F}^{\hat{x}}(\tau)\vec{\phi} \rangle = \langle \vec{\phi} \mid \widehat{\chi}_{(-\infty,\tau]}\vec{\phi} \rangle. \tag{22.13}$$

Explicitly we have

$$\begin{aligned} \mathcal{F}^{\hat{x}}(\vec{\phi}, \tau) : &= \int_{-\infty}^{\infty} \phi^*(x)\ \chi_{(-\infty,\tau]}(x)\phi(x)\, dx \\ &= \int_{-\infty}^{\tau} |\phi(x)|^2\, dx. \end{aligned} \tag{22.14}$$

This probability distribution function is absolutely continuous. The corresponding probability density function given by

$$w^{\hat{x}}(\vec{\phi}, \tau) := \frac{d}{d\tau} \int_{-\infty}^{\tau} |\phi(x)|^2\, dx = |\phi(\tau)|^2. \tag{22.15}$$

These results make it possible for us to pursue our previous model theories in Chapter 14 for observables with a continuous spectrum. We know that the position of a quantum particle does not have a definite value in any state and that a state can only determine the probability distribution of the particle's position values. Then following the discussion in §14.2 we may represent the position of a particle by the operator $\widehat{x}$ and a state of the particle by a unit vector $\vec{\phi} \in \vec{L}^2(I\!R)$. The probability distribution of position values will then be determined by the probability distribution function $\mathcal{F}^{\hat{x}}(\vec{\phi}, \tau)$ with the position expectation value for any $\vec{\phi} \in \mathcal{D}(\widehat{x})$ given by

$$\begin{aligned}\mathcal{E}(\widehat{x}, \vec{\phi}\,) :&= \int_{-\infty}^{\infty} \tau\, d_\tau \mathcal{F}^{\hat{x}}(\vec{\phi}, \tau) = \int_{-\infty}^{\infty} \tau w^{\hat{x}}(\vec{\phi}, \tau) d\tau \\ &= \int_{-\infty}^{\infty} \tau |\phi(\tau)|^2 \, d\tau = \langle \vec{\phi} \mid \widehat{x}\vec{\phi}\,\rangle \end{aligned} \tag{22.16}$$

with uncertainty

$$\Delta(\widehat{x}, \vec{\phi}\,) := \sqrt{||\widehat{x}\vec{\phi}\,||^2 - \langle \vec{\phi} \mid \widehat{x}\vec{\phi}\,\rangle^2}. \tag{22.17}$$

22.3.2 Momentum Operator in $\vec{L}^2(I\!R)$

In the momentum representation the momentum appears as a multiplication operator in the same way that the position acts as a multiplication operator in the coordinate representation. It follows that we can treat the momentum in the momentum representation in the same way as position in the coordinate representation. Using the expression in Eq. (20.29) for the momentum spectral function $\widehat{\underset{\sim}{F}}^{\hat{p}}(\tau)$ we obtain the momentum probability distribution function $\mathcal{F}^{\hat{p}}(\vec{\underset{\sim}{\phi}}, \tau)$ in the momentum representation

$$\begin{aligned}\mathcal{F}^{\hat{p}}(\vec{\underset{\sim}{\phi}}, \tau) &:= \int_{-\infty}^{\infty} \underset{\sim}{\phi}^*(p)\, \underset{\sim}{\chi}_{(-\infty,\tau]}(p) \underset{\sim}{\phi}(p)\, dp \\ &= \int_{-\infty}^{\tau} |\underset{\sim}{\phi}(p)|^2 \, dp. \end{aligned} \tag{22.18}$$

This probability distribution is absolutely continuous. The momentum probability density function is

$$w^{\hat{p}}(\vec{\phi}, \tau) := \frac{d}{d\tau}\mathcal{F}^{\hat{p}}(\vec{\phi}, \tau) = \frac{d}{d\tau}\int_{-\infty}^{\tau} |\phi(p)|^2\, dp = |\phi(\tau)|^2. \tag{22.19}$$

The resulting expectation value for any $\vec{\phi} \in \mathcal{D}(\widehat{p})$ is given by

$$\mathcal{E}(\widehat{p}, \vec{\phi}) := \int_{-\infty}^{\infty} \tau d_\tau \widetilde{\mathcal{F}}^{\hat{p}}(\vec{\phi}, \tau) = \int_{-\infty}^{\infty} \tau w^{\hat{p}}(\vec{\phi}, \tau) d\tau = \int_{-\infty}^{\infty} \tau |\phi(\tau)|^2\, d\tau = \langle \vec{\phi} \mid \widehat{p}\vec{\phi} \rangle, \tag{22.20}$$

with uncertainty

$$\Delta(\widehat{p}, \vec{\phi}) = \sqrt{||\widehat{p}\vec{\phi}||^2 - \langle \vec{\phi} \mid \widehat{p}\vec{\phi} \rangle^2}\,. \tag{22.21}$$

Alternatively we can also work in the coordinate representation. Using the spectral function $\widehat{F}^{\hat{p}}(\tau)$ in Eq. (20.32) we can evaluate the probability distribution function in the coordinate representation. Denoting he probability distribution function in the coordinate representation by $\mathcal{F}^{\hat{p}}(\vec{\phi}, \tau)$ we have

$$\begin{aligned}
\mathcal{F}^{\hat{p}}(\vec{\phi}, \tau) &= \int_{-\infty}^{\infty} \phi^*(x) \left(\int_{-\infty}^{\tau} f_p(x)\phi(p)\, dp \right) dx \\
&= \int_{-\infty}^{\tau} \left(\int_{-\infty}^{\infty} \phi^*(x) f_p(x)\, dx \right) \phi(p)\, dp \\
&= \int_{-\infty}^{\tau} \left(\int_{-\infty}^{\infty} \phi(x) f_p^*(x)\, dx \right)^* \phi(p)\, dp \\
&= \int_{-\infty}^{\tau} \phi^*(p)\phi(p)\, dp = \int_{-\infty}^{\tau} |\phi(p)|^2\, dp.
\end{aligned} \tag{22.22}$$

This agrees with previous result, i.e.,

$$\mathcal{F}^{\hat{p}}(\vec{\phi}, \tau) = \mathcal{F}^{\hat{p}}(\vec{\phi}, \tau). \tag{22.23}$$

The probability density function is the same as before, i.e.,

$$w^{\hat{p}}(\vec{\phi}, \tau) := \frac{d}{d\tau}\mathcal{F}^{\hat{p}}(\vec{\phi}, \tau) = |\phi(\tau)|^2 = w^{\hat{p}}(\vec{\phi}, \tau). \tag{22.24}$$

The corresponding expectation value is given by

$$\mathcal{E}(\widehat{p}, \vec{\phi}) = \langle \vec{\phi} \mid \widehat{p}\,\vec{\phi} \rangle \tag{22.25}$$

with the uncertainty

$$\Delta(\widehat{p}, \vec{\phi}) = \sqrt{||\widehat{p}\,\vec{\phi}||^2 - \langle \vec{\phi} \mid \widehat{p}\,\vec{\phi} \rangle^2}. \tag{22.26}$$

These values are the same as before, i.e.,[5]

$$\mathcal{E}(\underset{\sim}{\widehat{p}}, \underset{\sim}{\vec{\phi}}) = \mathcal{E}(\widehat{p}, \vec{\phi}), \quad \Delta(\underset{\sim}{\widehat{p}}, \underset{\sim}{\vec{\phi}}) = \Delta(\widehat{p}, \vec{\phi}). \tag{22.27}$$

All these results are as expected since probability distributions and expectation values are physically measurable quantities which are not dependent on any particular mathematical description.

Finally we should mention that the uncertainties of the position and momentum associated with an appropriate unit vector $\vec{\phi}$ in $\vec{L}^2(I\!R)$ are related by the following uncertainty relation[6]

$$\Delta(\widehat{x}, \vec{\phi})\Delta(\widehat{p}, \vec{\phi}) \geq \frac{1}{2}\hbar, \tag{22.28}$$

Exercises and Problems

Q22(1) The commutation relation between the selfadjoint position and momentum operators $\widehat{x} = x$ and $\widehat{p} = -i\hbar\, d/dx$ in the Hilbert space $\vec{L}^2(I\!R)$ is often written as

$$[\widehat{x},\ \widehat{p}\,] = i\hbar. \tag{22.29}$$

As pointed out in the discussion in §17.7 this relation should be expressed as an inequality, i.e.,

$$[\widehat{x},\ \widehat{p}\,] \subset i\hbar\,\widehat{I\!I}, \tag{22.30}$$

[5] These results confirm that fact that Fourier transform as a unitary transform preserves quadratic forms.

[6] Similar to the commutation relation in Eq. (22.31) this uncertainty relation does not apply to every unit vector in $\vec{L}^2(I\!R)$. Various proofs of the uncertainty relations using the Schwarz inequality are available in many textbooks on quantum mechanics.

or

$$[\widehat{x},\ \widehat{p}\ ]\,\vec{\phi} = i\hbar\,\vec{\phi} \tag{22.31}$$

for an appropriate set of vectors $\vec{\phi}$ in $\vec{L}^2(I\!R)$. What are the conditions $\vec{\phi}$ must satisfy in order for the equality to hold?

Q22(2) Consider a particle confined in an infinite potential well of width $[0,\ L]$. All its wave functions $\phi(x)$ must vanish outside the well, i.e., $\phi(x) = 0\ \ \forall x \notin [0,\ L]$. Hence the state space of the trapped particle is taken to be $\vec{L}^2(\Lambda)$, $\Lambda = [0,\ L]$ rather than $\vec{L}^2(I\!R)$.

(a) Taking the operator $\widehat{x}(\Lambda)$ in Eq. (17.22) as the position operator show that the uncertainty in position cannot be bigger than the width of the well.

(b) Taking the operator $\widehat{p}_{\lambda=0}(\Lambda)$ in Eq. (17.36) as the momentum operator show that the momentum uncertainty

$$\Delta\big(\widehat{p}_{\lambda=0}(\Lambda),\, \vec{\varphi}_{\lambda=0,n}(\Lambda)\big)$$

is zero. Here $\vec{\varphi}_{\lambda=0,n}(\Lambda)$ are eigenvectors of $\widehat{p}_{\lambda=0}(\Lambda)$ in Eq. (19.34).

(c) Bearing in mind the above results investigate whether or not an uncertainty relation similar to that shown in Eq. (22.28) remains valid for eigenvectors of $\widehat{p}_{\lambda=0}(\Lambda)$.[7]

[7] Fano pp. 407–408 for a similar problem with the uncertainty relation in the Hilbert space $\vec{L}^2(\mathcal{C}_a)$. *See* also §28.3.3 on a particle in circular motion.

Chapter 23

Physics of Unitary Transformations

A physical theory often admits different mathematical descriptions. Classical mechanics has three familiar formulations: the traditional Newton's formulation, the Lagrange's and the Hamilton's formulations.[1] It is easy to see that these are physically equivalent. Quantum mechanics also admits many different formulations. The physical equivalence of different formulations in quantum mechanics are often not obvious, since the mathematics involved is rather abstract. This chapter discusses how different mathematical descriptions of a probability theory based on a Hilbert space can be physically equivalent.

Following the model theory for spin in Chapter 14 we can establish a probabilistic theory for quantum systems satisfying properties QMP5.3(1) to QMP5.3(4) stated in §5.3 as follows:

(1) Choose an appropriate Hilbert space $\vec{\mathcal{H}}$ as the state space of the system in that a quantum state ϕ^s is described by a unit vector $\vec{\phi}$ in $\vec{\mathcal{H}}$.

[1] *See* §27.1 for a summary of the Lagrangian and the Hamilton's formulations. Hamilton (1805–1865) was an Irish physicist. Lagrange (1736–1813) was a French mathematician born in Italy.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

(2) Choose appropriate selfadjoint operators to associate with the observables of the system in that the measurable values of an observable A are identified with elements of the spectrum of its associated selfadjoint operator $\widehat{A}$.

(3) The probability distribution of measurable values of observable A in state ϕ^s is given by the probability distribution function $\mathcal{F}^{\hat{A}}(\vec{\phi}, \tau)$ in Theorem 22.1(1).

The above prescription does not single out a unique Hilbert space nor a unique set of unit vectors and operators for the description of a given quantum system. This non-uniqueness does not matter since operators and unit vectors themselves are not physically measurable quantities. What is required is for physically measurable values to be independent of any particular choice of those abstract mathematical quantities. Let us make this clear by setting out the requirements for two different mathematical descriptions of a quantum system to be physically equivalent.

Physical equivalence

Two different mathematical descriptions of a quantum system are physically equivalent if and only if they lead to the same measurable properties for the system. In other words the predicted measurable values together with their probability distribution for every observable in every state must be the same in the two descriptions.

Now consider a description of the system in terms of a set of selfadjoint operators $\widehat{A}$ and unit vectors $\vec{\phi}$ and another description in terms of a set of selfadjoint operators $\widehat{A}'$ and unit vectors $\vec{\phi}'$. The physics is contained in the various numerical values, e.g., eigenvalues, probabilities, expectation values. Everyone of these values is expressible in terms of the quadratic form generated by a selfadjoint operator in conjunction with a unit vector $\vec{\phi}$, i.e., $\mathcal{Q}(\widehat{A}, \vec{\phi})$ for the first description and $\mathcal{Q}(\widehat{A}', \vec{\phi}')$ for the second description. If the operators and unit vectors of two descriptions are related by a simultaneous unitary transformation, i.e., there is a unitary operator $\widehat{U}$ such that

$$\vec{\phi}' = \widehat{U}\,\vec{\phi} \quad \text{and} \quad \widehat{A}' = \widehat{U}\,\widehat{A}\widehat{U}^{\dagger}, \tag{23.1}$$

for every unit vector $\vec{\phi}$ and every selfadjoint operator $\widehat{A}$,[2] then the values of the quadratic forms $\mathcal{Q}(\widehat{A}, \vec{\phi}\,)$ and $\mathcal{Q}(\widehat{A}', \vec{\phi}\,')$ would be the same, i.e.,

$$\mathcal{Q}(\widehat{A}, \vec{\phi}\,) = \mathcal{Q}(\widehat{A}', \vec{\phi}\,') \quad \Leftrightarrow \quad \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle = \langle \vec{\phi}\,' \mid \widehat{A}'\vec{\phi}\,' \rangle. \tag{23.2}$$

This leads us to the following requirement.

Mathematical requirement for physical equivalence:

> *Two different mathematical descriptions of a quantum system are physically equivalent if the unit vectors and selfadjoint operators for the representation of states and observables in the two descriptions are related by a simultaneous unitary transformation.*

The descriptions of position and momentum in the coordinate momentum representations in §22.3.1 and §22.3.2 are examples. Many more examples will be presented in Chapters 29 and 35 which will illustrate the physical equivalence of different mathematical descriptions and formulations of quantum theory.

Exercises and Problems

Q23(1) Explain why a coordinate representation space and its corresponding momentum representation space introduced in §18.4.2 give rise to two mathematically different descriptions of the position and momentum which are physically equivalent.

[2] Theorem 15.3(2) also tells us that the spectral functions and the spectral measures of $\widehat{A}$ and $\widehat{A}'$ are also related by a unitary transformation generated by $\widehat{U}$.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 24

Direct Sums and Tensor Products of Hilbert Spaces and Operators

Given two Hilbert spaces we can construct further Hilbert spaces. There are two physically important constructions leading to a direct sum space and a tensor product space.

24.1 Direct Sums of Hilbert Spaces and Operators

24.1.1 Direct Sums of Hilbert Spaces

Let $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$ be two Hilbert spaces. The set of ordered pairs

$$\left\{\{\vec{\phi}^{(1)}, \vec{\phi}^{(2)}\} : \vec{\phi}^{(1)} \in \vec{\mathcal{H}}^{(1)},\ \vec{\phi}^{(2)} \in \vec{\mathcal{H}}^{(2)}\right\} \tag{24.1}$$

can be given a natural Hilbert space structure with algebraic operations and scalar product of any two elements $\{\vec{\phi}^{(1)}, \vec{\phi}^{(2)}\}$ and $\{\vec{\psi}^{(1)}, \vec{\psi}^{(2)}\}$ of the set defined by

$$c\,\{\vec{\phi}^{(1)}, \vec{\phi}^{(2)}\} := \{c\vec{\phi}^{(1)}, c\vec{\phi}^{(2)}\}, \quad c \in \mathbb{C}, \tag{24.2}$$

$$\{\vec{\phi}^{(1)}, \vec{\phi}^{(2)}\} + \{\vec{\psi}^{(1)}, \vec{\psi}^{(2)}\} := \{\vec{\phi}^{(1)} + \vec{\psi}^{(1)}, \vec{\phi}^{(2)} + \vec{\psi}^{(2)}\}, \tag{24.3}$$

$$\langle\{\vec{\phi}^{(1)}, \vec{\phi}^{(2)}\} \,|\, \{\vec{\psi}^{(1)}, \vec{\psi}^{(2)}\}\rangle^{\oplus} := \langle\vec{\phi}^{(1)} \mid \vec{\psi}^{(1)}\rangle^{(1)} + \langle\vec{\phi}^{(2)} \mid \vec{\psi}^{(2)}\rangle^{(2)}. \tag{24.4}$$

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

The set of ordered pair then forms a Hilbert space.[1]

Definition 24.1.1(1) *The* **direct sum** *of* $\vec{\mathcal{H}}^{(1)}$ *and* $\vec{\mathcal{H}}^{(2)}$*, denoted by* $\vec{\mathcal{H}}^{\oplus}$*, is the Hilbert space formed by the set of ordered pairs in* Eq. (24.1) *with the Hilbert space structure defined by* Eqs. (24.2), (24.3) and (24.4)*.*

The notation for direct sums are set out below:

(1) The direct sum is also denoted explicitly by $\vec{\mathcal{H}}^{(1)} \oplus \vec{\mathcal{H}}^{(2)}$ with elements $\{\vec{\phi}^{(1)}, \vec{\phi}^{(2)}\}$ of the direct sum $\vec{\mathcal{H}}^{(1)} \oplus \vec{\mathcal{H}}^{(2)}$ denoted by $\vec{\phi}^{(1)} \oplus \vec{\phi}^{(2)}$ which is also called the direct sum of $\vec{\phi}^{(1)}$ and $\vec{\phi}^{(2)}$. In keeping with the notation of $\vec{\mathcal{H}}^{\oplus}$ an element of the direct sum space is also denoted by $\vec{\phi}^{\oplus}$, i.e., we have

$$\vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{H}}^{(1)} \oplus \vec{\mathcal{H}}^{(2)}, \tag{24.5}$$

$$\vec{\phi}^{\oplus} = \vec{\phi}^{(1)} \oplus \vec{\phi}^{(2)}, \tag{24.6}$$

$$\vec{\psi}^{\oplus} = \vec{\psi}^{(1)} \oplus \vec{\psi}^{(2)}, \tag{24.7}$$

$$\langle \vec{\phi}^{\oplus} \mid \vec{\psi}^{\oplus} \rangle^{\oplus} := \langle \vec{\phi}^{(1)} \mid \vec{\psi}^{(1)} \rangle^{(1)} + \langle \vec{\phi}^{(2)} \mid \vec{\psi}^{(2)} \rangle^{(2)}. \tag{24.8}$$

(2) Vectors $\vec{\phi}^{(1)}$ in $\vec{\mathcal{H}}^{(1)}$ and $\vec{\phi}^{(2)}$ in $\vec{\mathcal{H}}^{(2)}$ correspond to vectors $\vec{\phi}^{(1)\oplus}$ and $\vec{\phi}^{(2)\oplus}$ in $\vec{\mathcal{H}}^{\oplus}$, i.e.,

$$\vec{\phi}^{(1)} \leftrightarrow \vec{\phi}^{(1)\oplus} := \vec{\phi}^{(1)} \oplus \vec{0}^{(2)}, \tag{24.9}$$

$$\vec{\phi}^{(2)} \leftrightarrow \vec{\phi}^{(2)\oplus} := \vec{0}^{(1)} \oplus \vec{\phi}^{(2)}, \tag{24.10}$$

where $\vec{0}^{(1)}$ and $\vec{0}^{(2)}$ are the zero vectors in $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$, respectively. We have

$$\langle \vec{\phi}^{(1)\oplus} \mid \vec{\phi}^{(2)\oplus} \rangle^{\oplus} = 0. \tag{24.11}$$

(3) The constituent spaces $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$ correspond to two subspaces of the direct sum space, e.g., $\vec{\mathcal{H}}^{(1)}$ corresponds to the subspace[2]

$$\vec{\mathcal{H}}^{(1)\oplus} := \left\{ \vec{\phi}^{(1)\oplus} = \vec{\phi}^{(1)} \oplus \vec{0}^{(2)} : \ \vec{\phi}^{(1)} \in \mathcal{H}^{(1)} \right\}. \tag{24.12}$$

[1] The superscript $\oplus$ in Eq. (24.4) indicates the scalar product in $\vec{\mathcal{H}}^{\oplus}$ while the superscripts (1) and (2) signify scalar products in $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$.

[2] Kadison and Ringrose Vol. 1 p. 111, pp. 121–122. The correspondence in Eqs. (24.9) and (24.12) can be made precise in terms of a unitary mapping.

We can extend the definition to construct the direct sum of three or more Hilbert spaces in a straightforward manner.[3] We can also form the direct sum of a Hilbert space with itself, e.g.,

$$\vec{\mathbb{C}}^{\oplus} := \vec{\mathbb{C}} \oplus \vec{\mathbb{C}}. \tag{24.13}$$

The dimension of the direct sum space is equal to the sum of the dimensions of the constituent spaces, e.g., $\vec{\mathbb{C}}^{\oplus}$ is two-dimensional and is identifiable with $\vec{\mathbb{C}}^{2}$.[4]

The concept of direct sum can be used to decompose a given Hilbert space in the following manner:

(1) Let $\vec{\mathcal{S}}$ be a subspace of a Hilbert space $\vec{\mathcal{H}}$ and let $\vec{\mathcal{S}}^{\perp}$ be its orthogonal complement. Vectors in $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^{\perp}$ as subspaces of $\mathcal{H}$ are denoted by $\vec{\phi}_{\vec{\mathcal{S}}}$ and $\vec{\phi}_{\vec{\mathcal{S}}^{\perp}}$ as in Eq. (16.84). These subspaces are Hilbert spaces in their own right so that we can form their direct sums

$$\vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{S}} \oplus \vec{\mathcal{S}}^{\perp}, \quad \vec{\phi}^{\oplus} = \vec{\phi}_{\vec{\mathcal{S}}} \oplus \vec{\phi}_{\vec{\mathcal{S}}^{\perp}}. \tag{24.14}$$

The following correspondence are obvious:

$$\vec{\mathcal{S}} \text{ and } \vec{\mathcal{S}}^{\perp} \quad \leftrightarrow \quad \vec{\mathcal{H}}^{(1)} \text{ and } \vec{\mathcal{H}}^{(2)}, \tag{24.15}$$

$$\vec{\phi}_{\vec{\mathcal{S}}} \in \vec{\mathcal{S}} \subset \vec{\mathcal{H}} \quad \leftrightarrow \quad \vec{\phi}^{\oplus}_{\vec{\mathcal{S}}} = \vec{\phi}_{\vec{\mathcal{S}}} \oplus \vec{0}^{(2)} \in \vec{\mathcal{H}}^{\oplus} \tag{24.16}$$

$$\vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{S}}^{\perp} \subset \vec{\mathcal{H}} \quad \leftrightarrow \quad \vec{\phi}^{\oplus}_{\vec{\mathcal{S}}^{\perp}} = \vec{0}^{(1)} \oplus \vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \mathcal{H}^{\oplus}, \tag{24.17}$$

$$\vec{\phi} = \vec{\phi}_{\vec{\mathcal{S}}} + \vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{H}} \quad \leftrightarrow \quad \vec{\phi}^{\oplus} = \vec{\phi}_{\vec{\mathcal{S}}} \oplus \vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{H}}^{\oplus}. \tag{24.18}$$

The direct sum $\vec{\mathcal{H}}^{\oplus}$ in Eq. (24.14) is identifiable with $\vec{\mathcal{H}}$. We can write $\vec{\mathcal{H}} = \vec{\mathcal{H}}^{\oplus}$, i.e., $\vec{\mathcal{H}} = \vec{\mathcal{S}} \oplus \vec{\mathcal{S}}^{\perp}$.

(2) Since the pair of subspaces $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^{\perp}$ constitutes a complete orthogonal family of subspaces of $\vec{\mathcal{H}}$,[5] we say that $\vec{\mathcal{H}}$ *is decomposable as a direct sum of the complete orthogonal family of subspaces* $\vec{\mathcal{S}}$ *and* $\vec{\mathcal{S}}^{\perp}$ *of* $\vec{\mathcal{H}}$.

---

[3] Roman Vol. 2 pp. 450–455.

[4] The two spaces $\vec{\mathbb{C}}^{\oplus}$ and $\vec{\mathbb{C}}^{2}$ are unitarily related.

[5] Definition 13.2.2(2).

(3) The decomposition can be extended to a general complete orthogonal family of subspaces of $\vec{\mathcal{H}}$. Treating these subspaces as Hilbert spaces with the notation $\vec{\mathcal{S}}^{(n)}$ we can form their direct sum, i.e.,[6]

$$\vec{\mathcal{H}}^{\oplus} := \oplus_n \vec{\mathcal{S}}^{(n)} = \vec{\mathcal{S}}^{(1)} \oplus \vec{\mathcal{S}}^{(2)} \oplus \cdots . \tag{24.19}$$

Elements of the direct sum space are of the form

$$\vec{\phi}^{\,\oplus} := \oplus_n \vec{\phi}^{\,(n)}, \quad \vec{\phi}^{\,(n)} \in \vec{\mathcal{S}}^{\,(n)}. \tag{24.20}$$

Subspaces of $\vec{\mathcal{H}}^{\oplus}$ corresponding to $\vec{\mathcal{S}}^{\,(n)}$ are denoted by $\vec{\mathcal{S}}^{(n)\oplus}$, e.g.,

$$\vec{\mathcal{S}}^{\,(1)\oplus} := \vec{\mathcal{S}}^{\,(1)} \oplus \vec{0}^{(2)} \oplus \vec{0}^{(3)} \oplus \cdots . \tag{24.21}$$

The above direct sum decomposition has the following properties:

P24.1.1(1) The direct sum space $\vec{\mathcal{H}}^{\oplus}$ is identifiable with $\mathcal{H}$. We say that

a Hilbert space $\vec{\mathcal{H}}$ is decomposable as a direct sum of a complete orthogonal family of subspaces $\vec{\mathcal{S}}^{\,(n)}$ of $\vec{\mathcal{H}}$.

P24.1.1(2) The scalar product of $\vec{\phi}^{\,\oplus}$ in Eq. (24.20) and another vector $\vec{\psi}^{\,\oplus}$ in $\mathcal{H}^{\oplus}$ is given by[7]

$$\langle \vec{\phi}^{\,\oplus} \mid \vec{\psi}^{\,\oplus} \rangle^{\oplus} = \sum_n \langle \vec{\phi}^{\,(n)} \mid \vec{\psi}^{\,(n)} \rangle^{(n)}. \tag{24.22}$$

P24.1.1(3) Generalising the notation in Eqs. (24.9) and (24.10), e.g.,

$$\vec{\phi}^{\,(1)\oplus} := \vec{\phi}^{\,(1)} \oplus \vec{0}^{\,(2)} \oplus \vec{0}^{\,(3)} \oplus \cdots , \tag{24.23}$$

$$\vec{\phi}^{(2)\oplus} := \vec{0}^{\,(1)} \oplus \vec{\phi}^{\,(2)} \oplus \vec{0}^{\,(3)} \oplus \cdots . \tag{24.24}$$

For $n \neq m$ we get

$$\langle \vec{\phi}^{\,(m)\oplus} \mid \vec{\phi}^{\,(n)\oplus} \rangle^{\oplus} = 0. \tag{24.25}$$

[6] Kadison and Ringrose Vol. 1 p. 111, pp. 121–122. The decompositions in Eqs. (24.5) and (24.19) are often referred to as an *external direct sum* and an *internal direct sum*, respectively.

[7] The scalar product $\langle \vec{\phi}^{\,(n)} \mid \vec{\psi}^{\,(n)} \rangle^{(n)}$ is evaluated in $\vec{\mathcal{S}}^{\,(n)}$.

P24.1.1(4) Equations (24.20) and (24.22) can rewritten as

$$\vec{\phi}^{\,\oplus} = \sum_n \vec{\phi}^{\,(n)\oplus} \quad \text{and} \quad \vec{\psi}^{\,\oplus} = \sum_n \vec{\psi}^{\,(n)\oplus} \tag{24.26}$$

and

$$\langle \vec{\phi}^{\,\oplus} \mid \vec{\psi}^{\,\oplus} \rangle^{\oplus} = \sum_n \langle \vec{\phi}^{\,(n)\oplus} \mid \vec{\psi}^{\,(n)\oplus} \rangle^{\oplus}. \tag{24.27}$$

24.1.2 Direct Sums of Operators

Let $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$ be two Hilbert spaces, and let $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ be two operators with domains $\vec{\mathcal{D}}(\widehat{A}^{(1)}) \subset \mathcal{H}^{(1)}$ and $\vec{\mathcal{D}}(\widehat{A}^{(2)}) \subset \mathcal{H}^{(2)}$, respectively. We can define the direct sum operator as an operator acting in the direct sum space $\vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{H}}^{(1)} \oplus \vec{\mathcal{H}}^{(2)}$.

Definition 24.1.2(1) *The* **direct sum** *of $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$, denoted by $\widehat{A}^{\oplus}$ or explicitly by $\widehat{A}^{(1)} \oplus \widehat{A}^{(2)}$, is the operator acting in the direct sum space $\vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{H}}^{(1)} \oplus \vec{\mathcal{H}}^{(2)}$ by*

$$\left(\widehat{A}^{(1)} \oplus \widehat{A}^{(2)}\right)\left(\vec{\phi}^{(1)} \oplus \vec{\phi}^{(2)}\right) := \widehat{A}^{(1)}\vec{\phi}^{(1)} \oplus \widehat{A}^{(2)}\vec{\phi}^{(2)}, \tag{24.28}$$

where $\vec{\phi}^{(1)} \in \mathcal{D}(\widehat{A}^{(1)})$ and $\vec{\phi}^{(2)} \in \mathcal{D}(\widehat{A}^{(2)})$.

If $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are selfadjoint then their direct sum $\widehat{A}^{\oplus}$ is also selfadjoint.[8] Moreover, the spectrum $sp(\widehat{A}^{\oplus})$ of $\widehat{A}^{\oplus}$ is related to the spectra $sp(\widehat{A}^{(1)})$ and $sp(\widehat{A}^{(2)})$ by[9]

$$sp(\widehat{A}^{\oplus}) = sp(\widehat{A}^{(1)}) \cup sp(\widehat{A}^{(2)}). \tag{24.29}$$

When a Hilbert space is decomposed as a direct sum of a complete orthogonal family of subspaces there are operators which can be similarly decomposed. An example is a reducible operator. Let $\widehat{A}$ be a bounded operator on $\vec{\mathcal{H}}$ which admits a reducing subspace $\vec{\mathcal{S}}$. Then $\widehat{A}$ leaves the subspaces $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^{\perp}$ invariant, i.e.,

$$\widehat{A}\vec{\phi}_{\vec{\mathcal{S}}} = \vec{\psi}_{\vec{\mathcal{S}}} \in \vec{\mathcal{S}} \quad \forall \vec{\phi}_{\vec{\mathcal{S}}} \in \vec{\mathcal{S}}, \tag{24.30}$$

$$\widehat{A}\vec{\phi}_{\vec{\mathcal{S}}^{\perp}} = \vec{\psi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{S}}^{\perp} \quad \forall \vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{S}}^{\perp}. \tag{24.31}$$

[8] Blank, Exner and Havlíček p. 145.
[9] Roman Vol. 2 (1975) p. 592.

The part of $\widehat{A}$ in $\vec{\mathcal{S}}$ is given by $\widehat{A}_{\vec{\mathcal{S}}} = \widehat{P}_{\vec{\mathcal{S}}}\widehat{A}\widehat{P}_{\vec{\mathcal{S}}}$ in accordance with Eq. (17.111). The part of $\widehat{A}$ in $\vec{\mathcal{S}}^{\perp}$ is given by $\widehat{A}_{\vec{\mathcal{S}}^{\perp}} = \widehat{P}_{\vec{\mathcal{S}}^{\perp}}\widehat{A}\widehat{P}_{\vec{\mathcal{S}}^{\perp}}$. The operator $\widehat{A}$ and its parts are related by

$$\widehat{A} = \widehat{A}_{\vec{\mathcal{S}}} + \widehat{A}_{\vec{\mathcal{S}}^{\perp}} \quad \text{and} \quad \widehat{A}\vec{\phi} = \widehat{A}_{\vec{\mathcal{S}}}\,\vec{\phi}_{\vec{\mathcal{S}}} + \widehat{A}_{\vec{\mathcal{S}}^{\perp}}\vec{\phi}_{\vec{\mathcal{S}}^{\perp}}. \tag{24.32}$$

The above results can be recast in terms of direct sums as follows:

(1) Applying Eq. (24.14) we get the decomposition $\vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{S}} \oplus \vec{\mathcal{S}}^{\perp}$. A vector $\vec{\phi} = \vec{\phi}_{\vec{\mathcal{S}}} + \vec{\phi}_{\vec{\mathcal{S}}^{\perp}}$ in $\vec{\mathcal{H}}$ corresponds to

$$\vec{\phi}^{\oplus} = \vec{\phi}_{\vec{\mathcal{S}}} \oplus \vec{\phi}_{\vec{\mathcal{S}}^{\perp}} \in \vec{\mathcal{H}}^{\oplus}. \tag{24.33}$$

(2) Because of Eqs. (24.30) and (24.31) we can regard $\widehat{A}_{\vec{\mathcal{S}}}$ and $\widehat{A}_{\vec{\mathcal{S}}^{\perp}}$ as operators acting on $\vec{\mathcal{S}}$ and $\vec{\mathcal{S}}^{\perp}$, respectively. We can then define an operator $\widehat{A}^{\oplus} := \widehat{A}_{\vec{\mathcal{S}}} \oplus \widehat{A}_{\vec{\mathcal{S}}^{\perp}}$ on $\vec{\mathcal{H}}^{\oplus}$ by

$$\widehat{A}^{\oplus}\vec{\phi}^{\oplus} := \widehat{A}_{\vec{\mathcal{S}}}\,\vec{\phi}_{\vec{\mathcal{S}}} \oplus \widehat{A}_{\vec{\mathcal{S}}^{\perp}}\vec{\phi}_{\vec{\mathcal{S}}^{\perp}}. \tag{24.34}$$

The operator $\widehat{A}^{\oplus}$ is identifiable with $\widehat{A}$ due to Eq. (24.32).

Let $\vec{\mathcal{H}}$ be a Hilbert space which is decomposable into a direct sum of a complete orthogonal family of subspaces $\vec{\mathcal{S}}^{(n)}$, i.e., $\vec{\mathcal{H}} = \vec{\mathcal{H}}^{\oplus}$ as shown in Eq. (24.19). Let $\widehat{A}^{(n)}$ be a bounded operator acting on $\vec{\mathcal{S}}^{(n)}$ which is treated as a Hilbert space in its own right.[10] Then we have the following definition.

Definition 24.1.2(2) *An operator $\widehat{A}^{\oplus}$ on $\vec{\mathcal{H}}^{\oplus}$ is said to be*[11]

(1) **decomposable** *if it is a direct sum of operators on $\vec{\mathcal{S}}^{(n)}$, i.e., if*

$$\widehat{A}^{\oplus} = \oplus_n \widehat{A}^{(n)}, \tag{24.35}$$

(2) **diagonalisable** *if it is of the form*

$$\widehat{A}^{\oplus} = \oplus_n c_n \widehat{I\!I}^{(n)}, \tag{24.36}$$

where $c_n \in \mathbb{C}$ and $\widehat{I\!I}^{(n)}$ is the identity operator on $\vec{\mathcal{S}}^{(n)}$.

[10] An bounded operator on $\vec{\mathcal{H}}$ which leaves $\vec{\mathcal{S}}^{(n)}$ invariant would correspond to a bounded operator $\widehat{A}^{(n)}$ on $\vec{\mathcal{S}}^{(n)}$.

[11] It is often more transparent to write down a decomposable Hilbert space as $\vec{\mathcal{H}}^{\oplus}$, i.e., by rewriting $\vec{\mathcal{H}}$ as $\vec{\mathcal{H}}^{\oplus}$.

Decomposable operators possess the following properties[12]:

P24.1.2(1) If $\widehat{A}^{\oplus}$ is selfadjoint then $\widehat{A}^{(n)}$ is selfadjoint on $\vec{\mathcal{S}}^{(n)}$. Conversely, if every $\widehat{A}^{(n)}$ is selfadjoint on $\vec{\mathcal{S}}^{(n)}$ then $\widehat{A}^{\oplus}$ is also selfajoint.[13]

P24.1.2(2) The direct sum operator $\widehat{A}^{\oplus}$ leaves every subspace $\vec{\mathcal{S}}^{(n)\oplus}$ invariant, i.e., we have, for all $\vec{\phi}^{\,(n)\oplus} \in \vec{\mathcal{S}}^{(n)\oplus}$,

$$\widehat{A}^{\oplus}\vec{\phi}^{(n)\oplus} = \vec{\psi}^{\,(n)\oplus} \in \vec{\mathcal{S}}^{(n)\oplus}. \tag{24.37}$$

In other words every subspace $\vec{\mathcal{S}}^{\,(n)\oplus}$ reduces the operator $\widehat{A}^{\oplus}$. For any $\vec{\phi}^{\,\oplus} = \oplus_n \vec{\phi}^{\,(n)} = \sum_n \vec{\phi}^{(n)\oplus}$ we have

$$\widehat{A}^{\oplus}\vec{\phi}^{\,\oplus} = \sum_n \widehat{A}^{\oplus}\vec{\phi}^{(n)\oplus} = \sum_n \vec{\psi}^{\,(n)\oplus}. \tag{24.38}$$

P24.1.2(3) We can define the quadratic form $\mathcal{Q}\big(\widehat{A}^{\oplus}, \vec{\phi}^{\,\oplus}\big)$ by

$$\mathcal{Q}\big(\widehat{A}^{\oplus}, \vec{\phi}^{\,\oplus}\big) := \langle \vec{\phi}^{\,\oplus} \mid \widehat{A}^{\oplus}\vec{\phi}^{\,\oplus}\rangle^{\oplus}. \tag{24.39}$$

Given $\vec{\phi}^{\,\oplus} = \oplus_n c_n \vec{\phi}^{\,(n)}$ we have, by Eqs. (24.22),

$$\begin{aligned}\mathcal{Q}\big(\widehat{A}^{\oplus}, \vec{\phi}^{\,\oplus}\big) &:= \langle \vec{\phi}^{\,\oplus} \mid \widehat{A}^{\oplus}\vec{\phi}^{\,\oplus}\rangle^{\oplus} \\ &= \sum_n c_n^* c_n \,\langle \vec{\phi}^{\,(n)} \mid \widehat{A}\vec{\phi}^{\,(n)}\rangle^{(n)}.\end{aligned} \tag{24.40}$$

P24.1.2(4) A decomposable operator $\widehat{A}^{\oplus}$ is unable to relate vectors in different subspaces, i.e., by Eq. (24.37) we have

$$\langle \vec{\phi}^{\,(m)\oplus} \mid \widehat{A}^{\oplus}\vec{\phi}^{\,(n)\oplus}\rangle^{\oplus} = 0, \quad m \neq n. \tag{24.41}$$

P24.1.2(5) A Hilbert space can always be decomposed as a direct sum of any chosen complete orthogonal family of subspaces. However, *not all operators are decomposable* in such a chosen decomposition of the Hilbert space.

The mathematics of direct sums of Hilbert spaces and operators are necessary in the formulation of superselection rules in quantum mechanics to be presented in Chapter 32.

---

[12] Wan pp. 213–218. Note that not all operators in $\vec{\mathcal{H}}^{\oplus}$ are decomposable.

[13] Weidmann p. 128. Blank, Exner and Havliček p. 145. Naimark Part 2 p. 209. If Eq. (24.35) consists of an infinite sum $\widehat{A}^{\oplus}$ can become unbounded.

24.2 Tensor Products of Hilbert Spaces and Operators

24.2.1 Definitions

We shall introduce the tensor product space construction in an intuitive manner.[14] Let $\vec{\mathcal{H}}^{(1)}$ be a Hilbert space and let the vectors and scalar product in $\vec{\mathcal{H}}^{(1)}$ be denoted by

$$\vec{\phi}^{(1)}, \quad \vec{\psi}^{(1)} \quad \text{and} \quad \langle \vec{\phi}^{(1)} \mid \vec{\psi}^{(1)} \rangle^{(1)}. \tag{24.42}$$

Let $\vec{\mathcal{H}}^{(2)}$ be another Hilbert space, and let the vectors and scalar product in $\vec{\mathcal{H}}^{(2)}$ be similarly denoted.

Definition 24.2.1(1) *A Hilbert space $\vec{\mathcal{H}}^{\otimes}$ is called the* **tensor product** *of two Hilbert spaces $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$, with its relation to $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$ denoted by*

$$\vec{\mathcal{H}}^{\otimes} = \vec{\mathcal{H}}^{(1)} \otimes \vec{\mathcal{H}}^{(2)}, \tag{24.43}$$

if the following requirements are satisfied:

(1) *Every pair of vectors $\vec{\phi}^{(1)} \in \vec{\mathcal{H}}^{(1)}$ and $\vec{\phi}^{(2)} \in \vec{\mathcal{H}}^{(2)}$ corresponds a unique vector in $\vec{\mathcal{H}}^{\otimes}$, denoted by $\vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)}$, such that the scalar product of $\vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)}$ with itself is given by*[15]

$$\begin{aligned} &\langle \vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)} \mid \vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)} \rangle^{\otimes} \\ &= \langle \vec{\phi}^{(1)} \mid \vec{\phi}^{(1)} \rangle^{(1)} \langle \vec{\phi}^{(2)} \mid \vec{\phi}^{(2)} \rangle^{(2)}. \end{aligned} \tag{24.44}$$

(2) *The set of all of vectors of the form $\vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)}$ spans $\vec{\mathcal{H}}^{\otimes}$. In particular, if $\{\vec{\phi}_j^{(1)}\}$ and $\{\vec{\phi}_k^{(2)}\}$ are orthonormal bases for $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$, respectively then the set of tensor product vectors $\vec{\phi}_j^{(1)} \otimes \vec{\phi}_k^{(2)}$ form an orthonormal basis in the tensor product space $\vec{\mathcal{H}}^{\otimes}$.*

---

[14] Sewell pp. 41–42. Isham pp. 144–147. Wan pp. 95–97. *See* Reed and Simon Vol. 1 pp. 49–54, Blank, Exner and Havlíček pp. 54–57 for a more general and abstract definition

[15] $\langle \vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)} \mid \vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)} \rangle^{\otimes}$ denotes scalar product of $\vec{\phi}^{(1)} \otimes \vec{\phi}^{(2)}$ with itself in the Hilbert spaces $\vec{\mathcal{H}}^{\otimes}$.

(3) *The tensor product operation is linear, i.e.,*

$$\left(\sum_j c_j^{(1)}\,\vec{\phi}_j^{\,(1)}\right)\otimes\left(\sum_k c_k^{(2)}\,\vec{\phi}_k^{\,(2)}\right)$$
$$=\sum_{j,k} c_j^{(1)}c_k^{(2)}\left(\vec{\phi}_j^{\,(1)}\otimes\vec{\phi}_k^{\,(2)}\right),\quad c_j^{(1)},\ c_k^{(2)}\in\mathbb{C}. \tag{24.45}$$

The following comments aim to clarify the definition:

C24.2.1(1) The notation $\vec{\phi}^{\,(1)}\otimes\vec{\phi}^{\,(2)}$ is a symbolic one. The concept of a product lies in the scalar product operation which involves a multiplication of $\langle\vec{\phi}^{\,(1)}\mid\vec{\phi}^{\,(1)}\rangle^{(1)}$ and $\langle\vec{\phi}^{\,(2)}\mid\vec{\phi}^{\,(2)}\rangle^{(2)}$ in Eq. (24.44).

C24.2.1(2) The vectors $\vec{\phi}_j^{\,(1)}\otimes\vec{\phi}_k^{\,(2)}$ are orthonormal since

$$\langle\vec{\phi}_j^{\,(1)}\otimes\vec{\phi}_k^{\,(2)}\mid\vec{\phi}_m^{\,(1)}\otimes\vec{\phi}_n^{\,(2)}\rangle^{\otimes}$$
$$=\langle\vec{\phi}_j^{\,(1)}\mid\vec{\phi}_m^{\,(1)}\rangle^{(1)}\langle\vec{\phi}_k^{\,(2)}\mid\vec{\phi}_n^{\,(2)}\rangle^{(2)}=\delta_{jm}\,\delta_{kn}. \tag{24.46}$$

C24.2.1(3) A general element $\Phi^{\otimes}$ in $\vec{\mathcal{H}}^{\otimes}$ can be written as

$$\vec{\Phi}^{\otimes}=\sum_{j,k} c_{jk}\,\vec{\phi}_j^{\,(1)}\otimes\vec{\phi}_k^{\,(2)},\quad c_{jk}=\langle\vec{\phi}_j^{\,(1)}\otimes\vec{\phi}_k^{\,(2)}\mid\Phi^{\otimes}\rangle^{\otimes}, \tag{24.47}$$

since $\vec{\phi}_j^{\,(1)}\otimes\vec{\phi}_k^{\,(2)}$ form an orthonormal basis for $\vec{\mathcal{H}}^{\otimes}$.

C24.2.1(4) If $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$ are of finite dimensions N_1 and N_2, respectively then their tensor product is of dimension N_1N_2.[16]

Definition 24.2.1(2)[17] *Let $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ be two bounded operators acting on $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$, respectively. The operator, denoted by $\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}$, defined on $\vec{\mathcal{H}}^{\otimes}=\vec{\mathcal{H}}^{(1)}\otimes\vec{\mathcal{H}}^{(2)}$ by*[18]

$$\left(\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}\right)\vec{\Phi}^{\otimes}=\sum_{j,k} c_{jk}\left(\widehat{A}^{(1)}\vec{\phi}_j^{\,(1)}\right)\otimes\left(\widehat{A}^{(2)}\vec{\phi}_k^{\,(2)}\right) \tag{24.48}$$

is called the **tensor product** *of $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$.*

---

[16] In contrast the dimension of the direct sum $\vec{\mathcal{H}}^{(1)}\oplus\vec{\mathcal{H}}^{(2)}$ is N_1+N_2.
[17] Prugovečki pp. 303–304, 311. Amrein, Jauch and Sinha p. 85.
[18] Here $\vec{\Phi}^{\otimes}$ is given by Eq. (24.47).

The resulting operator $\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}$ is bounded on $\vec{\mathcal{H}}^{\otimes}$. The simplest operation of $\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}$ is on a single product vector $\vec{\phi}^{(1)}\otimes\vec{\phi}^{(2)}$. We have

$$\left(\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}\right)\left(\vec{\phi}^{(1)}\otimes\vec{\phi}^{(2)}\right)=\left(\widehat{A}^{(1)}\vec{\phi}^{(1)}\right)\otimes\left(\widehat{A}^{(2)}\vec{\phi}^{(2)}\right). \quad (24.49)$$

Tensor products of bounded operators possess the following properties:

$$c\left(\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}\right)=\left(c\widehat{A}^{(1)}\right)\otimes\widehat{A}^{(2)}=\widehat{A}^{(1)}\otimes\left(c\widehat{A}^{(2)}\right), \quad (24.50)$$

$$\left(\widehat{A}^{(1)}+\widehat{B}^{(1)}\right)\otimes\widehat{A}^{(2)}=\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}+\widehat{B}^{(1)}\otimes\widehat{A}^{(2)}, \quad (24.51)$$

$$\left(\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}\right)\left(\widehat{B}^{(1)}\otimes\widehat{B}^{(2)}\right)=\widehat{A}^{(1)}\widehat{B}^{(1)}\otimes\widehat{A}^{(2)}\widehat{B}^{(2)}, \quad (24.52)$$

$$\left(\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}\right)^{\dagger}=\left(\widehat{A}^{(1)}\right)^{\dagger}\otimes\left(\widehat{A}^{(2)}\right)^{\dagger}. \quad (24.53)$$

It follows from Eq. (24.53) that if $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are selfadjoint then $\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}$ is selfadjoint. Tensor product operators of the forms

$$\widehat{A}^{(1)}\otimes\widehat{\mathbb{I}}^{(2)}, \quad \widehat{\mathbb{I}}^{(1)}\otimes\widehat{A}^{(2)} \quad (24.54)$$

and

$$\widehat{A}^{(1)}\otimes\widehat{\mathbb{I}}^{(2)}+\widehat{\mathbb{I}}^{(1)}\otimes\widehat{A}^{(2)}, \quad (24.55)$$

where $\widehat{\mathbb{I}}^{(1)}$ and $\widehat{\mathbb{I}}^{(2)}$ are, respectively, the identity operators on $\vec{\mathcal{H}}^{(1)}$ and $\vec{\mathcal{H}}^{(2)}$ have many physical applications.[19]

Definition 24.2.1(2) cannot be applied in a straightforward manner if $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are unbounded since we have to take the domains of the operators into consideration.[20] Before we can define the tensor product $\widehat{A}^{(1)}\otimes\widehat{A}^{(2)}$ we need to introduce an operator $\widehat{A}^{(1)}\hat{\otimes}\widehat{A}^{(2)}$:

(1) Let $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ be two possibly unbounded operators defined on the domains $\vec{\mathcal{D}}(\widehat{A}^{(1)})\subset\mathcal{H}^{(1)}$ and $\vec{\mathcal{D}}(\widehat{A}^{(2)})\subset\mathcal{H}^{(2)}$, respectively, and let $\vec{\mathcal{D}}(\widehat{A}^{(1)})\hat{\otimes}\vec{\mathcal{D}}(\widehat{A}^{(2)})$ denote the set of vectors $\vec{\Phi}^{\otimes}$ in $\vec{\mathcal{H}}^{\otimes}=\vec{\mathcal{H}}^{(1)}\otimes\vec{\mathcal{H}}^{(2)}$ of the form

$$\vec{\Phi}^{\otimes}=\sum_{j}^{n_1}\sum_{k}^{n_2}c_{jk}\vec{\phi}_j^{(1)}\otimes\vec{\phi}_k^{(2)}, \quad (24.56)$$

[19] *See* Eqs. (24.71), (33.4), (33.9), (33.11).
[20] Wan pp. 96, 105.

where $\vec{\phi}_j^{(1)} \in \vec{\mathcal{D}}(\widehat{A}^{(1)})$, $\vec{\phi}_k^{(2)} \in \vec{\mathcal{D}}(\widehat{A}^{(2)})$ and n_1 and n_2 are finite. In other words, $\vec{\Phi}^{\otimes}$ is a finite linear combination of $\vec{\phi}_j^{(1)} \otimes \vec{\phi}_k^{(2)}$. The resulting set $\vec{\mathcal{D}}(\widehat{A}^{(1)}) \hat{\otimes} \vec{\mathcal{D}}(\widehat{A}^{(2)})$ is dense in $\vec{\mathcal{H}}^{\oplus}$.[21]

(2) Define an operator, denoted by $\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}$, on the domain $\vec{\mathcal{D}}(\widehat{A}^{(1)}) \hat{\otimes} \vec{\mathcal{D}}(\widehat{A}^{(2)})$ in $\vec{\mathcal{H}}^{\otimes}$ by

$$\left(\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}\right) \vec{\Phi}^{\otimes} = \sum_{j}^{n_1} \sum_{k}^{n_2} c_{jk} \left(\widehat{A}^{(1)} \vec{\phi}_j^{(1)}\right) \otimes \left(\widehat{A}^{(2)} \vec{\phi}_k^{(2)}\right). \tag{24.57}$$

Theorem 24.2.1(1)[22] *If $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are closable then $\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}$ is also closable.*

Definition 24.2.1(3)[23] *The tensor product of two closable operators $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$, to be denoted by $\widehat{A}^{(1)} \otimes \widehat{A}^{(2)}$, is defined to be the closure $\overline{\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}}$ of $\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}$ in $\mathcal{H}^{\otimes}$, i.e.,*

$$\widehat{A}^{(1)} \otimes \widehat{A}^{(2)} = \overline{\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}} \tag{24.58}$$

Theorem 24.2.1(2)[24] *If $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are selfadjoint then:*

(1) *$\widehat{A}^{(1)} \hat{\otimes} \widehat{A}^{(2)}$ is also selfadjoint.*

(2) *$\widehat{A}^{(1)} \otimes \widehat{I\!I}^{(2)} + \widehat{I\!I}^{(1)} \otimes \widehat{A}^{(2)}$ is essentially selfadjoint.*[25]

(3) *$\widehat{A}^{(1)} \otimes \widehat{I\!I}^{(2)} + \widehat{I\!I}^{(1)} \otimes \widehat{A}^{(2)}$ is selfadjoint if $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are bounded below.*[26]

---

[21] Amrein, Jauch and Sinha pp. 84–86. Reed and Simon Vol. 1 p. 298. The domains of $\widehat{A}^{(1)}$ and $\widehat{A}^{(2)}$ are assumed to be dense. Note that the set of vectors $\vec{\Phi}^{\otimes}$ are obtained by all values of n_1 and n_2.

[22] Reed and Simon Vol. 1 (1972) pp. 298.

[23] Amrein, Jauch and Sinha pp. 84–86. Reed and Simon Vol. 1 pp. 298–302. Weidmann pp. 262–268. Different authors employ the notation $\widehat{A}^{(1)} \otimes \widehat{A}^{(2)}$ to mean different things. We follow the usage of Amrein, Jauch and Sinha.

[24] Amrein, Jauch and Sinha p. 599. Reed and Simon Vol. 1 pp. 298–302. Here the operators are possibly unbounded.

[25] *See* Definition 19.5(1) for essential selfadjointness.

[26] *See* Definition 19.1(6) for operators which are bounded below.

The tensor products of unbounded selfadjoint operators can also be defined through their spectral measures and spectral functions.[27] As an example consider the operators in Eq. (24.54) where $\widehat{A}^{(1)}$ is unbounded but selfadjoint. Let $\widehat{F}^{\hat{A}^{(1)}}$ be the spectral function of $\widehat{A}^{(1)}$. We can use Definition 24.2.1(2) to define the tensor product

$$\widehat{F}^{\hat{A}^{(1)}\otimes}(\lambda) = \widehat{F}^{\hat{A}^{(1)}}(\lambda) \otimes \widehat{I\!I}^{(2)} \qquad \forall \lambda \in I\!R. \tag{24.59}$$

This can be shown to be a spectral function on the product space $\vec{\mathcal{H}}^{\otimes}$. It follows that we can define a selfadjoint operator in $\mathcal{H}^{\otimes}$ with $\widehat{F}^{\hat{A}^{(1)}\otimes}(\lambda)$ as its spectral function. The resulting operator can be identified with $\widehat{A}^{(1)} \otimes \widehat{I\!I}^{(2)}$.

24.2.2 Examples

The examples listed here have direct physical applications.

E24.2.2(1) The space $\vec{L}^2(I\!R^2)$ can be identified with the tensor product of $\vec{L}^2(I\!R)$ with itself. For clarity let us denote the Hilbert space defined by square-integrable functions on the x-axis by $\vec{L}^2(I\!R, dx)$ and Hilbert space defined by square-integrable functions on the y-axis by $\vec{L}^2(I\!R, dy)$ and the Hilbert space defined by square-integrable functions on the x-y plane by $\vec{L}^2(I\!R^2, dxdy)$. Then $\vec{L}^2(I\!R^2, dxdy)$ is related to $\vec{L}^2(I\!R, dx)$ and $\vec{L}^2(I\!R, dy)$ by

$$\vec{L}^2(I\!R^2, dxdy) = \vec{L}^2(I\!R, dx) \otimes \vec{L}^2(I\!R, dy). \tag{24.60}$$

Similarly we have

$$\vec{L}^2(I\!R^3, dxdydz) = \vec{L}^2(I\!R, dx) \otimes \vec{L}^2(I\!R, dy) \otimes \vec{L}^2(I\!R, dz). \tag{24.61}$$

We can have other tensor product expressions for $\vec{L}^2(I\!R^2, dxdy)$ and $\vec{L}^2(I\!R^3, dxdydz)$, e.g.,[28]

$$\vec{L}^2(I\!R^2, dxdy) = \vec{L}^2(I\!R^+, rdr) \otimes \vec{L}^2(\mathcal{C}_u), \tag{24.62}$$

$$\vec{L}^2(I\!R^3, dxdydz) = \vec{L}^2(I\!R^+, r^2dr) \otimes \vec{L}^2(\mathcal{S}_u). \tag{24.63}$$

[27] Prugovečki p. 304. Weidman pp. 265–267.

[28] Amrein, Jauch and Sinha p. 459. Prugovečki p. 151. Wan and Menzies. Here $\mathcal{S}_u$ denotes the unit sphere as before, and $\mathcal{C}_u$ denotes the unit circle, i.e., $\mathcal{C}_{a=1}$.

These expressions are useful when polar coordinates are used in $I\!R^2$ and spherical coordinates are used in $I\!R^3$. Here $r \in [0, \infty)$ is the radial variable. Integration in r in the plane $I\!R^2$ is respect to the volume element rdr, hence the notation $\vec{L}^2(I\!R^+, rdr)$. Integration in r in $I\!R^3$ is respect to the volume element r^2dr, hence the notation $\vec{L}^2(I\!R^+, r^2dr)$.[29]

Further examples are the tensor product of $\vec{L}^2(I\!R^3, dxdydz)$ and a finite-dimensional complex Hilbert space $\vec{\mathbb{W}}^N$, e.g.,

$$\vec{L}^2(I\!R^3, dxdydz) \otimes \vec{\mathbb{W}}^2. \tag{24.64}$$

This tensor product is relevant to the study of spin.[30]

E24.2.2(2) We can relate operators in $\vec{L}^2(I\!R, dx)$ and $\vec{L}^2(I\!R, dy)$ to that in $\vec{L}^2(I\!R^2, dxdy)$ as follows:

E24.2.2(2)(a) The position and momentum operators:

$$\hat{x}(I\!R^2) = \hat{x}(I\!R) \otimes \hat{I\!I}_y, \quad \hat{p}_x(I\!R^2) = \hat{p}(I\!R) \otimes \hat{I\!I}_y; \tag{24.65}$$

$$\hat{y}(I\!R^2) = \hat{I\!I}_x \otimes \hat{y}(I\!R), \quad \hat{p}_y(I\!R^2) = \hat{I\!I}_x \otimes \hat{p}(I\!R). \tag{24.66}$$

These results can be readily extended to $\vec{L}^2(I\!R^3, dxdydz)$.[31]

E24.2.2(2)(b) Annihilation, creation and number operators: Let $\hat{a}_x$ and $\hat{a}_x^\dagger$ be a pair of creation and annihilation operators defined in $\vec{L}^2(I\!R, dx)$, and let $\hat{a}_y$ and $\hat{a}_y^\dagger$ be a pair of annihilation and creation operators defined in $\vec{L}^2(I\!R, dy)$. We can define their corresponding operators in $L^2(I\!R^2, dxdy)$ treated as a tensor product space shown in Eq. (24.60) by

$$\hat{a}_x^\otimes := \hat{a}_x \otimes \hat{I\!I}_y, \quad \hat{a}_x^{\otimes\dagger} := \hat{a}_x^\dagger \otimes \hat{I\!I}_y; \tag{24.67}$$

$$\hat{a}_y^\otimes := \hat{I\!I}_x \otimes \hat{a}_y, \quad \hat{a}_y^{\otimes\dagger} := \hat{I\!I}_x \otimes \hat{a}_y^\dagger. \tag{24.68}$$

Let $\hat{N}_x = \hat{a}_x^\dagger \hat{a}_x$ and $\hat{N}_y = \hat{a}_y^\dagger \hat{a}_y$. Then we have

$$\hat{N}_x^\otimes := \hat{N}_x \otimes \hat{I\!I}_y, \quad \hat{N}_y^\otimes := \hat{I\!I}_x \otimes \hat{N}_y; \tag{24.69}$$

$$\hat{N}^\otimes := \hat{N}_x \otimes \hat{I\!I}_y + \hat{I\!I}_x \otimes \hat{N}_y. \tag{24.70}$$

---

[29] Amrein, Jauch and Sinha p. 459. The corresponding set of functions are denoted by $L^2(I\!R^+, rdr)$ and $L^2(I\!R^+, r^2dr)$.

[30] *See* Eq. (33.10).

[31] These are the same as the corresponding operators introduced earlier, e.g., in Eq. (17.52).

The operator $\widehat{N}^{\otimes}$ is selfadjoint by Theorem 24.2.1(2) since $\widehat{N}_x$ and $\widehat{N}_y$ are bounded below.

E24.2.2(2)(c) Let $\widehat{H}_{ho,x}$ denote the operator in Eq. (19.54) defined in $\vec{L}^2(I\!R, dx)$, and let $\widehat{H}_{ho,y}$ denote the corresponding operator defined in $\vec{L}^2(I\!R, dy)$. Then we can define an operator in $\vec{L}^2(I\!R, dxdy)$ by

$$\widehat{H}_{2iso} := \widehat{H}_{ho,x} \otimes \widehat{I\!I}_y + \widehat{I\!I}_x \otimes \widehat{H}_{ho,y}. \tag{24.71}$$

The resulting operator is the Hamiltonian for an *isotropic oscillator* described in §35.4.

E24.2.2(3) The operators $\widehat{L}_z(\mathcal{S}_u)$ and $\widehat{L}^2(\mathcal{S}_u)$ in $\vec{L}^2(\mathcal{S}_u)$ defined by Eq. (17.42) and the operator expression in Eq. (19.50) can be extended to $\vec{L}^2(I\!R^3, dxdydz)$, i.e., we have

$$\widehat{L}_z(I\!R^3, dxdydz) := \widehat{I\!I}(I\!R^+) \otimes \widehat{L}_z(\mathcal{S}_u), \tag{24.72}$$
$$\widehat{L}^2(I\!R^3, dxdydz) := \widehat{I\!I}(I\!R^+) \otimes \widehat{L}^2(\mathcal{S}_u), \tag{24.73}$$

where $\widehat{I\!I}(I\!R^+)$ is the identity operator on $\vec{L}^2(I\!R^+, r^2dr)$. This operator can be identified with the operator for the total angular momentum square discussed in §36.1.1.

E24.2.2(4) If $\widehat{A}^{(1)}$ in $\vec{\mathcal{H}}^{(1)}$ possesses eigenvectors $\vec{\varphi}_m^{(1)}$ corresponding to eigenvalues $a_m^{(1)}$, and $\widehat{A}^{(2)}$ in $\vec{\mathcal{H}}^{(2)}$ possesses eigenvectors $\vec{\varphi}_n^{(2)}$ corresponding to eigenvalues $a_n^{(2)}$, then $\widehat{A}^{(1)} \otimes \widehat{A}^{(2)}$ admits $\vec{\varphi}_m^{(1)} \otimes \vec{\varphi}_n^{(2)}$ as eigenvectors corresponding to eigenvalues $a_m^{(1)} a_n^{(2)}$, i.e.,

$$\widehat{A}^{(1)} \otimes \widehat{A}^{(2)} \left(\vec{\varphi}_m^{(1)} \otimes \vec{\varphi}_n^{(2)} \right) = a_m^{(1)} a_n^{(2)} \left(\vec{\varphi}_m^{(1)} \otimes \vec{\varphi}_n^{(2)} \right). \tag{24.74}$$

We can define the tensor products of a finite number of Hilbert spaces and operators in a similar manner.

The importance of tensor products lies in the study of many-particle systems. An n-particle system would require a Hilbert space formed by the tensor product of the n Hilbert spaces of the individual particle. As will be seen in §33.4 later these tensor product spaces for various number of particles can be "added up", using the concept of direct sum, to form a new Hilbert space, known a *Fock space*, capable

of describing a system with indefinite numbers of particles.[32] Such a construction provides the basis for the formulation of quantum filed theory.

Exercises and Problems

Q24(1) Consider a three-dimensional Hilbert space $\vec{\mathcal{H}}$ with a preferred decomposition as the direct sum of a complete orthogonal family of one-dimensional subspaces $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$, i.e.,

$$\vec{\mathcal{H}} = \vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{S}}^{(-)} \oplus \vec{\mathcal{S}}^{(0)} \oplus \vec{\mathcal{S}}^{(+)}. \tag{24.75}$$

Let $\vec{\eta}^{(-)}$, $\vec{\eta}^{(0)}$ and $\vec{\eta}^{(+)}$ be unit vectors in $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$, respectively.[33] A vector in $\vec{\mathcal{H}}^{\oplus}$ is of the form

$$\begin{aligned} \vec{\eta}^{\oplus} &= c_- \vec{\eta}^{(-)} \oplus c_0 \vec{\eta}^{(0)} \oplus c_+ \vec{\eta}^{(+)} \\ &= c_- \vec{\eta}^{(-)\oplus} + c_0 \vec{\eta}^{(0)\oplus} + c_+ \vec{\eta}^{(+)\oplus}, \end{aligned} \tag{24.76}$$

where $c_-,\ c_0,\ c_+ \in \mathbb{C}$.

(a) Show that selfadjoint decomposable operators $\widehat{A}^{\oplus}$ on $\vec{\mathcal{H}}^{\oplus}$ are diagonalisable and of the form

$$\widehat{A}^{\oplus} = a_- \widehat{I\!I}^{(-)} \oplus a_0 \widehat{I\!I}^{(0)} \oplus a_+ \widehat{I\!I}^{(+)}, \quad a_-, a_0, a_+ \in I\!R. \tag{24.77}$$

What are the eigenvalues and eigenvectors of $\widehat{A}^{\oplus}$?

(b) Define three operators $\widehat{L}_-$, $\widehat{L}_+$ and $\widehat{L}$ on $\vec{\mathcal{H}}^{\oplus}$ by[34]

$$\widehat{L}_- \vec{\eta}^{(-)\oplus} = \vec{\eta}^{(0)\oplus}, \quad \widehat{L}_- \vec{\eta}^{(0)\oplus} = \vec{\eta}^{(-)\oplus}, \quad \widehat{L}_- \vec{\eta}^{(+)\oplus} = \vec{0}^{\oplus}. \tag{24.78}$$

[32] Fock (1896–1974) was a Soviet physicist.

[33] The notation such as $\vec{\eta}^{(-)\oplus}$ follows that of Eq. (24.23), i.e., $\vec{\eta}^{(-)\oplus} = \vec{\eta}^{(-)} \oplus \vec{0}^{(0)} \oplus \vec{0}^{(+)}$ which is a vector in $\vec{\mathcal{H}}^{\oplus}$.

[34] Wan p. 356. These operators are not the direct sums of operators on $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$. Hence they are not denoted with a superscript $\oplus$. Here $\vec{0}^{\oplus}$ is the zero operator on $\vec{\mathcal{H}}^{\oplus}$.

$$\widehat{L}_+\vec{\eta}^{\,(+)\oplus} = \vec{\eta}^{\,(0)\oplus}, \quad \widehat{L}_+\vec{\eta}^{\,(0)\oplus} = \vec{\eta}^{\,(+)\oplus}, \quad \widehat{L}_+\vec{\eta}^{\,(-)\oplus} = \vec{0}^{\oplus}. \tag{24.79}$$

$$\widehat{L} = \widehat{L}_- + \widehat{L}_+, \tag{24.80}$$

Show that these operators are selfadjoint but not decomposable.[35]

Q24(2) Show that the tensor products of two projectors, $\widehat{P}^{(1)}$ in Hilbert space $\vec{\mathcal{H}}^{(1)}$ and $\widehat{P}^{(2)}$ in Hilbert space $\vec{\mathcal{H}}^{(2)}$, is a projector in the tensor product space $\vec{\mathcal{H}}^{\otimes} = \vec{\mathcal{H}}^{(1)} \otimes \vec{\mathcal{H}}^{(2)}$.

Q24(3) Consider the tensor product $\vec{\mathcal{H}}^{\otimes} = \vec{\mathcal{H}} \otimes \vec{\mathcal{H}}$. Let $\{\vec{\varphi}_m\}$ be an orthonormal basis for $\vec{\mathcal{H}}$. Then $\{\vec{\varphi}_m \otimes \vec{\varphi}_n\}$ is an orthonormal basis for $\vec{\mathcal{H}}^{\otimes}$. Define the **permutation operator** $\widehat{U}_p$ on $\vec{\mathcal{H}}^{\otimes}$ by[36]

$$\widehat{U}_p\Big(\sum_{m,n} c_{mn}\vec{\varphi}_m \otimes \vec{\varphi}_n\Big) := \sum_{m,n} c_{mn}\,\vec{\varphi}_n \otimes \vec{\varphi}_m. \tag{24.81}$$

(a) Show that this is a bounded operator with $\vec{\mathcal{H}}^{\otimes}$ as its domain.

(b) Show that the square of $\widehat{U}_p$ is equal to the identity, and $\widehat{U}_p$ is unitary and selfadjoint, i.e.,

$$\widehat{U}_p^{\,2} = \widehat{\mathbb{I}}, \quad \widehat{U}_p^{\,\dagger} = \widehat{U}_p^{\,-1} = \widehat{U}_p. \tag{24.82}$$

(c) For two bounded operators $\widehat{A}$ and $\widehat{B}$ on $\vec{\mathcal{H}}^{\otimes}$ show that[37]

$$\widehat{U}_p\,(\widehat{A} \otimes \widehat{B})\,\widehat{U}_p^{\,\dagger} = \widehat{B} \otimes \widehat{A}. \tag{24.83}$$

(d) Show that

$$\widehat{P}^{(s)} := \frac{1}{2}\Big(\widehat{\mathbb{I}} + \widehat{U}_p\Big), \quad \widehat{P}^{(a)} := \frac{1}{2}\Big(\widehat{\mathbb{I}} - \widehat{U}_p\Big), \tag{24.84}$$

are projectors which are orthogonal to each other. Find examples of vectors in $\vec{\mathcal{H}}^{\otimes}$ which are unchanged by each of these two projectors.

(e) Find the eigenvalues and eigenvectors of $\widehat{U}_p$.

[35] *See* §32.3 and §34.7 for physical applications and Q32(3) for a similar operator in a two-dimensional space.
[36] *See* §33.3 for physical applications of these operators.
[37] *See* Eq. (33.15) in §33.3.

Section III

Quantum Formalism

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 25

Pure States

In Chapter 5, we have outlined the general properties of quantum systems. Some model systems have been presented in Chapters 10 and 14 to illustrate how we can formulate theories having probabilistic features in both real and complex vector spaces. Systems with observables taking on only a finite number of values are illustrated with model theories of spin in finite-dimensional complex vector spaces. The formulation follows from the general structure of physical theories, particularly the structure of quantum mechanics, given in Chapters 1 and 5. Observables of a general quantum system are capable of having an infinite number of different values. It follows that a general quantum theory must be formulated in an infinite-dimensional complex vector space, i.e., a Hilbert space. A general scheme for generating probability distributions in a Hilbert space is given in Chapter 22.

We are now in a position to present a systematic treatment of quantum theory in terms of six groups of postulates on the fundamental properties of quantum systems:

(1) Postulate on the description of pure states.
(2) Postulate on the description of observables and their values.
(3) Postulate on the transition from classical to quantum systems.
(4) Postulates on the probabilistic relationship between states and observables.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

(5) Postulates on the time evolution of the system.
(6) Postulate on states after measurement.

We then extend the basic theory to discuss mixed states, superselection rules and many-particle systems. A brief discussion of conceptual problems in quantum mechanics will also be presented.

25.1 Postulate (PS)

As pointed out in the discussion of quantum property QMP5.3(1) not all observables of quantum system are compatible, i.e., not all observables can be measured simultaneously. However, various complete sets of compatible observables exist. A maximum amount of information about the system corresponds to a set of simultaneously measured values of a complete set of discrete observables of the system.[1] Such an amount of information characterises a state of the system. States characterised by such a maximum amount of information about the system are called *pure states*. If we do not have a maximum amount of information about the system, we cannot determine a pure state. Such a situation does occur in many practical cases. It is still desirable to have a characterisation of the system based on the information practically available. Such a characterisation which is based on less than a maximum set of data about the system is called a *mixed state*. We shall start with a postulate on pure states. A more general postulate which includes mixed states will be given in Chapter 31.

Postulate 25.1(PS) *The* **state space** *of a quantum system is a complex Hilbert space. A* **pure state** ϕ^s *which corresponds to a maximum possible amount of information about the system is exhaustively describable by a unit vector* $\vec{\phi}$ *in the Hilbert space.*

The following comments aim to clarify this postulate:

C25.1(1) In quantum mechanics the description of a quantum state by a unit vector in the state space is exhaustive in the sense

[1] Recall that observables with a purely discrete set of values are referred to as *discrete observables* and those with a purely continuous set of values are referred to as *continuous observables*. These terms are introduced in §5.3.

that there is no finer or more detailed description possible. We call such unit vectors **state vectors**. A state is meant to be a pure state in what follows unless otherwise is specifically stated. The state ϕ^s and its state vector $\vec{\phi}$ are shown by the notation $\phi^s := \vec{\phi}$.

C25.1(2) There are theories proposing a more detailed description of a state in terms of *hidden variable*.[2]

C25.1(3) The description of states by unit vectors are motivated by the need to describe the probabilistic behaviour of quantum systems. As seen in the model theory for electron spin in §14.1.1 unit vectors can generate probability distributions. However, as pointed out in relation to Theorem 22.1(1) two unit vectors differing by a phase factor would generate the same probability distribution. This means that they are equivalent for the description of a state.[3] Non-unit vectors $\vec{\psi}$ can also describe a state provided we use its normalised counterpart $\vec{\psi}/||\vec{\psi}||$. All the vectors in the one-dimensional subspace spanned by a unit vector $\vec{\phi}$ would correspond to the same state. We can re-state the description of pure states in Postulate 25.1(PS) by saying that

> *a pure state is describable by a one-dimensional subspace of the state space.*

We can go one step further by saying that[4]

> *a pure state is describable by a one-dimensional projector on the state space.*

C25.1(4) While every pure state is described by a unit vector the converse is not assumed in the postulate. However, there are many systems for which the converse statement is true, i.e., there are

[2] Beltrametti and Cassinelli pp. 171–176 for a concise discussion of hidden variables theories.

[3] *See* also §28.1 and §28.2 .

[4] Beltrametti and Cassinelli p. 45. This is because one-dimensional subspaces correspond one-to-one to one-dimensional projectors. The probabilistic behaviour of quantum systems can be described directly in terms of one-dimensional projectors. A systematic approach to describe states in terms of operators will be presented in Chapter 32.

systems for which every unit vector of the Hilbert space describes a pure state. For such systems we have

> *pure states correspond one-to-one to one-dimensional subspaces of the state space $\vec{\mathcal{H}}$, or equivalently pure states correspond one-to-one to one-dimensional projectors on the state space $\vec{\mathcal{H}}$.*

C25.1(5) It is possible to formulate a theory in which not every unit vector describes a pure state. The resulting theory can lead to *superselection rules* which are applicable to some systems.[5] A formulation of superselection rules is presented in Chapter 32. The property stated in italic in C25.1(4) above is for systems which do not admit superselection rules.

C25.1(6) In §30.3 we shall discuss how a set of measured values of a complete set of compatible discrete observables determines a unit vector, or more precisely a one-dimensional subspace, for the description the state and why continuous observables are not generally suitable for the determination of a pure state.

C25.1(7) For systems without superselection rules a normalised linear combination of two or more state vectors would describe a new state. For example, if $\vec{\phi}_1$ and $\vec{\phi}_2$ are two orthogonal state vectors, then

$$\vec{\phi} = \frac{1}{\sqrt{3}}\,\vec{\phi}_1 + \sqrt{\frac{2}{3}}\,\vec{\phi}_2 \tag{25.1}$$

is normalised and $\vec{\phi}$ can serve as a state vector describing a new state, known as a **coherent superposition** of $\vec{\phi}_1$ and $\vec{\phi}_2$. Similarly, if $\vec{\phi}_1, \vec{\phi}_2, \vec{\phi}_3, \ldots$ are an orthogonal set of state vectors then

$$\vec{\phi} = \sum_\ell c_\ell\,\vec{\phi}_\ell, \quad \sum_\ell |c_\ell|^2 = 1 \tag{25.2}$$

is normalised and $\vec{\phi}$ can describe a new state. Such linear combinations of states are the basis of superposition principle described in quantum property QMP5.3(4). We shall explain how

[5] Superselection rules were introduced by Wick, Wightman and Wigner (1952) in particle physics.

a superposition would lead to information not contained in the individual constituent states in §31.5.

When a superselection rule exists, a normalised linear combination may not describe a pure state. *Such a combination is not referred to as a coherent superposition*.[6] The superposition principle does not apply to these states.

For clarity let us set out the notation and terminology:

(1) States are denoted by Greek letters with a superscript s, e.g., ϕ^s, ψ^s, ...

(2) Unit vectors representing states are denoted by the usual vector notation without the superscript, e.g., $\vec{\phi}$, $\vec{\psi}$, ... These are referred to as **state vectors**. The one-dimensional subspaces spanned by these vectors are denoted by $\vec{\mathcal{S}}_{\vec{\phi}}$, $\vec{\mathcal{S}}_{\vec{\psi}}$, ... and the projectors generated by these vectors are denoted by $\widehat{P}_{\vec{\phi}}$, $\widehat{P}_{\vec{\psi}}$, ...[7]

(3) A set of states is said to be a **complete set of states** if the corresponding state vectors form a complete set of vectors. If the corresponding state vectors do not form a complete set then the corresponding set of states is said to be *incomplete*. A *complete orthonormal set of states* corresponds to a complete orthonormal set of state vectors.

(4) It is a common practice to refer to state vectors simply as states with the symbol $\vec{\phi}$ denoting both the state and the vector representing the state.

(5) Complex-valued functions on $\mathbb{R}$ are denoted by showing their argument explicitly, e.g., $\phi(x)$, $\psi(x)$, ... As discussed in §16.1.2 these functions can define Hilbert spaces. For example, the set $L^2(\Lambda)$ of square-integrable functions $\phi(x)$, $\psi(x)$, ... on Λ defines a Hilbert space which is denoted by $\vec{L}^2(\Lambda)$. Vectors in $\vec{L}^2(\Lambda)$ corresponding to the functions $\phi(x)$, $\psi(x)$, ... are denoted by $\vec{\phi}$, $\vec{\psi}$, ... As in Eq. (12.24) we shall denote the relationship by $\vec{\phi} := \phi(x)$. The scalar product is given by Eq. (16.34).

[6] *See* Chapter 32.

[7] This is in line with the notation in Eqs. (6.52) and (9.13).

(6) Square-integrable functions $\phi(x), \psi(x), \ldots$ of the position variable used to define a state space of a quantum system are referred to as **wave functions**, a term first introduced in §10.3. A wave function may be real or complex.

Finally we should mention that states can be determined by measurements. This is discussed in §26.2.3 and C28.2(7) in Chapter 28.

Exercises and Problems

Q25(1) Explain the concept of pure states.

Q25(2) Explain why only unit vectors are used for state description in Postulate 25.1(PS).[8]

Q25(3) Explain why pure states do not correspond one-to-one to unit vectors in the state space.

[8] *See* also the discussion in §14.1.1 and §22.1.

Chapter 26

Observables and Their Values

26.1 Postulate (OV)

Postulate 26.1(OV) *An observable A of a quantum system is describable by a selfadjoint operator $\widehat{A}$ in the state space $\vec{\mathcal{H}}$ of the system. The measurable values of the observable are given by the elements of the spectrum $sp(\widehat{A})$ of the operator.*

The following comments serve to clarify the postulate:

C26.1(1) An observable A is also said to *correspond* to or to be *represented* by a selfadjoint operator $\widehat{A}$. When these terms are used we have in mind that the measurable values of the observable are the elements of the spectrum $sp(\widehat{A})$ of the operator. It is also common to use the word observable to mean both the physical quantity A and the corresponding operator $\widehat{A}$ when the usage does not cause any confusion.

C26.1(2) In view of the one-to-one correspondence between selfadjoint operators and spectral functions we can describe observables directly in terms of spectral functions.[1] We can rephrase the first statement of Postulate 26.1(OV) as

[1] This is the description adopted by the quantum logic approach.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

an observable of a quantum system is describable by a spectral function on the state space $\vec{\mathcal{H}}$.

C26.1(3) In classical mechanics observables are functions of the state while in quantum theory there is no directly mathematical relationship between a quantum state ϕ^s and an arbitrary observable A. Postulates 25.1(PS) and 26.1(OV) make no mention of how a quantum state ϕ^s may provide information on the values of observable A, e.g., the postulates do not tell us whether or not observable A has a definite value in a given state ϕ^s. The relationship has to be stated by a separate postulate.[2]

C26.1(4) There are two categories of quantum observables:

(1) Quantum observables without a classical counterpart. The spin of an electron is a familiar example. An axiomatic approach is used in selecting the Hilbert space and the operators for the description of spin in order to reflect the properties of spin obtained by experiments. The theory of electron spin presented in Chapter 14 serves to illustrate this approach.

(2) Quantum observables with a classical counterpart. The approach here is different. Based on some fundamental properties of the corresponding classical observables a *quantisation process* can often be established to obtain the corresponding quantum observables. An example is a harmonic oscillator. A classical harmonic oscillator is characterised by a classical Hamiltonian H_{ho} given in Eq. (27.11). To establish the quantised harmonic oscillator, we first associate the Hilbert space $\vec{L}^2(I\!R)$ with it. Observables correspond to selfadjoint operators in $\vec{L}^2(I\!R)$. For example, we have

   (a) The position and momentum correspond to operators $\widehat{x}$ in Eq. (17.12) and $\widehat{p}$ in Eq. (17.50), respectively.

   (b) The Hamiltonian corresponds to the operator $\widehat{H}_{ho}$ in Eq. (19.54).

   The quantisation process can be rather complicated and there is as yet no universal scheme capable of generating quantum observables to correspond to all classical observables, despite a

[2] *See* Chapter 28.

great deal of research into this area of studies over the years.[3] In the next chapter we shall present a generally accepted scheme for the quantisation of familiar observables such as position, momentum, angular momentum and the Hamiltonian.

C26.1(5) The set of all possible values of an observable A will be referred to as the **spectrum of the observable** to be denoted by $sp(A)$. We can classify observables according to the nature of their spectra into three different types based on the spectra of their corresponding operators set out in §20.2:

(1) **Discrete observables** These are observables A described by selfadjoint operators $\widehat{A}$ with a discrete spectrum $sp_d(\widehat{A}) = \{a_1, a_2, \cdots\}$. The eigenvalues a_ℓ of $\widehat{A}$ are the measurable values of A.

(2) **Continuous observables**[4] These are observables described by selfadjoint operators $\widehat{A}$ with a continuous spectrum $sp_c(\widehat{A}) = \{\tau \in \Lambda \subset I\!R\}$. A continuous observable has a continuous set of values λ which coincides with the spectrum $sp_c(\widehat{A})$.

(3) **Others** There are observables which can have both a discrete set and a continuous set of values. An example is the Hamiltonian operator of a particle in a finite square well potential.[5]

We shall concentrate on discrete and continuous observables in what follows. An observable which is partly discrete and partly continuous can be studied through discrete and continuous observables.

C26.1(6) A real-valued function of the observable $f(A)$ is described by the corresponding function of the selfadjoint operator $\widehat{A}$, i.e., by $f(\widehat{A})$.[6] For example, the observable A^2 is described by $\widehat{A}^2$.

---

[3] *See* Wan pp. 252–282, pp. 443–505.

[4] We have not attached the terms *discrete* and *continuous* to operators to avoid confusion, e.g., continuous operators are defined by Definition 17.1(3) which is not related to the continuity of their spectrum.

[5] Zettili p. 224.

[6] *See* §20.5 for the definition of functions of selfadjoint operators. This property is consistent with the function preserving quantisation rule associated with Postulate 27.2(CQ) in §27.2. *See* Isham pp. 81–84, 161–168.

C26.1(7) Projectors are selfadjoint. It follows that they can describe observables. Observables which are associated with projectors are called **propositions**. These observables are conceptually simple but important. Propositions are the basis for a *quantum logic* formulation of quantum mechanics.[7] We shall examine the physics of these observables in the next section.

C26.1(8) Postulates 25.1(PS) and 25.1(OV) are not explicitly related to quantum properties QMP5.3(1) to QMP5.3(2) on states and observables. These properties would emerge from a separate postulate on the probability distribution of values of an observable in a given state presented in Chapter 28.

C26.1(9) In §5.2 the concepts of *possessed values* and *measured values* of an observable are introduced. It is pointed out that generally a measured value may not be the value the observable has before the measurement. The question then arises as to whether the measured value is the value the observable has immediately after the measurement. For classical systems the answer to the question is always a yes, but the same cannot be said about quantum systems. However, for many quantum measurements the answer to this question is still a yes. This kind of measurements are already discussed in QMP5.3(3), i.e., these are **ideal measurements**. From now on a measurement is meant to be an ideal measurement, unless otherwise is specifically stated. Further discussions on this problem are given in §30.1.2 and §34.7.1.

C26.1(10) The postulate does not assume a one-to-one correspondence between observables of the quantum system and selfadjoint operators in $\vec{\mathcal{H}}$. As mentioned in relation to Postulate 25.1(PS) the absence of such a one-to-one correspondence can lead to a theory with superselection rules. Traditional quantum systems, e.g., electrons, spin systems, have the properties that every unit vector describes a state and every selfadjoint operator describes an observable. We call such systems orthodox quantum systems.

[7] Isham pp. 168–178. Mackey pp. 56–81. Jauch pp. 67–110. Beltrametti and Cassinelli Chapter 10. Cohen.

Definition 26.1(1) *A quantum system is called an* **orthodox quantum system** *if it satisfies*[8]:

(1) Postulate 25.1(PS). *In addition there is a one-to-one correspondence between pure states and one-dimensional subspaces.*
(2) Postulate 26.1(OV). *In addition there is a one-to-one correspondence between observables and selfadjoint operators.*

From now on a quantum system means an orthodox quantum system, unless otherwise is stated. For an orthodox quantum system, every normalised linear combination of state vectors is a coherent superposition., i.e., the superposition principle applies without any restriction.

26.2 On Propositions

26.2.1 Definition

Definition 26.2.1(1) *An observable P is called a* **proposition** *of a quantum system if it is represented by a projector $\widehat{P}$ on the state space of the system.*

A proposition is a discrete observable having only two values, i.e., 1 and 0, which are the eigenvalues of its associated projector. Physically[9]

> *a proposition is a statement about the system which is either true or false*.

An experiment can be performed to test whether a proposition is true or false. An experiment to measure a proposition is called a **yes-no experiment**, since we can arrange the experiment so that the measured value 1 would correspond to the answer of *yes* to the proposition and the value 0 would correspond to the answer of *no* to the proposition. Propositions are closely related to general

[8] Beltrametti and Cassinelli p. 45. Fano p. 391. These are in contrast to *mixed quantum systems* described in Definition 32.1(4).
[9] Isham pp. 168–178 for a discussion of various interpretations. This kind of observables also exist in classical mechanics (*see* Isham pp. 61–65).

observables and states as seen in the discussion in the following subsections.

26.2.2 Observables and Their Propositions

Since a selfadjoint operator has a spectral decomposition in terms of projectors, we can associate a set of propositions to each observable.

26.2.2.1 Discrete observables

As shown in Eq. (20.20), the selfadjoint operator $\widehat{A}$ associated with a discrete observable A has the following spectral decomposition:

$$\widehat{A} = \sum_m a_m \widehat{P}^{\hat{A}}(a_m). \tag{26.1}$$

The projector $\widehat{P}^{\hat{A}}(a_m)$ describes a proposition, i.e.,[10]

> *the proposition that a measurement of the observable A will result in the value a_m.*

The harmonic oscillator has a Hamiltonian $\widehat{H}_{ho}$ given by Eq. (19.54). This Hamiltonian represents the energy of the oscillator. The operator $\widehat{H}_{ho}$ has a nondegenerate spectrum with eigenvalues E_n given by Eq. (20.22) corresponding to eigenvectors $\vec{\varphi}_n$. Its spectral decomposition is

$$\widehat{H}_{ho} = \sum_{n=0}^{\infty} E_n \widehat{P}^{\hat{H}_{ho}}(E_n). \tag{26.2}$$

Each projector $\widehat{P}^{\hat{H}_{ho}}(E_n)$ describes a proposition, i.e.,

> *the proposition that a measurement the energy of the oscillator will result in the value E_n.*

These propositions are measured by a yes-no experiment, e.g., in a yes-no experiment of the proposition represented by $\widehat{P}^{\hat{H}_{ho}}(E_n)$ a yes answer means the energy value E_n is obtained.

[10] *See* Isham p. 85 for an alternative statement of the proposition.

The above discussion tells us that

(1) A discrete observable has a set of propositions associated with it. These propositions correspond to the set of eigenprojectors $\widehat{P}^{\hat{A}}(a_m)$ of its associated selfadjoint operator $\widehat{A}$.

(2) The observable can be physically characterised by its associated set of propositions in that measurement of the observable is equivalent to measurements of its propositions.

26.2.2.2 Continuous observables

Propositions of a continuous observable A are described by the spectral projectors $\widehat{M}^{\hat{A}}(\Lambda)$ of its associated selfadjoint operator $\widehat{A}$ with the interpretation that $\widehat{M}^{\hat{A}}(\Lambda)$ corresponds to[11]

> *the proposition that a measurement the observable A will result in a value τ in the Borel set Λ.*

These propositions are measured by yes-no experiment. An obvious example is the position operator $\widehat{x}$ in $\vec{L}^2(I\!R)$. Its spectral projectors $\widehat{M}^{\hat{x}}(\Lambda)$ are defined by characteristic functions as shown in Eq. (20.28). We interpret $\widehat{M}^{\hat{x}}(\Lambda)$ as

> *the proposition that a measurement of the position will result in a value τ in the Borel set Λ.*

We shall call the observables corresponding to these spectral projectors **position propositions** or more intuitively **local position observables**.[12] A discussion of yes-no experiments for local position observables will be presented in §30.2.2.

26.2.3 Propositions and States

From C25.1(4) we know that a pure state ϕ^s described by the state vector $\vec{\phi}$ is also describable by the projector $\widehat{P}_{\vec{\phi}}$ generated by $\vec{\phi}$, i.e., the projector $\widehat{P}_{\vec{\phi}}$ plays a dual role. In addition to describing a pure

[11] *See* Definition 15.3(1) and P15.3(2) for the distinction between spectral projectors and eigenprojectors.

[12] One can also call the observables corresponding to $\widehat{x}\widehat{M}^{\hat{x}}(\Lambda) := x\,\chi_\Lambda(x)$ *local position observables*. However, these are not propositions and they are not measured by yes-no experiments.

state it also represents an observable in its own right, since it is a selfadjoint operator. The observable corresponds to[13]

> *the proposition that the system is in the state ϕ^s described by the state vector $\vec{\phi}$.*

When the measured value of $\widehat{P}_{\vec{\phi}}$ is 1 in a yes-no experiment, the answer to the proposition is a *yes*, i.e., the system is in state ϕ^s immediately after the measurement.[14] A yes-no experiment to measure $\widehat{P}_{\vec{\phi}}$ is not a measurement of the state.[15]

Exercises and Problems

Q26(1) Discuss the fundamental differences between classical observables and quantum observables.

Q26(2) What are the measurable values of a function $f(A)$ of a discrete observable A described by a selfadjoint operator $\widehat{A}$ which has a discrete spectrum $sp_d = \{a_1, a_2, \ldots\}$?

Q26(3) Give a brief account of the concept of propositions in quantum mechanics.

Q26(4) The spectral projector $\widehat{M}^{\hat{x}}(\Lambda)$ of the position operator $\widehat{x}$ for an interval Λ is defined by a characteristic function in Eq. (20.28). What is the physical meaning of $\widehat{M}^{\hat{x}}(\Lambda)$ as a proposition? What physical devices are capable of measuring $\widehat{M}^{\hat{x}}(\Lambda)$?

Q26(5) A state ϕ^s corresponds to the one-dimensional projector $\widehat{P}_{\vec{\phi}}$ generated by the state vector $\vec{\phi}$. What is the meaning of the proposition corresponding to the projector $\widehat{P}_{\vec{\phi}}$?

---

[13] Wan pp. 242–243.

[14] We assume ideal measurement here. *See* QMP5.3(3) on ideal measurement and Chapter 30 for discussion on state after a measurement.

[15] A state of a classical particle can be numerically identified, i.e., by a set of values of the particle's position and momentum which can be directly measured. In contrast a quantum state ϕ^s has no value in its own right. A quantum state is considered measurable in terms of an appropriate set of yes-no experiments corresponding to the set of propositions of the system (*see* C28.2(7) of Chapter 28 and Jauch pp. 93–94).

Chapter 27

Canonical Quantisation

Many important quantum observables, e.g., angular momentum, have their origin in corresponding classical observables. Postulate 26.1(OV) does not tell us how to find the operators to describe these observables. We can tackle this problem by formulating a process of *quantisation* to establish the operators for the corresponding quantum observables. Before we can formulate a quantisation process we must choose an appropriate description of classical observables. This is done in the following section.

27.1 Hamiltonian Formulation of Classical Mechanics

Newton's theory is not suitable when it comes to relate classical theory to quantum theory. As discussed in §5.1 Newton's theory is based on the notion of instantaneous velocity which is not appropriate for quantum particles. It is Hamilton's formulation which enables us to make a direct transition to quantum theory, since Hamilton's formulation is based position and momentum rather than velocity. Hamilton's formulation is evolved from an earlier Lagrangian theory. A general formulation of Hamilton's

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

theory is rather complicated.[1] For our purposes we only need to confine our discussion to some specific systems which can be treated easily.

27.1.1 Conservative Mechanical Systems

Consider the motion of a particle in the physical space under a conservative force $\vec{f}$. A force $\vec{f}$ is conservative if it is derivable from the gradient of a scalar function $V(\vec{x})$ on the physical space, i.e., by $\vec{f} = -\nabla V(\vec{x})$.[2] In Cartesian coordinates (x, y, z) the gradient is given in the usual notation by

$$\nabla V = \frac{\partial V}{\partial x}\vec{i} + \frac{\partial V}{\partial y}\vec{j} + \frac{\partial V}{\partial z}\vec{k}. \tag{27.1}$$

The scalar function V is defined to be the *potential energy* of the system, up to an arbitrary additive constant. Here the potential energy is assumed to be a function of $\vec{x}$ and is not dependent on the velocity $\vec{v} := d\vec{x}/dt$. Lagrange's idea is that a function $\mathcal{L}$ defined to be the difference between the kinetic energy K and the potential energy V can determine the motion of the particle.[3] Rewriting the Cartesian coordinates as x_j we can express this new quantity as a function of x_j and $\dot{x}_j := dx_j/dt$, i.e.,

$$\mathcal{L} = \mathcal{L}(x_i, \dot{x}_i) := K - V. \tag{27.2}$$

One can verify that the following equations[4]:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}(x_i, \dot{x}_i)}{\partial \dot{x}_j}\right) = \frac{\partial \mathcal{L}(x_i, \dot{x}_i)}{\partial x_j}, \tag{27.3}$$

can serve as a set of equations of motion for the particle as they are equivalent to Newton's second law. This function is known as the *Lagrangian* and Eqs. (27.3) are known as the *Lagrange's equations*. Lagrange's theory is not suitable for quantisation since it based on

[1] Wan pp. 63–78.
[2] Note that a force means a force field defined on the physical space.
[3] In contrast it is the function H which is the sum of K and V in the Hamiltonian formulation which determines the particle's motion.
[4] Taylor pp. 238–245.

velocity variables. We need to eliminate these velocity variables. This is achieved by introducing the concept of canonical momentum. The **canonical momentum** p_{cj} conjugate to coordinate x_j is defined by

$$p_{cj} := \frac{\partial \mathcal{L}(x_i, \dot{x}_i)}{\partial \dot{x}_j}. \tag{27.4}$$

The subscript c distinguishes a canonical momentum from the linear momentum $\vec{p}_l := md\vec{x}/dt$.[5] The six conjugate variables, e.g., the position variables x_j and their canonical momentum variables p_{cj} are called **canonical variables**.

Next we need to replace Lagrange's equations with equations of motions involving the canonical variables. This is done by introducing a new quantity, known as the **Hamiltonian**, which is defined in terms of the canonical variables by

$$H := \sum_j p_{cj}\dot{x}_j - \mathcal{L}(x_i, \dot{x}_i). \tag{27.5}$$

When the velocity variables $\dot{x}_i$ are eliminated in terms x_i and p_{ci} the resulting expression is a function of x_i and p_{ci}. The Hamiltonian will be regarded as a function of x_i and p_{ci} from now on, i.e.,

$$H = H(x_i, p_{ci}). \tag{27.6}$$

Unlike the Lagrangian this new quantity has an obvious physical meaning as the total energy of the system since we can verify that[6]

$$H(x_i, p_{ci}) = K + V. \tag{27.7}$$

The equations of motion of the system can now be written in terms of the Hamiltonian, i.e., the second order Lagrange's equations of motion in Eq. (27.3) are equivalent to the following first-order equations[7]:

$$\frac{dx_j}{dt} = \frac{\partial H(x_i, p_{ci})}{\partial p_{cj}}, \quad \frac{dp_{cj}}{dt} = -\frac{\partial H(x_i, p_{ci})}{\partial x_j}. \tag{27.8}$$

[5] *See* Eq. (2.2). The subscript l highlights the difference between linear and canonical momentum variables. As shown in Eq. (27.44), the linear momentum is different from the canonical momentum, i.e., $p_{lj} \neq p_{cj}$ for some systems.

[6] Dicke and Wittke pp. 83–84.

[7] Taylor pp. 521–535. The Hamiltonian must be treated as a function of x_j and p_{cj}.

These are known as **Hamilton's equations** of motion. In Cartesian coordinates we have the following results:

(1) The canonical momentum coincides with the linear momentum since the potential energy is assumed to be velocity-independent, i.e., we have

$$p_{cj} = m\dot{x}_j = p_{lj} \quad \text{or} \quad \vec{p}_c = m\vec{v} = \vec{p}_l. \tag{27.9}$$

We shall denote both the canonical momentum and the linear momentum by $\vec{p} = (p_x,\ p_y,\ p_z)$ in such cases.[8]

(2) The Hamiltonian of the system is equal to

$$H(x,\ p) = \frac{1}{2m}\,\vec{p}^{\,2} + V(\vec{x}\,) = \frac{1}{2m}\left(p_x^2 + p_y^2 + p_z^2\right) + V(\vec{x}\,). \tag{27.10}$$

(3) An example which we shall exploit repeatedly is that of a harmonic oscillator. This is a particle of mass m in one-dimensional motion along the x-axis under a harmonic force with a Hamiltonian given by[9]

$$H_{ho}(x,\ p) = \frac{1}{2m}\,p^2 + \frac{1}{2}\,m\omega^2 x^2, \tag{27.11}$$

where ω is the angular frequency of the oscillatory motion of the particle. The Hamilton's equations are

$$\frac{dx}{dt} = \frac{p}{m}, \qquad \frac{dp}{dt} = -m\omega^2\,x. \tag{27.12}$$

The solution which describes a harmonic motion of the particle released from an initial position $x(0)$ with zero initial momentum is

$$x(t) = x(0)\cos\omega t, \qquad p(t) = -m\omega x(0)\sin\omega t, \tag{27.13}$$

where $x(0)$ is the initial position corresponding to the initial momentum $p(0) = 0$.

[8] This is not the case when the potential energy term is velocity-dependent. We shall discuss such a situation in §27.1.2 (*see* Eq. (27.44)).

[9] As customary p stands for p_x one-dimensional motion.

In Lagrange's approach it is not necessary to specify the position by Cartesian coordinates x_i. Other coordinate systems can also be used to fix the position of a particle. These are known as **generalised coordinates**. Generalised coordinates are variables q_i which can determine the position of the particle, i.e., we have

$$x_i = x_i(q_1, q_2, q_3) \quad \text{and} \quad q_i = q_i(x_1, x_2, x_3). \tag{27.14}$$

The Lagrangian in Eq. (27.2) becomes a function of q_j and $\dot{q}_j$, i.e.,

$$\mathcal{L} = \mathcal{L}(q_i, \dot{q}_i), \quad \dot{q}_i := \frac{dq_i}{dt}. \tag{27.15}$$

Lagrange's equation becomes[10]

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}(q_i, \dot{q}_i)}{\partial \dot{q}_j}\right) = \frac{\partial \mathcal{L}(q_i, \dot{q}_i)}{\partial q_j}, \tag{27.16}$$

which serves as a set of equations of motion for the particle.

We can also define canonical momentum and the Hamiltonian in generalised coordinates:

(1) The *canonical momentum* p_{cj} conjugate to each generalised coordinate q_j is defined by

$$p_{cj} := \frac{\partial \mathcal{L}(q_i, \dot{q}_i)}{\partial \dot{q}_j}. \tag{27.17}$$

The six conjugate variables, e.g., q_j and p_{cj} are also referred as *canonical variables*.

(2) The Hamiltonian is defined in terms of these new canonical variables by

$$H(q_i, p_{ci}) := \sum_j p_{cj}\dot{q}_j - \mathcal{L}(q_i, \dot{q}_i) \tag{27.18}$$

where the variables $\dot{q}_j$ must be eliminated in terms q_i and p_{ci}. The Hamiltonian is regarded as a function of q_j and p_{cj}, i.e.,

$$H = H(q_i, p_{ci}). \tag{27.19}$$

[10] Taylor pp. 238–245.

The equations of motion of the system can now be written in terms of the Hamiltonian, i.e., we have[11]

$$\frac{dq_j}{dt} = \frac{\partial H(q_i,\, p_{ci})}{\partial p_{cj}}, \qquad \frac{dp_{cj}}{dt} = -\frac{\partial H(q_i,\, p_{ci})}{\partial q_j}. \tag{27.20}$$

These are Hamilton's equations in generalised coordinates.

The relations between generalised coordinates and their canonically conjugate momenta can be quite different from that of Cartesian coordinates. An example is the spherical coordinate system with spherical coordinates (r, θ, φ) which are related to the Cartesian coordinates (x, y, z) by Eq. (16.41).[12]

The kinetic energy in spherical coordinates are[13]

$$K = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta\, \dot{\varphi}^2\right). \tag{27.21}$$

The conjugate canonical momenta to r, θ and φ are given in accordance with Eq. (27.17) by[14]

$$p_{cr} = m\dot{r}, \qquad p_{c\theta} = mr^2\dot{\theta}, \qquad p_{c\varphi} = mr^2 \sin^2\theta\, \dot{\varphi}. \tag{27.22}$$

In terms of canonical momenta the kinetic energy becomes

$$K = \frac{1}{2m}\, p_{cr}^2 + \frac{1}{2mr^2}\left(p_{c\theta}^2 + \frac{1}{\sin^2\theta}\, p_{c\varphi}^2\right). \tag{27.23}$$

A particle has a **canonical angular momentum** $\vec{L}_c$ defined in terms of the canonical momentum $\vec{p}_c$ by[15]

$$\vec{L}_c := \vec{x} \times \vec{p}_c. \tag{27.24}$$

The components and the square of $\vec{L}_c$ can be expressed in terms of spherical coordinates and their conjugate canonical momenta, e.g.,

[11] Taylor pp. 521–535.

[12] Spherical coordinates are not well-defined everywhere, e.g., at $r = 0$. But they work well in practical applications. Polar coordinates in the x-y plane have similar problems some of which are discussed in §16.1.2.

[13] As before an overhead dot indicate a time derivative, e.g., $\dot{r} := dr/dt$.

[14] Goldstein p. 299. Wan p. 17. $\widehat{p}_{cr}$ is known as the *radial* momentum.

[15] The components of $\vec{L}_c$ are similarly given by Eq. (2.4) to (2.6) for the kinematic angular momentum.

the square of the total canonical angular momentum in spherical coordinates is[16]

$$L_c^2 = p_{c\theta}^2 + \frac{1}{\sin^2\theta}\, p_{c\varphi}^2. \tag{27.25}$$

The expression for the kinetic energy reduces to

$$K = \frac{1}{2m}\, p_{cr}^2 + \frac{1}{2mr^2}\, L_c^2. \tag{27.26}$$

The Hamiltonian defined by Eq. (27.18) is

$$H = p_{cr}\dot{r} + p_{c\theta}\dot{\theta} + p_{c\varphi}\dot{\varphi} - \mathcal{L}(q_i, \dot{q}_i). \tag{27.27}$$

This is equal to the total energy. In terms of canonical variables, we have

$$H = \frac{1}{2m}\, p_{cr}^2 + \frac{1}{2mr^2}\left(p_{c\theta}^2 + \frac{1}{\sin^2\theta}\, p_{c\varphi}^2\right) + V(r, \theta, \varphi). \tag{27.28}$$

Similarly in cylindrical coordinates (r, θ, z) we have

$$K = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2\right), \tag{27.29}$$

$$p_{cr} = m\dot{r}, \quad p_{c\theta} = mr^2\dot{\theta}, \quad p_{cz} = m\dot{z}, \tag{27.30}$$

$$H = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2\right) + V(r, \theta, \varphi). \tag{27.31}$$

27.1.2 Charged Particle in Magnetic Field

Hamilton's formulation can be generalised to certain non-conservative systems. An important example, which we shall encounter in Chapter 37, is the motion of a charged particle in a time-independent magnetic field $\vec{B}(\vec{x})$.[17] Such a magnetic field is derivable as the curl of a vector potential $\vec{A}(\vec{x})$, i.e.,

$$\vec{B}(\vec{x}) = \nabla \times \vec{A}(\vec{x}), \tag{27.32}$$

[16] The square of $\vec{L}_c$, denoted by $\vec{L}_c^2$ or simply by L_c^2, is defined by $\vec{L}_c \cdot \vec{L}_c$.
[17] Dicke and Wittke pp. 80–81. Greiner pp. 150–151. Goldstein pp. 19–21, pp. 48–49.

where the curl operation in Cartesian coordinates is given by

$$\nabla \times \vec{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\vec{i} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\vec{j} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\vec{k}. \tag{27.33}$$

The motion of the particle having a charge q is determined by the Lorentz force $\vec{f} = q\,\vec{v} \times \vec{B}$. In Newton's theory its equation of motion is

$$m\frac{d^2\vec{x}}{dt^2} = q\,\vec{v} \times \vec{B}. \tag{27.34}$$

This equation of motion can be derived from the following Lagrangian[18]:

$$\mathcal{L}(x_i, \dot{x}_i) = K + q\,\vec{v} \cdot \vec{A}. \tag{27.35}$$

In other words, the resulting Lagrange's equations would agree with Newton's equation of motion in Eq. (27.34). To show this, we first calculate the various terms in the Lagrange's equations in Cartesian coordinates:

$$\frac{\partial \mathcal{L}(x_i, \dot{x}_i)}{\partial x_i} = q \sum_{j=1}^{3} \dot{x}_j \frac{\partial A_j}{\partial x_i}, \tag{27.36}$$

$$\frac{\partial \mathcal{L}(x_i, \dot{x}_i)}{\partial \dot{x}_i} = m\dot{x}_i + qA_i, \tag{27.37}$$

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}(x_i, \dot{x}_i)}{\partial \dot{x}_i}\right) = m\frac{d^2x_i}{dt^2} + q\frac{d}{dt}A_i \tag{27.38}$$

$$= m\frac{d^2x_i}{dt^2} + q\sum_{j=1}^{3}\frac{\partial A_i}{\partial x_j}\dot{x}_j. \tag{27.39}$$

Equation (27.16) of Lagrange becomes

$$m\frac{d^2x_i}{dt^2} + q\sum_{j=1}^{3}\frac{\partial A_i}{\partial x_j}\dot{x}_j = q\sum_{j=1}^{3}\dot{x}_j\frac{\partial A_j}{\partial x_i}, \tag{27.40}$$

[18] Taylor pp. 273–275. Here $K = \frac{1}{2}\sum_j m\dot{x}_j^2$.

or

$$m\frac{d^2x_i}{dt^2} = q\sum_{j=1}^{3}\dot{x}_j\left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j}\right). \tag{27.41}$$

These equations agree with Newton's Eq. (27.34). As an illustration consider the equation for x_1. Rewriting x_j as x, y, z we get

$$m\frac{d^2x}{dt^2} = q\,\dot{y}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) - q\dot{z}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right). \tag{27.42}$$

This equation reduces to

$$m\frac{d^2x}{dt^2} = q\left(\dot{y}B_z - \dot{z}B_y\right), \tag{27.43}$$

which is the x-component of Newton's Eq. (27.34).

While the Lagrangian in Eq. (27.35) appears to be in the standard form of a difference of two terms the physical interpretation is different here. The second term $q\,\vec{v}\cdot\vec{A}$ is velocity-dependent and is not the potential energy of the particle. Moreover, the Lorentz force is not obtainable as the negative of the gradient of this term.[19] The reason for introducing this Lagrange's theory is to establish a Hamiltonian formulation for the motion of a charge particle in an external magnetic field which can then be quantised.

To establish Hamilton's formulation, we first define the canonical momenta as before, i.e., we have[20]

$$p_{cj} := \frac{\partial\mathcal{L}(x_i, \dot{x}_i)}{\partial\dot{x}_j} = m\dot{x}_j + q\,A_j, \tag{27.44}$$

or in vector notation,

$$\vec{p}_c = m\vec{v} + q\,\vec{A} = \vec{p}_l + q\,\vec{A}. \tag{27.45}$$

The canonical momentum $\vec{p}_c$ is seen to have a *magnetic component*. It is fundamentally different from the linear momentum $\vec{p}_l$. While the linear momentum is a kinematic quantity fixed by the motion,

[19] Taylor p. 274.

[20] Explicitly we have $p_{cx} = m\dot{x} + q\,A_x$, $p_{cy} = m\dot{y} + q\,A_y$, $p_{cz} = m\dot{z} + q\,A_z$.

i.e., velocity, the canonical momentum is a dynamical quantity not determined by the motion alone.[21] It is not even unique, due to the non-uniqueness of the vector potential.[22]

The Hamiltonian defined by Eq. (27.18) becomes

$$H = \frac{1}{2m}\left(\vec{p}_c - q\vec{A}\right)^2. \tag{27.47}$$

This is equal to the kinetic energy of the particle, i.e.,

$$H = \frac{1}{2}m\vec{v}^{\,2}. \tag{27.48}$$

In the form of Eq. (27.47), the Hamiltonian can be quantised.[23]

In the presence of a magnetic field the canonical angular momentum $\vec{L}_c$ of the particle is again defined by Eq. (27.24). This is different from the kinematic angular momentum in Eq. (2.3). To emphasise the difference, we shall denote the kinematic angular momentum by $\vec{L}_k$. Explicitly we have

$$\vec{L}_c = \vec{x} \times \vec{p}_c \quad \text{and} \quad \vec{L}_k = \vec{x} \times \left(\vec{p}_c - q\vec{A}\right). \tag{27.49}$$

27.1.3 Poisson Bracket and Structure of Classical Observables

In classical mechanics there is no need to set up separate equations of motion for observables since the time dependence of the state automatically leads to the time dependence of observables which are just functions of the state.[24] Still we can establish an explicit equation of motion of an arbitrary observable $A(x_j,\ p_{cj})$ which is not

[21] *See* §2.1.4. Feynman, Leighton and Sands (pp. 21–25) call $\vec{p}_l$ the *kinematic momentum* and p_c the *dynamical momentum*.

[22] The vector potential $\vec{A}'$ related to $\vec{A}$ through a function $f(\vec{x})$ by

$$\vec{A}' = \vec{A} + \nabla f \tag{27.46}$$

defines the same magnetic field, since the curl of a gradient of a function vanishes. The above relation is known as a *gauge transformation*.

[23] *See* §27.6.

[24] Observables in classical mechanics are functions of x_j and p_{cj}.

explicitly time dependent, i.e.,

$$\frac{dA}{dt} = \sum_{j=1}^{3} \left(\frac{\partial A}{\partial x_j} \frac{dx_j}{dt} + \frac{\partial A}{\partial p_{cj}} \frac{dp_{cj}}{dt} \right)$$
$$= \sum_{j=1}^{3} \left(\frac{\partial A}{\partial x_j} \frac{\partial H}{\partial p_{cj}} - \frac{\partial A}{\partial p_{cj}} \frac{\partial H}{\partial x_j} \right). \tag{27.50}$$

We can simplify the above expression by introducing a new mathematical expression known as the **Poisson bracket**. Given two observables A, B their Poisson bracket, denoted by $\{A, B\}$, is defined to be[25]

$$\{A, B\} := \sum_{j=1}^{3} \left(\frac{\partial A}{\partial x_j} \frac{\partial B}{\partial p_{cj}} - \frac{\partial A}{\partial p_{cj}} \frac{\partial B}{\partial x_{cj}} \right). \tag{27.51}$$

Equation (27.50) can then be rewritten as[26]

$$\frac{dA}{dt} = \{A, H\}. \tag{27.53}$$

In other words, the motion of an observable which is not explicitly time dependent is determined by the Poisson bracket of the observable and the Hamiltonian.

The Poisson bracket of two observables defined by Eq. (27.51) plays an important role in expressing the dynamics of observables. The characteristic properties of Poisson bracket operation of observables and canonical variables can be summarised as follows:

P27.1.3(1) For any observables A, B, C, we have

$$\{A, A\} = 0, \quad \{A, c\} = 0 \quad c \in \mathbb{R}, \tag{27.54}$$
$$\{A, B\} = -\{B, A\}, \tag{27.55}$$

[25] For observables which are differentiable with respect to x_j and p_{cj}.

[26] An observable can be explicitly time dependent. For such an observable we have $A = A(x_j, p_{cj}, t)$ and the equation of motion becomes

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A, H\}. \tag{27.52}$$

For simplicity we will assume all observables to be not explicitly time dependent in this book unless otherwise is explicitly stated.

$$\{A,\, B + C\} = \{A,\, B\} + \{A,\, C\}, \tag{27.56}$$

$$\{A + B,\, C\} = \{A,\, C\} + \{B,\, C\}, \tag{27.57}$$

$$\{A,\, BC\} = \{A,\, B\}C + B\{A,\, C\}. \tag{27.58}$$

$$\{AB,\, C\} = \{A,\, C\}B + A\{B,\, C\}. \tag{27.59}$$

P27.1.3(2) For the canonical variables x_j and p_{cj} we have

$$\{x_i,\, x_j\} = 0, \quad \{p_{ci},\, p_{cj}\} = 0, \quad \{x_i,\, p_{cj}\} = \delta_{ij}. \tag{27.60}$$

These are characteristic of canonical variables. For example, the Poisson bracket of two canonical momenta is zero. The three components of the canonical angular momentum in Eq. (27.24) do not satisfy this condition. They satisfy the following relations[27]:

$$\{L_{cx},\, L_{cy}\} = L_{cz},\ \{L_{cz},\, L_{cx}\} = L_{cy},\ \{L_{cy},\, L_{cz}\} = L_{cx}. \tag{27.61}$$

Hence they are not regarded as canonical momenta.[28]

27.2 Postulate (CQ)

Our aim here is to establish a quantisation process to obtain operators to represent quantum observables. The Poisson bracket of two classical observables defined by Eq. (27.51) is seen to play an important role in expressing the dynamics of observables. The characteristic properties of Poisson bracket operation are summarised in Eqs. (27.54) to (27.59). These are fundamental structural properties of classical observables.

The commutator operation for operators in a Hilbert space possess similar properties. Given three operators $\widehat{A}$, $\widehat{B}$ and $\widehat{C}$ we have the following formal commutation relations:

$$[\widehat{A},\, \widehat{A}\,] = \widehat{0}, \quad [\widehat{A},\, c\,] = \widehat{0}, \quad c \in \mathbb{C}, \tag{27.62}$$

$$[\widehat{A},\, \widehat{B}\,] = -[\widehat{B},\, \widehat{A}\,], \tag{27.63}$$

$$[\widehat{A},\, \widehat{B} + \widehat{C}\,] = [\widehat{A},\, \widehat{B}\,] + [\widehat{A},\, \widehat{C}\,], \tag{27.64}$$

$$[\widehat{A},\, \widehat{B}\widehat{C}\,] = [\widehat{A},\, \widehat{B}]\widehat{C} + \widehat{B}[\widehat{A},\, \widehat{C}\,]. \tag{27.65}$$

[27] *See* Eqs. (2.3) to (2.7) for the expressions for the components of the angular momentum.

[28] Goldstein pp. 265–266.

We also have

$$[\widehat{A}, c_1\widehat{B} + c_2\widehat{C}\,] = c_1[\widehat{A}, \widehat{B}] + c_2[\widehat{A}, \widehat{C}\,], \quad c_1, c_2 \in \mathbb{C}, \tag{27.66}$$

$$[c_1\widehat{A} + c_2\widehat{B}, \widehat{C}\,] = c_1[\widehat{A}, \widehat{C}] + c_2[\widehat{B}, \widehat{C}\,], \quad c_1, c_2 \in \mathbb{C}, \tag{27.67}$$

$$[\widehat{A}\widehat{B}, \widehat{C}\,] = \widehat{A}[\widehat{B}, \widehat{C}\,] + [\widehat{A}, \widehat{C}\,]\widehat{B}. \tag{27.68}$$

It is then reasonable to hypothesise that operators representing quantum observables should have the same structural properties of classical observables. In other words, the commutator operation of operators representing quantum observables should correspond to the classical Poisson bracket operation of their associated classical observables. In particular

> *selfadjoint operators representing classical canonical variables should have commutation relations which would correspond to the Poisson bracket relations of the classical canonical variables.*

In Cartesian coordinates the canonical variables x_i and p_{ci} satisfy the Poisson bracket relations in Eq. (27.60). are characteristic of canonical variables. The *canonical quantisation scheme* is based on the idea that the Poisson bracket relations of the canonical variables in Eq. (27.60) should be carried over to quantum theory.[29] Hence the transition from a classical particle with canonical variables x_i and p_{cj} to a quantum particle with a Hilbert space as its state space can be achieved in accordance with the following postulate:

Postulate 27.2(CQ) *In quantum theory both x_i and p_{cj} are represented by selfadjoint operators $\widehat{Q}_i$ and $\widehat{P}_j$ in the state space of the quantum particle satisfying the following commutation relations*[30]:

$$[\widehat{\mathsf{Q}}_i, \widehat{\mathsf{Q}}_j] = \widehat{0}, \quad [\widehat{\mathsf{P}}_i, \widehat{\mathsf{P}}_j] = \widehat{0}, \quad [\widehat{\mathsf{Q}}_i, \widehat{\mathsf{P}}_j] = i\hbar\delta_{ij}\widehat{\mathbb{I}}. \tag{27.69}$$

This quantisation scheme turns out to be surprisingly powerful even though it does not tell us explicitly what state space and what

[29] Dirac pp. 84–89. This idea is generally attributed to Dirac, who presented it as *quantum conditions*. Dirac even called a commutator a *quantum Poisson bracket*.

[30] The commutation relations are subject to qualifications discussed in §17.7 for unbounded operators.

operators to choose for a given physical system. We would have a choice, subject to the commutation relations. These commutation relations are known as **canonical commutation relations**. Note that the imaginary unit i is necessary in the commutation relation of $\widehat{\mathsf{Q}}_i$ and $\widehat{\mathsf{P}}_j$ since

$$[\widehat{\mathsf{Q}}_i, \widehat{\mathsf{P}}_j]^\dagger = -[\widehat{\mathsf{Q}}_i, \widehat{\mathsf{P}}_j]. \tag{27.70}$$

To carry out the quantisation scheme for a particle in three-dimensional motion in the physical space we live in, we proceed with the following canonical quantisation procedure:

P27.2(1) Choose $\vec{L}^2(I\!R^3)$ as the state space for the system. This space corresponds to functions $\phi(x, y, z)$ on the physical space which are square-integrable with respect to x, y and z. The space $\vec{L}^2(I\!R^3)$ is referred to as a *coordinate representation space*.[31]

P27.2(2) Choose six selfadjoint operators in $\vec{L}^2(I\!R^3)$ which satisfy the canonical commutation relations. A familiar choice is

$$\widehat{x}(I\!R^3),\ \widehat{y}(I\!R^3),\ \widehat{z}(I\!R^3) \quad \text{for } \widehat{\mathsf{Q}}_i, \tag{27.71}$$

$$\widehat{p}_x(I\!R^3),\ \widehat{p}_y(I\!R^3),\ \widehat{p}_z(I\!R^3) \quad \text{for } \widehat{\mathsf{P}}_j. \tag{27.72}$$

These are the selfadjoint position and momentum operators in Eqs. (17.20), and (17.54).[32] In line with the notation in §20.4 we can conveniently denote these operators by

$$\widehat{\vec{x}} = \big(\widehat{x},\ \widehat{y},\ \widehat{z}\big) \quad \text{and} \quad \widehat{\vec{p}}_c = \big(\widehat{p}_{cx},\ \widehat{p}_{cy},\ \widehat{p}_{cy}\big). \tag{27.73}$$

The subscript c in $\widehat{\vec{p}}_c$ indicates quantised canonical momentum variables.[33] These operators satisfy the canonical commutation relations. The subscript c is often omitted when no confusion arises, e.g., writing $\widehat{p}_{cx}$ as $\widehat{p}_x$. Since all the operators are unbounded, their commutation relations cannot produce the equalities shown in Eq. (27.69). This problem has already been pointed out in relation to

[31] This term is first introduced in §18.4.2.1.
[32] *See* also Eqs. (24.65) and (24.66).
[33] *See* Eq. (27.49) and Eqs. (27.84) to (27.99) for the reason of this notation.

Eq. (17.81). It follows that we must relax the equalities demanded in the canonical commutation relations shown in Eq. (27.69) and interpret the commutation relations in the same way we do for Eq. (17.81), i.e., the equalities hold in a dense subset such as $\vec{C}_c^{\infty}(I\!R^3)$ or $\vec{S}_s(I\!R^3)$ of $\vec{L}^2(I\!R^3)$. Following convention, we shall continue to use the equality sign in the canonical commutation relations and other commutation relations resulting from them. With this understanding the canonical quantisation scheme can be realised. We call the position and momentum operators in Eq. (27.72) a *coordinate representation* of the canonical commutation relations in the coordinate representation space $\vec{L}^2(I\!R^3)$.

P27.2(3) To extend to other observables, three more quantisation rules are made[34]:

(1) **Linearity rule** If A and B are quantised as selfadjoint operators $\widehat{A}$ and $\widehat{B}$, then the classical observable $aA + bB$ is quantised as the operator $a\widehat{A} + b\widehat{B}$.

(2) **Functional relationship rule** Let $f(\tau)$ be a real-valued function of a real variable $\tau \in I\!R$. If A is quantised as selfadjoint operator $\widehat{A}$ then observable $f(A)$ is quantised as the operator $f(\widehat{A})$.[35]

(3) **Symmetrisation rule**[36] If A and B are quantised, respectively, as selfadjoint operators $\widehat{A}$ and $\widehat{B}$ then AB is quantised as the operator

$$\big(\widehat{A}\widehat{B} + \widehat{A}\widehat{B}\big)/2. \tag{27.74}$$

This rule is used to overcome the ambiguity arising from quantising a product, i.e., classically $AB = BA$ but $\widehat{A}\widehat{B} \neq \widehat{A}\widehat{B}$ unless $\widehat{A}$, $\widehat{B}$ commute. This is like quantising the symmetrised classical observable $\big(AB + AB\big)/2$.

P27.2(4) Care has to be taken when using rules presented above. For example, $a\widehat{A} + b\widehat{B}$ may only be symmetric and not selfadjoint since generally $\big(a\widehat{A} + b\widehat{B}\big)^{\dagger} \supset \big(a^*\widehat{A}^{\dagger} + b^*\widehat{B}^{\dagger}\big)$ by Eq. (17.101) when

[34] Isham p. 77. Wan pp. 256–259.
[35] *See* E27.10.2(2) for a generalisation of this rule.
[36] Isham pp. 78–81.

$\widehat{A}$ and $\widehat{B}$ are unbounded. The way to proceed is as follow:

(1) First check whether Eq. (17.101) becomes an equality for the given $\widehat{A}$ and $\widehat{B}$. Fortunately this is the case for many applications, e.g., those in Eqs. (27.82) and (35.9).

(2) Generally it is not obvious whether Eq. (17.101) in an equality. It is easier to start by considering the restriction of $(a\widehat{A}+b\widehat{B})$ to an appropriate dense set in $\vec{L}^2(I\!R^3)$.[37] A first attempt would be to choose $\vec{C}_c^\infty(I\!R^3)$. This set lies within the domain of any powers of the momentum and position operators such as $\widehat{p}_x^m$ and $\widehat{x}^n$ and the set is also invariant under these operators.[38] However, the set $\vec{C}_c^\infty(I\!R^3)$ is very restrictive as it excludes functions such as Hermite functions which are not of compact support. A more natural choice in many cases is the Schwartz space $\vec{S}_s(I\!R^3)$ which is also invariant under $\widehat{p}_x^m$ and $\widehat{x}^n$. The restriction to the Schwartz space is denoted by[39]

$$(a\widehat{A}+b\widehat{B})_{\vec{S}_s} := a\widehat{A}_{\vec{S}_s}+b\widehat{B}_{\vec{S}_s}. \tag{27.75}$$

(3) We want an appropriate selfadjoint extension of the above operator to serve as the quantised operator for the classical observable $aA + bB$, $a, b \in I\!R$. There are three cases to consider[40]:

   (a) Selfadjoint extension may not exist. Then there is no quantised operator to correspond to the particular classical observable.

   (b) When the operator in Eq. (27.75) is essentially selfadjoint it would have a unique selfadjoint extension.[41] This unique selfadjoint extension is taken to be the quantised operator.

   (c) The operator in Eq. (27.75) is not essentially selfadjoint, but it admits many selfadjoint extensions. Then a suitable selfadjoint extension has to be chosen on physical ground to represent the quantised observable.

---

[37] *See* Definition 17.6(1).
[38] *See* Definition 17.6(1). Here $\widehat{p}_x$ is the operator $\widehat{p}_{cx}$ in Eq. (27.73).
[39] For operators in $\vec{L}^2(I\!R)$ the restriction would be to the Schwartz space $\vec{S}_s(I\!R)$.
[40] We would examine many examples of these in the following sections. Hall pp. 190–196 is devoted to the study of sums of selfadjoint operators.
[41] *See* Definition 19.5(1).

27.3 Canonical Quantisation Scheme

Postulate 27.2(CQ) supplemented by the rules in P27.2(3) is known as a **canonical quantisation scheme**. Despite the successes the canonical quantisation scheme encounters many difficulties. We shall discuss some of the limitations of the scheme in §27.10. Here we shall examine one obvious problem, i.e., the canonical quantisation scheme does not produce a unique set of operators even after the state space has been chosen. Consider the case of a particle in one-dimensional motion along the x-axis. The position and momentum operators must obey the canonical commutation relation $[\widehat{\mathsf{Q}}, \widehat{\mathsf{P}}] = i\hbar$. In the coordinate representation in $\vec{L}^2(I\!R)$ the position and its conjugate canonical momentum can be quantised as the operators $\widehat{x} := \widehat{x}(I\!R)$, and $\widehat{p} := \widehat{p}(I\!R)$ which are defined, respectively, by Eqs. (17.13) and (17.50). Clearly the operators $\widehat{x}$ and $\widehat{p} + f(x)$ also satisfy the canonical commutation relations so that the position and momentum can be equally quantised as the operators $\widehat{x}$ and $\widehat{p} + f(x)$.[42] The crucial question is whether these different representations are physically equivalent in the sense of Chapter 23. The answer in this particular case is given by a theorem of von Neumann which says that any two representations of the above canonical commutation relations in $\vec{L}^2(I\!R)$ are unitarily related.[43] In other words, if two pairs of selfadjoint operators in $\vec{L}^2(I\!R)$ satisfy the above commutation relations, i.e., if[44]

$$[\widehat{\mathsf{Q}}, \widehat{\mathsf{P}}] = i\hbar, \quad \text{and} \quad [\widehat{\mathsf{Q}}', \widehat{\mathsf{P}}'] = i\hbar, \tag{27.76}$$

then there exists a unitary operator $\widehat{U}$ on $\vec{L}^2(I\!R)$ such that

$$\widehat{\mathsf{Q}}' = \widehat{U}\,\widehat{\mathsf{Q}}\widehat{U}^\dagger \quad \text{and} \quad \widehat{\mathsf{P}}' = \widehat{U}\,\widehat{\mathsf{P}}\widehat{U}^\dagger. \tag{27.77}$$

An example is the pairs mentioned above, i.e.,

$$\widehat{\mathsf{Q}} = \widehat{x}, \quad \widehat{\mathsf{P}} = \widehat{p} \quad \text{and} \quad \widehat{\mathsf{Q}}' = \widehat{x}, \quad \widehat{\mathsf{P}}' = \widehat{p} + f(\widehat{x}). \tag{27.78}$$

[42] Here $f(x)$ is a real-valued function of x, and $\widehat{p}$ stands for $\widehat{p}_c$.

[43] von Neumann (1903–1957) was an Hungarian American mathematical physicist well known for his mathematical formulation of quantum mechanics. A postulate on measurement, the *projection postulate*, is named after him (*see* §30.1.1).

[44] Jauch pp. 200–201. Reed and Simon Vol. 1 pp. 274–275. Prugovečki p. 342. As shown in §27.10.2, the von Neumann theorem does not apply to other choices of Hilbert spaces, e.g., $\vec{L}^2(\Lambda)$.

We can verify that they are related with the unitary operator[45]

$$\widehat{U} := \exp\left\{-\dot{\imath}\int^{x} f(x)dx\right\}, \quad \dot{\imath} = i/\hbar. \tag{27.79}$$

As an illustration consider the case $f(x) = x$. Letting $f(x) = x$ in the above equation we get

$$\widehat{U} := \exp\left\{-\dot{\imath}\int_0^{x} x dx\right\} = e^{-\dot{\imath}x^2/2}. \tag{27.80}$$

We can verify by explicit calculation that

$$\widehat{U}\,\widehat{p}\,\widehat{U}^{\dagger} = \widehat{p} + \widehat{x}. \tag{27.81}$$

The unitary relationship ensures that $\widehat{p}+\widehat{x}$ are selfadjoint since $\widehat{p}$ is selfadjoint.[46]

It is clear now that there are many different representations of the canonical commutation relations. For example, we have a representation arises from the Fourier transformation from the *coordinate representation space* $\vec{L}^2(I\!R)$ to the *momentum representation space* $\vec{L}^2(\underset{\sim}{I\!R})$ described in §18.4.2.1 and §18.4.2.2.

In the momentum representation space the position and momentum are described by the operators $\widehat{\underset{\sim}{x}}$ and $\widehat{\underset{\sim}{p}}$ in $\vec{L}^2(\underset{\sim}{I\!R})$ in Eqs. (18.57) and (18.60). The operators $\widehat{\underset{\sim}{x}}$ and $\widehat{\underset{\sim}{p}}$ satisfy the canonical commutation relations. All these quantities are related to the corresponding quantities in the coordinate representation by a simultaneous Fourier transformation, i.e., we have

$$\vec{\underset{\sim}{\varphi}} = \widehat{U}_F\vec{\varphi}, \quad \widehat{\underset{\sim}{x}} = \widehat{U}_F\,\widehat{x}\,\widehat{U}_F^{\dagger}, \quad \widehat{\underset{\sim}{p}} = \widehat{U}_F\,\widehat{p}\,\widehat{U}_F^{\dagger}. \tag{27.83}$$

We can conclude that the momentum representation in terms of operators $\widehat{\underset{\sim}{x}}$ and $\widehat{\underset{\sim}{p}}$ is a representation of the canonical commutation relation physically equivalent to the coordinate representation in

[45] Wan p. 270.

[46] Since

$$(\widehat{p}+\widehat{x})^{\dagger} = \widehat{p}+\widehat{x} = \widehat{p}^{\dagger}+\widehat{x}^{\dagger}, \tag{27.82}$$

we have a special case of Eq. (17.101) on the adjoint of the sum of two unbounded operators. We also have the result that the restriction of $\widehat{p}+\widehat{x}$ to $\vec{S}_s(I\!R)$ is essentially selfadjoint (*see* Hall pp. 192–193).

terms of operators $\widehat{x}$ and $\widehat{p}$. The physical equivalence of different representations tells us that in many cases

> *the physical properties of the position and momentum operators and indeed of many other operators are effectively determined by their commutation relations.*

Once the state space is chosen and position and momentum are quantised, we can proceed to establish other operators which are functions of position and momentum by the canonical quantisation scheme.[47]

27.4 Quantisation of Orbital Angular Momentum

The components L_{cx}, L_{cy}, L_{cz} of the canonical angular momentum $\vec{L}_c$ in Eq. (27.24) are given by expressions similar to that for the kinematic angular momentum in Eqs. (2.4) to (2.6). Following P27.2(3) the components of the canonical angular momentum are quantised as the following operators in $\vec{L}^2(I\!R^3)$:

$$\widehat{L}_{cx} := \widehat{y}\widehat{p}_{cz} - \widehat{z}\widehat{p}_{cy} = -i\hbar\,(y\partial/\partial z - z\partial/\partial y)\,, \tag{27.84}$$

$$\widehat{L}_{cy} := \widehat{z}\widehat{p}_{cx} - \widehat{x}\widehat{p}_{cz} = -i\hbar\,(z\partial/\partial x - x\partial/\partial z)\,, \tag{27.85}$$

$$\widehat{L}_{cz} := \widehat{x}\widehat{p}_{cy} - \widehat{y}\widehat{p}_{cx} = -i\hbar\,(x\partial/\partial y - y\partial/\partial z)\,. \tag{27.86}$$

These expressions can be conveniently written as

$$\widehat{\vec{L}}_c = \widehat{\vec{x}} \times \widehat{\vec{p}}_c. \tag{27.87}$$

The operator for the square of the total angular momentum is

$$\widehat{L}_c^2 := \widehat{L}_{cx}^2 + \widehat{L}_{cy}^2 + \widehat{L}_{cz}^2. \tag{27.88}$$

Here we have the fortunate case that all the above operators are selfadjoint.[48] It is easier to appreciate the selfadjointness of the

[47] A straightforward application of the rules in P27.2(3) will not necessarily produce selfadjoint operators, because of Eqs. (17.100) and (17.101). Fortunately in many applications we do obtain selfadjoint operators by a direct application of these rules. See Chapter 35 for more examples.

[48] Takhtajan p. 182.

operators in spherical coordinates. Take the examples of $\widehat{L}_{cz}$ and $\widehat{L}_c^2$. When expressed in terms of spherical coordinates these two operators have the following familiar expressions:

$$-i\hbar\frac{\partial}{\partial\varphi}, \qquad -\hbar^2\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right). \tag{27.89}$$

The radial variable r does not appear. The above operator expressions show that $\widehat{L}_{cz}$ and $\widehat{L}_c^2$ are related to two operators in the space $\vec{L}^2(S_u)$[49]:

(1) $\widehat{L}_{cz}$ is related to $\widehat{L}_z(\mathcal{S}_u)$ in Eq. (17.42). More explicitly $\widehat{L}_{cz}$ in Eq. (27.86) is the same as $\widehat{L}_z(I\!R^3, dxdydz)$ in Eq. (24.72) which is selfadjoint since $\widehat{L}_z(\mathcal{S}_u)$ is selfadjoint.[50]

(2) Similarly $\widehat{L}_c^2$ in Eq. (27.88) is identifiable with $\widehat{L}^2(I\!R^3, dxdydz)$ in Eq. (24.73) and this operator is also selfadjoint.[51]

It can be shown that the restrictions of $\widehat{L}_{cx}$, $\widehat{L}_{cy}$, $\widehat{L}_{cz}$ and $\widehat{L}_c^2$ to the domain defined by the Schwartz space $\vec{\mathcal{S}}_s(I\!R^3)$ are essentially selfadjoint.[52] It follows that $\widehat{L}_{cx}$, $\widehat{L}_{cy}$, $\widehat{L}_{cz}$ and $\widehat{L}_c^2$ are their unique selfadjoint extensions. These selfadjoint operators can then be taken to be the quantised angular momentum operators without any ambiguity.

In the absence of external magnetic field, the classical canonical momentum $\vec{p}_c$ and canonical angular momentum $\vec{L}_c$ agree with the classical linear (kinematic) momentum $\vec{p}_l$ and kinematic angular momentum $\vec{L}_k$. We simply call both $\vec{p}_l$ and $\vec{p}_c$ as the momentum to be denoted by $\vec{p}$, and call both $\vec{L}_c$ and $\vec{L}_k$ as the orbital angular momentum to be denoted by $\vec{L}$. The quantised operators $\widehat{p}_{cx}$, $\widehat{p}_{cy}$, $\widehat{p}_{cz}$, $\widehat{L}_{cx}$, $\widehat{L}_{cy}$, $\widehat{L}_{cz}$ can then be written as

$$\widehat{p}_x,\ \widehat{p}_y,\ \widehat{p}_z \quad \text{and} \quad \widehat{L}_x, \widehat{L}_y,\ \widehat{L}_z, \tag{27.90}$$

without the subscript c, as we shall do in the next section.

---

[49] Amrein, Jauch and Sinha pp. 458–459. $\vec{L}^2(S_u)$ is the space of square-integrable functions on the unit sphere introduced in §16.1.2. *See* also Eq. (19.50).

[50] *See* §19.3.2.

[51] *See* E19.4(3).

[52] Amrein, Jauch and Sinha p. 32. Blank, Exner and Havliček p. 365. Moretti p. 458. We shall not delve into the mathematical analysis here. *See* E16.1.2.3(5) for $\mathcal{S}_s(I\!R^3)$ and §16.2.3 for $\vec{\mathcal{S}}_s(I\!R^3)$.

These operators are far more complicated than momentum operators. It is not just that they appear more complicated. The crucial difference is that these operators do not commute, e.g., one can easily check that $[\widehat{L}_x, \widehat{L}_y] \neq \widehat{0}$.[53] Hence $\widehat{L}_x$, $\widehat{L}_y$ and $\widehat{L}_z$ do not share a complete orthonormal set of eigenvectors. The square of the total angular momentum operator commutes with each component, e.g., $\widehat{L}^2$ commutes with $\widehat{L}_z$. Similar results apply to $\widehat{L}^2$, $\widehat{L}_x$ and for $\widehat{L}^2$, $\widehat{L}_y$. However, the eigenvectors of $\widehat{L}^2$, $\widehat{L}_z$ are different from that of $\widehat{L}^2$, $\widehat{L}_x$ and $\widehat{L}^2$, $\widehat{L}_y$. We shall study these operators in greater details in §36.1 in Chapter 36.

27.5 Quantisation of Hamiltonians

Consider the motion of a chargeless particle of mass m in the absence of magnetic magnetic field. Classically the canonical momentum coincides with the linear momentum. We shall denote both by $\vec{p}$. The kinetic energy K is given by Eq. (27.10). It follows that the potential energy is quantised as the multiplication operator $\widehat{V} := V(\vec{x})$, and the kinetic energy is quantised as[54]

$$\widehat{K} := \frac{1}{2m}\left(\widehat{p}_x^2 + \widehat{p}_y^2 + \widehat{p}_z^2\right) = -\frac{\hbar^2}{2m}\nabla^2 \tag{27.91}$$

in $\vec{L}^2(I\!R^3)$. This operator is selfadjoint since it is a real function of a three commuting selfadjoint operators.[55] It is convenient to express $\widehat{K}$ as

$$\widehat{K} := \widehat{\vec{p}}^{\,2}/2m \qquad \text{where} \quad \widehat{\vec{p}}^{\,2} := \widehat{p}_x^2 + \widehat{p}_y^2 + \widehat{p}_z^2. \tag{27.92}$$

The Hamiltonian (total energy) operator is then given by

$$\widehat{H} := \widehat{K} + \widehat{V}. \tag{27.93}$$

This operator, often referred to as the Hamiltonian for short, is an important characterisation of a system. It is a challenging

[53] Their commutation relations are given by Eqs. (27.111) to (27.114).
[54] $\nabla^2 := \left(\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2\right)$.
[55] *See* §15.4.

mathematical problem to investigate the selfadjointness of the operator for any given potential.[56] The following cases serve to illustrate the situation:

(1) For many familiar potentials in quantum mechanics the operator $\widehat{H}$ is selfadjoint. Moreover, its restriction to Schwartz space is essentially selfadjoint. The unique selfadjoint extension coincides with $\widehat{H}$. Then we have a unique quantised Hamiltonian. The Hamiltonian for the harmonic oscillator discussed in Chapter 35 serves to illustrate this case.

(2) *Infinite square barriers and potential well* When the potential function is discontinuous and rises to infinity abruptly over extended regions. We will be confronted with many ambiguities in quantisation. Let us consider two examples:

   (a) *Infinite square potential barrier* This is the case when the potential function is given by $V(x) = \infty$ for $x \leq 0$ and $V(x) = 0$ for $x > 0$. Such a potential confines the particle to the region $x > 0$. It follows that the state space for the particle is $\vec{L}^2(I\!R^+)$. As pointed out in §19.3.3 there is no selfadjoint momentum operator in $\vec{L}^2(I\!R^+)$ since $\widehat{p}_D(I\!R^+)$ admits no selfadjoint extension in $\vec{L}^2(I\!R^+)$. This shows that not every classical observable has an obvious quantum counterpart in the orthodox quantum theory.[57]
   However, as pointed out in E19.4(4), we can have a selfadjoint Hamiltonian in the form of $\widehat{p}_D^{\dagger}(I\!R^+)\widehat{p}_D(I\!R^+)$ which is selfadjoint.[58]

   (b) *Infinite square potential well* This is the case when the potential function is given by $V(x) = 0$ for x inside a finite closed interval $\Lambda = [0, L]$ and $V(x) = \infty$ for x outside the interval and at the boundary points $x = 0$ and $x = L$. Such a potential confines the particle to the region $0 < x < L$. The state space is $\vec{L}^2(\Lambda)$. There are

[56] Hamiltonian operators are also known as Schrödinger operators in mathematics literature.

[57] The converse is also true, e.g., spin has no classical counterpart. In a *generalised quantum theory*, we can accept certain symmetric operators such as $\widehat{p}_D(I\!R^+)$ to represent observables (*see* Wan pp. 395–426).

[58] Note that this operator is not equal to $\widehat{p}_D^2(I\!R^+)$.

now many possible selfadjoint momentum operators. The reason is that the operator $\widehat{p}_D(\Lambda)$ in E17.3.2.1(2)(a) is shown in §19.2 to be only symmetric.[59] This operator is not essentially selfadjoint since it admits a one-parameter family of selfadjoint extensions $\widehat{p}_\lambda(\Lambda)$.[60] These operators possess different sets of eigenvalues and they are not unitarily related to each other.

There are many choices of Hamiltonians also. The standard choice of a particle trapped in an infinite square potential well is taken to be the operator $\widehat{H}^\infty_D(\Lambda)$ defined by Eq. (19.42). Other choices are $\widehat{H}^\infty_\lambda(\Lambda)$ given by Eq. (19.46).[61] This operator is the square of $\widehat{p}_\lambda(\Lambda)$ while $\widehat{H}^\infty_D(\Lambda)$ is not. These operators are distinguishable by their different eigenvalues.

A choice of momentum and the Hamiltonian has to be made by physical consideration of a given system.[62]

(c) Our discussion here shows that the rule on the preservation of functional relation in §27.2 should be amended when the quantised operator is not selfadjoint. When A is quantised as the operator $\widehat{A}$ the rule means that A^2 should be quantised as the operator $\widehat{A}^2$. This would produce a selfadjoint operator for A^2 if $\widehat{A}$ is selfadjoint. When $\widehat{A}$ is only symmetric and closed we should quantise $\widehat{A}^2$ as the selfadjoint operator $\widehat{A}^\dagger \widehat{A}$.

(3) *Point interactions* This is when the potential is discontinuous and rises to infinity abruptly at some isolated point. An example is a potential given by a Dirac delta function, i.e., $V(x) := \lambda\,\delta(x)$ where $\lambda \in \mathbb{R}$.[63] There are a host of other different interactions which occur at a single point which cannot be expressed

[59] The same applies to $\widehat{p}_0(\Lambda)$ introduced in E17.3.2.1(1).

[60] Each $\widehat{p}_\lambda(\Lambda)$ is clearly an extension of $\widehat{p}_D(\Lambda)$ and each $\widehat{p}_\lambda(\Lambda)$ is selfadjoint, as shown in §19.3.1. *See* Wan p. 490 for motion along a circle.

[61] Merzbacher p. 66. $\widehat{H}^\infty_{\lambda=0}(\Lambda)$ is chosen for a particle in a box problem for some applications. Schiff p. 49 has a paragraph on the physics of the periodic boundary condition.

[62] Reed and Simon Vol. 1, p. 303.

[63] Merzbacher pp. 107–108.

explicitly in terms of a continuous potential function. We call all these *point interactions*. The canonical quantisation scheme cannot be applied to such cases. A scheme of **quantisation by parts** can be formulated to deal with such interactions.[64]

It should be pointed out that the non-uniqueness of in quantising the classical Hamiltonian exists even without infinite potentials. An example is a particle in one-dimensional motion along the x-axis with potential energy $-gx^4$ where g is a positive real constant. In accordance with P27.2(4) the Hamiltonian operator is first quantised as[65]

$$\widehat{H}_{\vec{S}_s} = \frac{1}{2m}\,\widehat{p}^{\,2}_{\vec{S}_s} - g\,\widehat{x}^{\,4}_{\vec{S}_s} \tag{27.94}$$

acting on the Schwartz space $\vec{S}_s(I\!R) \subset \vec{L}^2(I\!R)$. This operator is not essentially selfadjoint.[66] It follows that we cannot have a unique selfadjoint Hamiltonian.[67]

In classical mechanics the Hamiltonian determines the time evolution of the system in the form of Hamilton's equations. The Hamiltonian operator also plays a crucial role in determining the time evolution of the quantum system, as shown in Chapter 28.

27.6 Charged Particles in Magnetic Field

For a particle of mass m and charged q in a magnetic field characterised by a vector potential $\vec{A}$, we have to distinguish the canonical momentum $\vec{p}_c$ from the linear (kinematic) momentum $\vec{p}_l$, and the canonical angular momentum from the kinematic angular momentum. The canonical momentum and the canonical angular momentum are quantised as before as $\widehat{\vec{p}}_c$ and $\widehat{\vec{L}}_c$.[68] On account

[64] Wan pp. 431–506 for detailed study of the scheme of *quantisation by parts* and classification of point interactions.

[65] $\widehat{x}_{\vec{S}_s}$, $\widehat{p}_{\vec{S}_s}$ are the restrictions of $\widehat{x}$ and $\widehat{p}$ to the Schwartz space $\vec{S}_s(I\!R)$ previously denoted by $\widehat{x}_{\vec{S}_s}(I\!R)$, $\widehat{p}_{\vec{S}_s}(I\!R)$ in E17.6(2), Eqs. (17.48) and (17.80).

[66] Hall p. 194 where the operator defined on $C_c^\infty(I\!R)$ is shown not to be essentially selfadjoint. As a consequence $\widehat{H}_{\vec{S}_s}$ is also not essentially selfadjoint.

[67] Reed and Simon Vol. 1, p. 303. We have to make a choice on physical grounds.

[68] *See* Eq. (27.87).

of Eqs. (27.45), (27.49) and in accordance with P27.2(4) the linear momentum $\vec{p}_l$ and the kinematic angular momentum $\vec{L}_k$ are quantised, respectively, as

$$\hat{\vec{p}}_l := \hat{\vec{p}}_c - q\hat{\vec{A}} \quad \text{and} \quad \hat{\vec{L}}_k := \hat{\vec{x}} \times \left(\hat{\vec{p}}_c - q\hat{\vec{A}} \right). \tag{27.95}$$

We can write down their components explicitly, e.g.,[69]

$$\hat{p}_{lx} = (\hat{p}_{cx} - q\hat{A}_x), \tag{27.96}$$
$$\hat{L}_{kz} = \hat{x}\,(\hat{p}_{cy} - q\hat{A}_y) - \hat{y}\,(\hat{p}_{cx} - q\hat{A}_x). \tag{27.97}$$

The kinematic angular momentum is dependent on the vector potential. The Hamiltonian in Eq. (27.47) is quantised as

$$\hat{H} := \frac{1}{2m}\left(\hat{\vec{p}}_c - q\hat{\vec{A}}\right)^2, \tag{27.98}$$

where $\left(\hat{\vec{p}}_c - q\hat{\vec{A}}\right)^2$ is defined by

$$\left(\hat{p}_{cx} - q\hat{A}_x\right)^2 + \left(\hat{p}_{cy} - q\hat{A}_y\right)^2 + \left(\hat{p}_{cz} - q\hat{A}_z\right)^2. \tag{27.99}$$

27.7 Manipulations of Commutation Relations

It is important to be able to evaluate commutation relations. The properties listed in Eqs. (27.62) to (27.68) are useful for this purpose. We can make use of all these properties when calculating commutation relations as seen in the following examples:

E27.7(1) When there are commuting operators involved, Eqs. (27.65) and (27.68) can be simplified, i.e., we have

$$[\hat{A}, \hat{B}] = \hat{0} \;\Rightarrow\; [\hat{A}, \hat{B}\hat{C}] = \hat{B}\,[\hat{A}, \hat{C}], \tag{27.100}$$
$$[\hat{A}, \hat{C}] = \hat{0} \;\Rightarrow\; [\hat{A}, \hat{B}\hat{C}] = [\hat{A}, \hat{B}]\,\hat{C}, \tag{27.101}$$
$$[\hat{A}, \hat{C}] = \hat{0} \;\Rightarrow\; [\hat{A}\hat{B}, \hat{C}] = \hat{A}\,[\hat{B}, \hat{C}], \tag{27.102}$$
$$[\hat{B}, \hat{C}] = \hat{0} \;\Rightarrow\; [\hat{A}\hat{B}, \hat{C}] = [\hat{A}, \hat{C}]\,\hat{B}. \tag{27.103}$$

[69] Here $\hat{\vec{A}}$ is defined by its components $\hat{A}_x$, $\hat{A}_y$ and $\hat{A}_z$ which are multiplication operators in $\vec{L}^2(I\!R^3)$.

E27.7(2) Extension of the relation $[\hat{x},\ \hat{p}\,] = i\hbar$ in $\vec{L}^2(I\!R)$ to[70]

$$[\hat{x},\ \hat{p}^{\,n}] = i\hbar\, n\, \hat{p}^{\,(n-1)}, \quad [\,\hat{p},\ \hat{x}^{\,n}] = -i\hbar\, n\, \hat{x}^{\,(n-1)}. \tag{27.104}$$

These commutation relations are well-defined on the Schwartz space $\vec{S}(I\!R)$.

E27.7(3) Let $\hat{A}(\hat{x},\ \hat{p}\,)$ be a polynomial in $\hat{x}$ and $\hat{p}$, i.e.,[71]

$$\hat{A}(\hat{x},\ \hat{p}\,) = \sum_{m,n} c_{mn}\, \hat{x}^{\,m}\, \hat{p}^{\,n}, \quad c_{mn} \in \mathbb{C}. \tag{27.105}$$

Using Eq. (27.104) we obtain the following formal expressions[72]:

$$[\hat{x},\ \hat{A}(\hat{x},\ \hat{p}\,)] = i\hbar \sum_{m,n} c_{mn}\, n\, \hat{x}^{\,m}\, \hat{p}^{\,(n-1)} = i\hbar \frac{\partial \hat{A}(\hat{x},\ \hat{p}\,)}{\partial \hat{p}}, \tag{27.106}$$

$$[\,\hat{p},\ \hat{A}(\hat{x},\ \hat{p}\,)] = -i\hbar \sum_{m,n} c_{mn}\, m\, \hat{x}^{\,(m-1)}\, \hat{p}^{\,n} = -i\hbar \frac{\partial \hat{A}(\hat{x},\ \hat{p}\,)}{\partial \hat{x}}. \tag{27.107}$$

These expressions are useful in formal evaluation of commutators involving the Hamiltonian of the form $\hat{H} = \hat{K} + \hat{V}$, where $\hat{K} = \hat{p}^2/2m$ is the kinetic energy and $\hat{V}$ is a polynomial function of $\hat{x}$. Then we have

$$[\hat{x},\ \hat{H}\,] = i\hbar \frac{\partial \hat{H}}{\partial \hat{p}} = \frac{i\hbar}{m}\, \hat{p}, \tag{27.108}$$

$$[\,\hat{p},\ \hat{H}\,] = -i\hbar \frac{\partial \hat{H}}{\partial \hat{x}} = -i\hbar\, \frac{\partial \hat{V}}{\partial \hat{x}}. \tag{27.109}$$

For the simple harmonic oscillator we immediately get

$$[\hat{x},\ \hat{H}_{ho}] = \frac{i\hbar}{m}\, \hat{p}, \quad [\,\hat{p},\ \hat{H}_{ho}] = -i\hbar m\omega^2\, \hat{x}. \tag{27.110}$$

---

[70] The method of induction (Spiegel (1) p. 7) can provide the required proof.

[71] Messiah Vol. 1 p. 208. Roman (1965) p. 11. The order in which $\hat{x}^{\,m}$ and $\hat{p}^{\,n}$ appear in the sum is important. A sum of $\hat{p}^{\,m}\hat{x}^{\,n}$ would produce a different operator. de Lange and Raab pp. 18–21 has more examples.

[72] The derivatives are formal expressions, e.g., $d\,\hat{p}^{\,n}/d\,\hat{p} := n\,\hat{p}^{\,(n-1)}$.

E27.7(4) Using the commutation relations between position and momentum operators, we can verify the following canonical angular momentum commutation relations[73]:

$$[\widehat{L}_{cx}, \widehat{L}_{cy}] = i\hbar\widehat{L}_{cz}, \tag{27.111}$$
$$[\widehat{L}_{cz}, \widehat{L}_{cx}] = i\hbar\widehat{L}_{cy}, \tag{27.112}$$
$$[\widehat{L}_{cy}, \widehat{L}_{cz}] = i\hbar\widehat{L}_{cx}, \tag{27.113}$$
$$[\widehat{L}_{cx}, \widehat{L}_{c}^2] = [\widehat{L}_{cy}, \widehat{L}_{c}^2] = [\widehat{L}_{cz}, \widehat{L}_{c}^2] = \widehat{0}. \tag{27.114}$$

These angular momentum operators are markedly different from the spin angular momentum operators discussed in §14.1.1. However, by comparing the above equations with Eqs. (14.46) and (14.47) we can see that they share the same commutation relations.

27.8 A Particle in Circular Motion

It is possible to quantise the motion of a chargeless particle in the absence of magnetic field under certain geometric constraint. As an illustration consider a particle of mass m constrained to move freely along the circumference of a circle $\mathcal{C}_a$ of radius a on the x-y plane centred at the origin.[74] The position of the particle can be specified by the angular position variable $\theta(\mathcal{C}_a)$ or by the linear position variable $s(\mathcal{C}_a) = a\theta(\mathcal{C}_a)$.[75] The motion of the particle can be specified by its momentum or its angular momentum[76]:

(1) Linear momentum $p(\mathcal{C}_a)$ is given by mv, where $v = \dot{s} := ds/dt$ is the linear velocity along the circumference. The kinetic energy is $K = m\dot{s}^2/2$. For free motion this is equal to the Lagrangian $\mathcal{L}$ in accordance with Eq. (27.2). Hence the canonical momentum conjugate to s given by $\partial\mathcal{L}/\partial\dot{s}$ is equal to the linear momentum

[73] For brevity the subscript c is usually omitted, e.g., writing $\widehat{L}_{cx}$ as $\widehat{L}_x$.

[74] Martin pp. 46–47. Wan pp. 480–485. A constrained system can present many problems in quantisation. *See* §27.10.2 and §27.10.3 for more comments.

[75] *See* §16.1.2.8 for the notation.

[76] The terms *linear position* and *velocity* are meant to the components of the quantities. As vectors the linear position and velocity are given, respectively, by $a\theta\vec{s}^{\,(u)}$ and $v\vec{s}^{\,(u)}$, where $\vec{s}^{\,(u)}$ is the unit vector along the anti-clockwise tangential direction (*see* Eq. (6.17) for the notation). The same terminology applies to linear and angular momenta.

$p(\mathcal{C}_a)$. It follows that s and $p(\mathcal{C}_a)$ form a pair of canonical variables.

(2) Angular momentum $L(\mathcal{C}_a)$ is given by $ap(\mathcal{C}_a) = mva$.

Quantum mechanically the state space associated with particle is $\vec{L}^2(\mathcal{C}_a)$ with scalar product given by Eq. (16.40).[77] Following the canonical quantisation scheme, the canonical variables s and $p(\mathcal{C}_a)$ are quantised as selfadjoint operators $\widehat{s}(\mathcal{C}_a)$ and $\widehat{p}(\mathcal{C}_a)$ in $\vec{L}^2(\mathcal{C}_a)$ obeying the following commutation relation[78]:

$$[\,\widehat{s}(\mathcal{C}_a),\ \widehat{p}(\mathcal{C}_a)\,] = i\hbar\,\widehat{I\!I}. \tag{27.115}$$

The following comments aim to clarify the quantised system:

C27.8(1) The linear position operator is the bounded multiplication operator $\widehat{s}(\mathcal{C}_a)$ acting on $\vec{\phi} \in \vec{L}^2(\mathcal{C}_a)$ by $\widehat{s}(\mathcal{C}_a)\,\vec{\phi} := a\theta\,\phi(\theta)$. The angular position $\widehat{\theta}(\mathcal{C}_a)$ is defined by $\widehat{\theta}(\mathcal{C}_a)\,\vec{\phi} := \theta\,\phi(\theta)$.

C27.8(2) The momentum operator is the selfadjoint operator $\widehat{p}(\mathcal{C}_a)$ given in Eq. (17.38), i.e.,

$$\widehat{p}(\mathcal{C}_a) = -\frac{i\hbar}{a}\frac{d}{d\theta}. \tag{27.116}$$

The eigenvectors and eigenvalues of $\widehat{p}(\mathcal{C}_a)$ are $\vec{\varphi}_n(\mathcal{C}_a)$ and $p_n(\mathcal{C}_a)$ given, respectively, by Eqs. (19.36) and (19.37).

C27.8(3) The angular momentum operator is the selfadjoint operator $a\,\widehat{p}(\mathcal{C}_a)$, i.e.,

$$\widehat{L}(\mathcal{C}_a) := a\,\widehat{p}(\mathcal{C}_a) = -i\hbar\frac{d}{d\theta}. \tag{27.117}$$

This angular momentum operator also admits $\vec{\varphi}_n(\mathcal{C}_a)$ as its eigenvectors corresponding to eigenvalues

$$ap_n(\mathcal{C}_a) = n\hbar, \quad n = 0, \pm1, \pm2, \ldots. \tag{27.118}$$

These results for $n \neq 0$ agree with the quantised angular momentum values of the Bohr's atom.[79]

---

[77] For a particle confined to move on the surface of a unit sphere, the state space is $\vec{L}^2(\mathcal{S}_u)$. *See* Q36(3) on the rigid rotator.

[78] *See* §27.10.3 for a discussion on problems which can arise. We have omitted the subscript c (for canonical) in the quantised momentum operator.

[79] Bohr (1885–1962) was a Danish physicist famous for his model of the atom.

C27.8(4) The angular position and angular momentum operators satisfy the following commutation relation[80]:

$$[\,\widehat{\theta}(\mathcal{C}_a),\, \widehat{L}(\mathcal{C}_a)\,] = i\hbar\,\widehat{I\!I}. \tag{27.119}$$

C27.8(5) The Hamiltonian is

$$\widehat{H}(\mathcal{C}_a) := \frac{1}{2m}\,\widehat{p}^{\,2}(\mathcal{C}_a), \tag{27.120}$$

with eigenvectors $\vec{\varphi}_n(\mathcal{C}_a)$ corresponding to eigenvalues $E_n(\mathcal{C}_a) = p_n(\mathcal{C}_a)^2/2m$.

An application of the system described above to the motion of a particle in magnetic field which gives rise to the *Aharonov–Bohm effect* is discussed in §37.6.

27.9 Characterisation of Annihilation and Creation Operators

Annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$ and their number operator $\widehat{N}$ are introduced in Definitions 17.10(1), 17.10(2) and 19.1(4). The characteristic feature of a pair of annihilation and creation operators is that they are associated with an orthonormal basis $\{\vec{\varphi}_n\}$ on which they operate in accordance with Eqs. (17.114), (17.115) and (17.121). These operators satisfy the following commutation relations[81]:

$$[\,\widehat{a},\, \widehat{a}^\dagger\,] = \widehat{I\!I}, \quad [\,\widehat{a}^\dagger,\, \widehat{a}\,] = -\widehat{I\!I}, \tag{27.121}$$

$$[\,\widehat{a},\, \widehat{N}\,] = \widehat{a}, \quad [\,\widehat{a}^\dagger,\, \widehat{N}\,] = -\widehat{a}^\dagger. \tag{27.122}$$

A pair of annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$ possesses the following properties:

P27.9(1) They are defined on a common dense domain, i.e., $\mathcal{D}(\widehat{a}\,) = \mathcal{D}(\widehat{a}^\dagger)$ and they are adjoints of each other and hence they are closed operators.

---

[80] The commutation relation is valid only when applied to a restricted domain. *See* the discussion leading to Eq. (27.154) and Q27(15).

[81] *See* Eqs. (17.125) and (17.126).

P27.9(2) They obey the commutation relations in Eq. (27.121).

P27.9(3) There exists a vector $\vec{\varphi}_0 \neq \vec{0}$ which the annihilation operator will annihilate, i.e., $\widehat{a}\,\vec{\varphi}_0 = \vec{0}$.

P27.9(4) They form an irreducible set of operators in accordance with the conclusion in §20.7.

Equation (27.121) is similar to the commutation relation which characterises the position and momentum operators. This raises the question as to whether the commutation relation in Eq. (27.121) would help to characterise a pair of annihilation and creation operators, i.e.,

> *Is it possible to characterise a pair of operators in a Hilbert space having the properties of a pair of annihilation and creation operators without direct reference to an orthonormal basis?*

To answer the above question, we have to tackle the following mathematical problem:

> *Given an operator* $\widehat{a}$ *and its adjoint* $\widehat{a}^\dagger$ *in a Hilbert space* $\vec{\mathcal{H}}$ *satisfying properties* P27.9(1) *to* P27.9(4) *above show that there exists an orthonormal basis* $\{\vec{\varphi}_n\}$ *for* $\vec{\mathcal{H}}$ *on which* $\widehat{a}$ *and* $\widehat{a}^\dagger$ *act in accordance with* Eqs. (17.114), (17.115) *and* (17.121).

The main problem is to find an orthonormal basis to associate with the given operators. We know from Eq. (19.6) that the number operator associated with a pair of creation and annihilation operators has a discrete and nondegenerate spectrum $sp(\widehat{N}) = \{n = 0, 1, 2, 3, \cdots\}$ with the corresponding eigenvectors forming an orthonormal basis. This suggests that we can solve our problem in two stages[82]:

Stage 1 Given a pair of operators $\widehat{a}$ and $\widehat{a}^\dagger$ satisfying properties P27.9(1) to P27.9(4) we form the operator

$$\widehat{N} = \widehat{a}^\dagger \widehat{a}. \tag{27.123}$$

[82] Jauch pp. 211–214.

This operator has the following properties of a number operator:

(1) $\widehat{N}$ is selfadjont on account of Theorem 19.1(1).

(2) $\widehat{N}$ satisfies the commutation relations in Eq. (27.122) which is equivalent to

$$\widehat{N}\widehat{a} = \widehat{a}\widehat{N} - \widehat{a}, \quad \widehat{N}\widehat{a}^\dagger = \widehat{a}^\dagger\widehat{N} + \widehat{a}^\dagger. \tag{27.124}$$

(3) $\widehat{N}$ has a set of eigenvalues with a corresponding set of normalised eigenvectors on account of Eq. (27.124):

(a) $\widehat{N}$ possesses the eigenvalue 0, since according to P27.9(2) there exists a unit vector $\vec{\varphi}_0$ such that $\widehat{a}\vec{\varphi}_0 = 0$. In other words, $\vec{\varphi}_0$ is an eigenvector of $\widehat{N}$ corresponding to the eigenvalue 0.

(b) $\widehat{N}$ possesses a set of positive integer eigenvalues and eigenvectors which can be obtained by successive applications of $\widehat{a}^\dagger$ to $\vec{\varphi}_0$, e.g.,

$$\begin{aligned}
\widehat{N}\left(\widehat{a}^\dagger\vec{\varphi}_0\right) &= \left(\widehat{a}^\dagger\widehat{N} + \widehat{a}^\dagger\right)\vec{\varphi}_0 = 1\left(\widehat{a}^\dagger\vec{\varphi}_0\right), \\
\widehat{N}\left(\left(\widehat{a}^\dagger\right)^2\vec{\varphi}_0\right) &= \left(\widehat{a}^\dagger\widehat{N} + \widehat{a}^\dagger\right)\left(\widehat{a}^\dagger\vec{\varphi}_0\right) = 2\left(\left(\widehat{a}^\dagger\right)^2\vec{\varphi}_0\right), \\
\widehat{N}\left(\left(\widehat{a}^\dagger\right)^3\vec{\varphi}_0\right) &= \left(\widehat{a}^\dagger\widehat{N} + \widehat{a}^\dagger\right)\left(\left(\widehat{a}^\dagger\right)^2\vec{\varphi}_0\right) = 3\left(\left(\widehat{a}^\dagger\right)^3\vec{\varphi}_0\right).
\end{aligned}$$

We can continue the process to get

$$\widehat{N}\left(\left(\widehat{a}^\dagger\right)^n\vec{\varphi}_0\right) = n\left(\left(\widehat{a}^\dagger\right)^n\vec{\varphi}_0\right), \tag{27.125}$$

for any positive integer n. The vectors $\left(\widehat{a}^\dagger\right)^n\vec{\varphi}_0$ are not normalised. Their normalised counterparts are[83]

$$\vec{\varphi}_n = \frac{1}{\sqrt{n!}}\left(\widehat{a}^\dagger\right)^n\vec{\varphi}_0. \tag{27.126}$$

The conclusion is that starting from $\vec{\varphi}_0$ we can generate a set of integer eigenvalues together with their corresponding

[83] This can be proved by the method of induction.

normalised eigenvectors as listed below[84]:

eigenvalues	normalised eigenvectors	
0	$\vec{\varphi}_0$,	(27.127)
1	$\vec{\varphi}_1 = \hat{a}^\dagger \vec{\varphi}_0$,	(27.128)
2	$\vec{\varphi}_2 = \frac{1}{\sqrt{2}} (\hat{a}^\dagger)^2 \vec{\varphi}_0$	(27.129)
3	$\vec{\varphi}_3 = \frac{1}{\sqrt{6}} (\hat{a}^\dagger)^3 \vec{\varphi}_0$,	(27.130)
$\vdots$	$\vdots$	(27.131)
n	$\vec{\varphi}_n = \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^n \vec{\varphi}_0$.	(27.132)

(4) The operators $\hat{a}^\dagger$ and $\hat{a}$ act like creation and annihilation operators on $\vec{\varphi}_n$:

$$\begin{aligned} \hat{a}\vec{\varphi}_n &= \hat{a} \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^n \vec{\varphi}_0 = \hat{a} \frac{\hat{a}^\dagger}{\sqrt{n}} \frac{1}{\sqrt{(n-1)!}} (\hat{a}^\dagger)^{(n-1)} \vec{\varphi}_0 \\ &= \frac{\hat{a}\hat{a}^\dagger}{\sqrt{n}} \vec{\varphi}_{n-1} = \frac{\hat{a}^\dagger\hat{a} + 1}{\sqrt{n}} \vec{\varphi}_{n-1} = \frac{\hat{N}\vec{\varphi}_{n-1} + \vec{\varphi}_{n-1}}{\sqrt{n}} \\ \Rightarrow \quad & \hat{a}\vec{\varphi}_n = \sqrt{n}\, \vec{\varphi}_{n-1}. \end{aligned} \tag{27.133}$$

$$\begin{aligned} \vec{\varphi}_{n+1} &= \frac{1}{\sqrt{(n+1)!}} (\hat{a}^\dagger)^{(n+1)} \vec{\varphi}_0 = \frac{\hat{a}^\dagger}{\sqrt{n+1}} \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^n \vec{\varphi}_0 \\ &= \frac{\hat{a}^\dagger}{\sqrt{n+1}} \vec{\varphi}_n \\ \Rightarrow \quad & \hat{a}^\dagger \vec{\varphi}_n = \sqrt{n+1}\, \vec{\varphi}_{n+1}. \end{aligned} \tag{27.134}$$

Stage 2 We need to show that this set of eigenvectors $\vec{\varphi}_n$ forms an orthonormal basis for $\vec{\mathcal{H}}$. Let us first consider the subspace $\vec{\mathcal{S}}$ of $\vec{\mathcal{H}}$ spanned by these eigenvectors. The orthogonal complement of this subspace will be denoted by $\vec{\mathcal{S}}^\perp$. Then both these subspaces are invariant under $\hat{a}$ and $\hat{a}^\dagger$. To prove this, let

$$\vec{\phi} \in \mathcal{D}(\hat{a}) \cap \vec{\mathcal{S}} \quad \text{and} \quad \vec{\phi}^\perp \in \mathcal{D}(\hat{a}) \cap \vec{\mathcal{S}}^\perp, \tag{27.135}$$

[84] These eigenvectors are orthogonal to each other, being the eigenvectors of a selfadjoint operators corresponding to different eigenvalues.

where $\mathcal{D}(\hat{a})$ is the domain of $\hat{a}$ which coincides with the domain $\mathcal{D}(\hat{a}^\dagger)$ of $\hat{a}^\dagger$ according to P27.9(1). Then:

(1) $\vec{\mathcal{S}}$ is invariant under $\hat{a}$ and $\hat{a}^\dagger$ since[85]:

(a) A vector $\vec{\phi}$ in Eq. (27.135) must be a linear combination of $\vec{\varphi}_n$, i.e., we have $\vec{\phi} = \sum_{n=0}^{\infty} c_n \vec{\varphi}_n$. The output vectors $\hat{a}\,\vec{\phi}$ and $\hat{a}^\dagger\,\vec{\phi}$ are also linear combinations of $\vec{\varphi}_n$ $\vec{\varphi}_n$, i.e.,

$$\hat{a}\vec{\phi} = \sum_{n=1}^{\infty} c_n\, \hat{a}\, \vec{\varphi}_n = \sum_{n=1}^{\infty} \sqrt{n}\, c_n \vec{\varphi}_{n-1} \in \vec{\mathcal{S}}, \tag{27.136}$$

$$\hat{a}^\dagger\vec{\phi} = \sum_{n=0}^{\infty} c_n\, \hat{a}^\dagger\, \vec{\varphi}_n = \sum_{n=0}^{\infty} \sqrt{n+1}\, c_n \vec{\varphi}_{n+1} \in \vec{\mathcal{S}}. \tag{27.137}$$

It follows that $\vec{\mathcal{S}}$ is invariant under $\hat{a}$ and $\hat{a}^\dagger$.

(b) For $\vec{\phi}^\perp$ in Eq. (27.135) we have, for all n,

$$\begin{aligned} \langle\, \hat{a}\vec{\phi}^\perp \mid \vec{\varphi}_n \rangle &= \langle \vec{\phi}^\perp \mid \hat{a}^\dagger \vec{\varphi}_n \rangle \sqrt{(n+1)}\, \langle \vec{\phi}^\perp \mid \vec{\varphi}_{n+1} \rangle = 0 \\ &\Rightarrow \quad \hat{a}\vec{\phi}^\perp \in \vec{\mathcal{S}}^\perp, \end{aligned} \tag{27.138}$$

$$\begin{aligned} \langle\, \hat{a}^\dagger\vec{\phi}^\perp \mid \vec{\varphi}_n \rangle &= \langle \vec{\phi}^\perp \mid \hat{a}\vec{\varphi}_n \rangle = \sqrt{n}\, \langle \vec{\phi}^\perp \mid \vec{\varphi}_{n-1} \rangle = 0 \\ &\Rightarrow \quad \hat{a}^\dagger\vec{\phi}^\perp \in \vec{\mathcal{S}}^\perp. \end{aligned} \tag{27.139}$$

(2) $\mathcal{D}(\hat{a})$ is invariant under the projector $\hat{P}_{\vec{\mathcal{S}}}$.[86] To prove this, let $\vec{\psi} \in \mathcal{D}(\hat{a})$ and let $\vec{\psi}_{\vec{\mathcal{S}}} = \hat{P}_{\vec{\mathcal{S}}}\vec{\psi}$ be the projection of $\vec{\psi}$ onto $\vec{\mathcal{S}}$, i.e.,[87]

$$\vec{\psi}_{\vec{\mathcal{S}}} = \sum_{n=0}^{\infty} \langle \vec{\varphi}_n \mid \vec{\psi} \rangle\, \vec{\varphi}_n. \tag{27.140}$$

Then we have[88]

$$\begin{aligned} \hat{P}_{\vec{\mathcal{S}}}\big(\hat{a}\vec{\psi}\,\big) &= \sum_{n=0}^{\infty} \langle \vec{\varphi}_n \mid \hat{a}\vec{\psi} \rangle\, \vec{\varphi}_n = \sum_{n=0}^{\infty} \langle \hat{a}^\dagger\vec{\varphi}_n \mid \vec{\psi} \rangle\, \vec{\varphi}_n. \\ &= \sum_{n=0}^{\infty} \sqrt{n+1}\, \langle \vec{\varphi}_{n+1} \mid \vec{\psi} \rangle\, \vec{\varphi}_n. \end{aligned} \tag{27.141}$$

---

[85] *See* Definition 17.9(1).

[86] See Definition 17.9(2). $\hat{P}_{\vec{\mathcal{S}}}$ is the projector onto the subspace $\vec{\mathcal{S}}$.

[87] Here $\vec{\psi}$ is not necessarily in $\vec{\mathcal{S}}$ and $\vec{\psi}_{\vec{\mathcal{S}}}$ is not necessarily in $\mathcal{D}(\hat{a})$.

[88] The projection $\hat{P}_{\vec{\mathcal{S}}}\big(\hat{a}\vec{\psi}\,\big)$ would have a finite norm, i.e., $||\hat{P}_{\vec{\mathcal{S}}}\big(\hat{a}\vec{\psi}\,\big)|| \leq \infty$.

$$||\widehat{P}_{\vec{\mathcal{S}}}(\widehat{a}\vec{\psi}\,)||^2 = \sum_{n=0}^{\infty}(n+1)\,|\langle\vec{\varphi}_{n+1}\mid\vec{\psi}\,\rangle|^2$$
$$= \sum_{m=1}^{\infty} m\,|\langle\vec{\varphi}_m\mid\vec{\psi}\,\rangle|^2 < \infty. \qquad (27.142)$$

This implies that $\vec{\psi}_{\vec{\mathcal{S}}}$ is in $\mathcal{D}(\widehat{a}\,)$ since when acted on by $\widehat{a}$ the output vector $\widehat{a}\vec{\psi}_{\vec{\mathcal{S}}}$ is well-defined with a finite norm, i.e.,

$$\widehat{a}\vec{\psi}_{\vec{\mathcal{S}}} = \sum_{n=1}^{\infty}\langle\vec{\varphi}_n\mid\vec{\psi}\,\rangle\sqrt{n}\,\vec{\varphi}_{n-1} \qquad (27.143)$$

$$\Rightarrow\quad ||\widehat{a}\vec{\psi}_{\vec{\mathcal{S}}}||^2 = \sum_{n=1}^{\infty}|\langle\vec{\varphi}_n\mid\vec{\psi}\,\rangle|^2\,n. \qquad (27.144)$$

In accordance with Definitions 17.9(2), 17.9(3) and Theorem 17.9(1), we conclude that $\vec{\mathcal{S}}$ reduces the operators $\widehat{a}$, and $\widehat{a}^\dagger$.

(3) Since $\widehat{a}$ and $\widehat{a}^\dagger$ are irreducible by P27.9(3), the subspace $\vec{\mathcal{S}}$ has to be $\vec{\mathcal{H}}$ itself to avoid a contradiction. It follows that the eigenvectors $\vec{\varphi}_n$ spans $\vec{\mathcal{H}}$, i.e., the set of eigenvectors is a complete set in $\vec{\mathcal{H}}$. This implies that $\widehat{N}$ possesses a discrete and nondegenerate spectrum.[89]

We can conclude that the pair operators $\widehat{a}$ and $\widehat{a}^\dagger$ with properties P27.9(1) to P27.9(4) do have the properties of a pair of annihilation and creation operators.[90]

There are many occasions when we introduce a pair of operators satisfying P27.9(1), P27.9(2)and P27.9(3) but not P27.9(4).[91] Such a pair of operators would still possess many of the properties of annihilation and creation operators. Notably its associated number operator would still have the same set of eigenvalues as before, i.e., $0, 1, 2, \ldots$ and its eigenvectors can still be generated from the eigenvector corresponding to the eigenvalue 0. However, the eigenvalues would be degenerate. We would still call such a pair of operators a pair of creation and annihilation operators.

[89] This also implies the absence of any continuous part in the spectrum.
[90] *See* Jauch pp. 215–219 on explicit representations of these operators.
[91] See §35.3.3 and §36.2.1.

27.10 Limitations of Canonical Quantisation Scheme

The canonical quantisation scheme based on Postulate 27.2(CQ) and its extensions has its limitations.

27.10.1 Problems due to Non-Commuting Factors

An example is the classical observable $A = xp$ for a particle in one-dimensional motion along the x-axis. Clearly $A = xp = px$, but $\widehat{p}\,\widehat{x}$ is not equal to $\widehat{x}\,\widehat{p}$. We have to use the symmetrisation rule to obtain the following quantised operator:

$$\widehat{A} = \frac{1}{2}\left(\widehat{x}\,\widehat{p}_x + \widehat{p}_x\widehat{x}\right). \tag{27.145}$$

This operator is selfadjoint in $\vec{L}^2(I\!R)$. Such an ad hoc scheme cannot be expected to work in every case.[92] An important example is that of the radial momentum p_r conjugate to the radial coordinate r in a spherical coordinate system introduced in Eq. (27.22). When expressed in terms of quantities in the Cartesian coordinates this radial momentum has the following expression:

$$p_r = \frac{x}{r}\,p_x + \frac{y}{r}\,p_y + \frac{z}{r}\,p_z, \quad r = \sqrt{x^2 + y^2 + z^2}. \tag{27.146}$$

Each term can be symmetrised, e.g., we can rewrite the first term as

$$\frac{1}{2}\left(\frac{x}{r}\,p_x + \frac{x}{r}\,p_x\right). \tag{27.147}$$

We may then attempt to quantise p_r using the symmetrisation rule.[93] The resulting operator has an operator expression

$$-i\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) \tag{27.148}$$

[92] The symmetrised product of two selfadjoint operators may not be selfadjoint (*see* Wan p. 101).

[93] Merzbacher p. 255. Zettili p. 324. de Lange and Raab p. 151 for commutation relations involving $\widehat{p}_r$.

acting in $\vec{L}^2(I\!R^3)$. Following the standard procedure, we would specify a domain within $\vec{L}^2(I\!R^3)$, e.g., $\vec{C}_c^\infty(I\!R^3)$, for this operator expression to act so as to produce a symmetric operator. The domain can be extended beyond $\vec{C}_c^\infty(I\!R^3)$ to produce an extended symmetric operator, denoted by $\widehat{p}_r$ and referred to as the **radial momentum operator**.[94] We would then look for selfadjoint extensions. But such a standard approach fails to produce any selfadjoint extension. This is because there does not exist a domain for the above operator expression to act on to produce a selfadjoint operator.[95]

The operator $\widehat{p}_r$ can be shown to be closed. Theorem 19.1(1) then tells us that the operator $\widehat{p}_r^\dagger \widehat{p}_r$ which has the following operator expression:

$$\widehat{p}_r^\dagger \widehat{p}_r = -\hbar^2 \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) \tag{27.149}$$

is selfadjoint. It is this operator which appears in the traditional expression of the kinetic energy operator $\widehat{K}$ in spherical coordinates, i.e., we have

$$\widehat{K} = \frac{1}{2m} \widehat{p}_r^\dagger \widehat{p}_r + \frac{1}{2mr^2} \widehat{L}^2 = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \right) + \frac{1}{2mr^2} \widehat{L}^2. \tag{27.150}$$

Here $\widehat{L}^2$ is the square of the total angular momentum operator.

The above discussion shows that the canonical quantisation scheme is generally not applicable to non-Cartesian coordinates, e.g., the radial momentum p_r conjugate to the radial variable r cannot be quantised into a selfadjoint operator.[96]

27.10.2 Problems due to Constraints on Position

Any constraint in position variables can lead to the failure of the scheme, e.g., failure to establish selfadjointness or uniqueness:

Non-selfadjointness Mathematically the failure to produce a satisfactory quantisation for p_r is due to the limitation on the

[94] The operator $\widehat{p}_r$ has the differential expression in Eq. (27.148) but acting on a bigger domain. *See* Q27(14) for more details.
[95] Wan pp. 174–175, p. 274. Richtmyer Vol. 1 pp. 139–140.
[96] Wan pp. 75–76. *See* Wan pp. 279–281 for tensor product expressions.

radial position variable involved, e.g., r is restricted to the range $[0, \infty)$. There are other situations where the position variable is limited. Physically this limitation may correspond to a particle with its movement constrained within some region in the physical space. Consider a particle in one-dimensional motion confined to the positive x-axis, i.e., the particle's position is restricted to $x \in [0, \infty)$. Then the state space associated with the system would be $\vec{L}^2(I\!R^+)$ rather than $\vec{L}^2(I\!R)$. One may quantise the momentum p as an operator with an operator expression $-i\hbar d/dx$. This operator expression leads to the symmetric operator $\widehat{p}_D(I\!R^+)$ in $\vec{L}^2(I\!R^+)$ defined be Eq. (19.23) which admits no selfadjoint extension.[97] It follows that there is no usual selfadjoint momentum operator.[98]

Non-uniqueness The non-uniqueness of quantisation of both momentum and the Hamiltonian have already been discussed in §27.5.

27.10.3 Problems with Commutation Relations

As discussed in §17.7, care has to be taken when evaluating commutation relations involving unbounded operators. In the Hilbert space $\vec{L}^2(\Lambda)$ the position operator $\widehat{x}(\Lambda)$ is bounded. Let us investigate the commutator

$$[\,\widehat{x}(\Lambda),\, \widehat{p}_{\lambda=0}(\Lambda)\,] = \widehat{x}(\Lambda)\widehat{p}_{\lambda=0}(\Lambda) - \widehat{p}_{\lambda=0}(\Lambda)\widehat{x}(\Lambda), \qquad (27.151)$$

where $\widehat{p}_{\lambda=0}(\Lambda)$ is unbounded. To establish the domain of this commutator, we need to consider the domains of the products $\widehat{x}(\Lambda)\widehat{p}_{\lambda=0}(\Lambda)$ and $\widehat{p}_{\lambda=0}(\Lambda)\widehat{x}(\Lambda)$ in accordance with Eq. (17.58):

(1) The product $\widehat{x}(\Lambda)\widehat{p}_{\lambda=0}(\Lambda)$ acts on the domain $\mathcal{D}\big(\widehat{p}_{\lambda=0}(\Lambda)\big)$ of $\widehat{p}_{\lambda=0}(\Lambda)$, since $\widehat{x}(\Lambda)$ is bounded without any domain restrictions.

(2) The product operator $\widehat{p}_{\lambda=0}(\Lambda)\widehat{x}(\Lambda)$ needs careful consideration. For $\vec{\phi} \in \vec{L}^2(\Lambda)$ to be in the domain of this product operator

[97] Akhiezer and Glazman Vol. 1 pp. 106–112. This has already been pointed out in §27.5.

[98] *See* §27.5.

we require $\widehat{x}(\Lambda)\vec{\phi}$ to be in $\mathcal{D}\big(\widehat{p}_{\lambda=0}(\Lambda)\big)$. This means that the function $x\phi(x)$ must satisfy the periodic boundary condition, i.e., we must have[99]

$$0\phi(0) = L\phi(L) \quad \Rightarrow \quad \phi(L) = 0 \;\Rightarrow\; \phi(0) = 0. \tag{27.152}$$

(3) It follows that the eigenvectors $\vec{\varphi}_{\lambda=0,n}(\Lambda)$ of $\widehat{p}_{\lambda=0}(\Lambda)$ in Eq. (19.34) are not in the domain of the commutator.

A similar situation occurs when we consider the circular motion of a particle in the Hilbert space $\vec{L}^2(\mathcal{C}_a)$. The commutation relation in Eq. (27.119) is deceptively simple. However, we need to examine carefully the domains of product operators $\widehat{\theta}(\mathcal{C}_a)\,\widehat{L}(\mathcal{C}_a)$ and $\widehat{L}(\mathcal{C}_a)\widehat{\theta}(\mathcal{C}_a)$ in accordance with Eq. (17.58)[100]:

(1) The product $\widehat{\theta}(\mathcal{C}_a)\,\widehat{L}(\mathcal{C}_a)$ is an operator acting on the domain of $\mathcal{D}\big(\widehat{L}(\mathcal{C}_a)\big)$, since $\widehat{\theta}(\mathcal{C}_a)$, is bounded.

(2) Let $\vec{\phi}$ be a vector in the domain of the product operator $\widehat{L}(\mathcal{C}_a)\widehat{\theta}(\mathcal{C}_a)$. Then $\widehat{\theta}(\mathcal{C}_a)\vec{\phi} := \theta\phi(\theta)$ must satisfy the following periodic boundary condition,[101] i.e., we have

$$0\phi(0) = 2\pi\phi(2\pi) \quad \Rightarrow \quad \phi(2\pi) = 0 \;\Rightarrow\; \phi(0) = 0. \tag{27.153}$$

(3) The eigenvectors $\vec{\varphi}_n(\mathcal{C}_a)$ of $\widehat{L}(\mathcal{C}_a)$ are not in the domain of the commutator.

(4) As already shown in Eq. (17.81), the commutation relation should be rewritten as

$$[\,\widehat{\theta}(\mathcal{C}_a),\, \widehat{L}(\mathcal{C}_a)\,] \subset i\hbar\widehat{\mathbb{I}}. \tag{27.154}$$

The discussion above is not just a mathematical pedantry. It has physical implications. We shall return to this problem in §28.3.3.

There are other simple quantisation schemes such as the Weyl scheme and the Born–Jordan scheme.[102] None of these schemes

[99] Fano p. 408.
[100] Recall that $\widehat{L}(\mathcal{C}_a)$ is the angular momentum operator.
[101] Fano p. 408.
[102] Wan pp. 256–259.

works in all cases. Moreover, there are novel quantum systems, such as quantum circuit systems to which such simple quantisation scheme cannot be applied directly. More sophistication schemes have to be adopted.[103]

Exercises and Problems

Q27(1) Show that Hamilton's equations of motion in Eq. (27.8) for a classical harmonic oscillator whose Hamiltonian is given by Eq. (27.11) are equivalent to Newton's equation of motion.

Q27(2) Verify the properties of Poisson brackets shown in Eqs. (27.54) to (27.59).

Q27(3) Verify the Poisson bracket relations in Eq. (27.61) between the components of the canonical angular momentum $\vec{L}_c$.

Q27(4) Show that the equation of motion (27.53) in terms of Poisson bracket reduces to the Hamilton's equations when we replace A by x_i and p_i.

Q27(5) Show that a classical observable is a *constant of motion*, i.e., it is time-independent, if it has a zero Poisson bracket with the Hamiltonian.[104]

Q27(6) Verify Eq. (27.70). Show that Postulate 27.2(CQ) as expressed in Eq. (27.69) cannot be valid without the imaginary number i.

Q27(7) Show that $\widehat{U}$ in Eq. (27.79) is unitary and that $\widehat{\mathsf{Q}}', \widehat{\mathsf{P}}'$ in Eq. (27.78) are the unitary transforms of $\widehat{\mathsf{Q}}, \widehat{\mathsf{P}}$ generated by this unitary operator.

[103] There are numerous papers and many books devoted to the subject of quantisation over the years. *See* Wan §3.3, §6.4 for a review and further discussions on this subject.

[104] Recall that we confine ourselves to observables which are not explicitly time dependent unless otherwise is stated.

Q27(8) Prove Eq. (27.104) by the method of induction.

Q27(9) Verify the commutation relations in Eqs. (27.106) and (27.107).

Q27(10) Verify the commutation relations in Eqs. (27.108), (27.109) and (27.110).

Q27(11) Verify the commutation relations in Eqs. (27.111) to (27.114) for angular momentum operators.

Q27(12) Verify the commutation relations in Eq. (27.122).

Q27(13) Prove, by the method of induction, that the eigenvectors in Eq. (27.126) are normalised.

Q27(14) Let $\vec{\Phi}$ be a vector in $\vec{L}^2(I\!R^3)$ defined by the product of a function of the radial variable r and a spherical harmonics $Y_{\ell,m_\ell}(\theta,\varphi)$, i.e.,

$$\vec{\Phi} := \Phi(r,\theta,\varphi) = \phi(r)Y_{\ell,m_\ell}(\theta,\varphi). \tag{27.155}$$

Let $\vec{\Psi}$ be another vector defined in the same way, i.e.,

$$\vec{\Psi} := \Psi(r,\theta,\varphi) = \psi(r)Y_{\ell,m_\ell}(\theta,\varphi). \tag{27.156}$$

Furthermore the functions $\phi(r)$ and $\psi(r)$ satisfy the boundary condition

$$\lim_{r\to 0} r|\phi(r)| = 0 \quad \text{and} \quad \lim_{r\to 0} r|\psi(r)| = 0. \tag{27.157}$$

Working in spherical coordinates show that[105]

$$\langle \vec{\Psi} \mid \widehat{p}_r\vec{\Phi}\,\rangle = \langle \widehat{p}_r\vec{\Psi} \mid \vec{\Phi}\,\rangle,$$

where $\widehat{p}_r$ is the radial momentum operator introduced by Eq. (27.148).[106] Explain why $\widehat{p}_r$ can be symmetric but not selfadjoint.

---

[105] *See* Eq. (16.44). In spherical coordinates the scalar product is given by

$$\langle \vec{\Psi} \mid \widehat{p}_r\vec{\Phi}\,\rangle = -i\hbar\int_0^\infty\int_0^{2\pi}\int_0^\pi \Psi(r,\theta,\varphi)^*\left(\frac{\partial}{\partial r}+\frac{1}{r}\right)\Phi(r,\theta,\varphi)r^2\sin\theta dr d\theta d\varphi.$$

[106] We assume that $\phi(r)$ and $\psi(r)$ are differentiable with respect to r, i.e., they are absolutely continuous in r. *See* Wan pp. 174–175 for more details.

Q27(15) Explain why the product of operators $\widehat{L}(\mathcal{C}_a)$ and $\widehat{\theta}(\mathcal{C}_a)$ in $\vec{L}^2(\mathcal{C}_a)$, i.e., the operator $\widehat{L}(\mathcal{C}_a)\,\widehat{\theta}(\mathcal{C}_a)$, cannot operate on the eigenvectors $\vec{\varphi}_n(\mathcal{C}_a)$ of $\widehat{L}(\mathcal{C}_a)$ in Eq. (19.36), and that[107]

$$[\,\widehat{\theta}(\mathcal{C}_a),\,\widehat{L}(\mathcal{C}_a)\,]\,\vec{\varphi}_n(\mathcal{C}_a) \tag{27.158}$$

is not defined.[108]

[107] *See* Eqs. (17.23) and (27.117) for the definitions of $\widehat{\theta}(\mathcal{C}_a)$ and $\widehat{L}(\mathcal{C}_a)$.
[108] Fano pp. 407–408. *See* also Q22(2).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 28

States, Observables and Probability Distributions

Vectors and operators in a Hilbert space are independent quantities and there is no obvious link between them. The mathematical description of the states and observables of a quantum system given by Postulates 25.1(PS) and 26.1(OV) in terms of vectors and operators does not tell us how a given state would determine the value of an observable. Physically the relationship between states and observables is about how the measurable values of an observable is related to a state. This relationship should lead to the quantum properties QMP5.3(1) to QMP5.3(3). First we must know how a state can give rise to a probability distribution of measurable values of an observable. This would require a prescription to relate a unit vector to the spectrum of selfadjoint operators. The discussion in Chapter 22 tells us how to obtain such a prescription. Because of the intricacy of the relationship both physically and mathematically we shall set out the relationship for discrete and continuous observables separately.

Postulate 26.1(OV) tells us that an observable A has a discrete set of measurable values if the selfadjoint operator $\widehat{A}$ associated with it has a discrete spectrum and that the set of measurable values of A

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

coincides with the set of eigenvalues a_ℓ of $\widehat{A}$. We need to know how a state can determine the probability density function on this discrete set of values.

An observable A has a continuous set of measurable values if the selfadjoint operator $\widehat{A}$ associated with it has a continuous spectrum and that the set of measurable values of A coincides with the spectrum $sp_c(\widehat{A})$ of $\widehat{A}$. Here we need to know how a state can determine the probability distribution function on this continuous set of values of A.

28.1 Postulate (PDDO) on Discrete Observables

Postulate 28.1(PDDO) *When the system is in state ϕ^s the probability distribution of the measurable values a_m of a discrete observable A is given by a probability mass function $\wp^A(\phi^s, a_m)$ on the spectrum $sp_d(A)$ of A specified by* Theorem 22.2 (1)*.*

This postulate tells us that the probability of an individual measured value a_m is given by Eq. (22.9), i.e.,[1]

$$\wp^A(\phi^s, a_m) = \mathcal{Q}(\widehat{P}^{\hat{A}}(a_m), \vec{\phi}\,) = \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_m)\vec{\phi}\,\rangle. \tag{28.1}$$

The corresponding probability distribution function $\mathcal{F}^A(\phi^s, \tau)$ is piecewise constant and it is given explicitly by Eq. (22.11).

The expectation value is given by Eq. (22.12), i.e., we have

$$\mathcal{E}(A, \phi^s) = \mathcal{Q}(\widehat{A}, \vec{\phi}\,) = \langle \vec{\phi} \mid \widehat{A}\vec{\phi}\,\rangle. \tag{28.2}$$

It is often convenient and more transparent to denote the expectation value by $\mathcal{E}(\widehat{A}, \vec{\phi}\,)$. The expectation value is generally different from any of the eigenvalues. However, if the state vector happens to be an eigenvector of the operator $\widehat{A}$, then the expectation value is equal to the corresponding eigenvalue. The uncertainty is given by Eq. (22.4).

[1] Here $\widehat{P}^{\hat{A}}(a_m)$ is the eigenprojector given by Eq. (20.18) and $\phi^s := \vec{\phi}$. For a nondegenerate eigenvalue, the projector is given by Eq. (20.16).

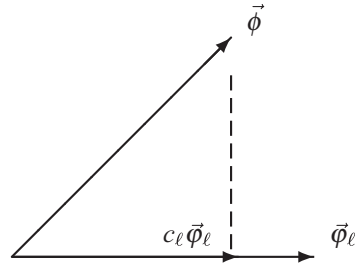


Figure 28.1 Projection of $\vec{\phi}$ onto $\vec{\varphi}_\ell$.

An intuitive understanding of Eq. (28.1) can be obtained by a geometric visualisation of Eq. (28.1). For simplicity let us assume that the eigenvalue a_ℓ is nondegenerate corresponding to the unit eigenvector $\vec{\varphi}_\ell$. Then we have

$$\widehat{P}^{\hat{A}}(a_\ell)\vec{\phi} = c_\ell \vec{\varphi}_\ell, \quad c_\ell = \langle \vec{\varphi}_\ell \mid \vec{\phi} \rangle. \tag{28.3}$$

$$\wp^A(\phi^s, a_\ell) = \langle \vec{\phi} \mid \widehat{P}^{\hat{A}}(a_\ell)\vec{\phi} \rangle = |\, c_\ell \,|^2. \tag{28.4}$$

The output vector $c_\ell \vec{\varphi}_\ell$ is the projection of the input vector $\vec{\phi}$ onto the eigenvector $\vec{\varphi}_\ell$ and the coefficient c_ℓ is component of $\vec{\phi}$ on $\vec{\varphi}_\ell$. A schematic illustration of the geometric situation is shown in Fig. 28.1.

The following comments aim to clarify the postulate:

C28.1(1) When the state vector $\vec{\phi}$ coincides with an eigenvector $\vec{\varphi}_\ell$ the projection of $\vec{\phi}$ onto $\vec{\varphi}_\ell$ coincides with $\vec{\varphi}_\ell$, resulting in $\wp^A(\phi^s, a_\ell) = 1$. Such states are called **eigenstates** of A. They are denoted by φ^s_ℓ. In the eigenstate $\varphi^s_\ell := \vec{\varphi}_\ell$ the probability of a measurement of A resulting in the eigenvalue a_ℓ is 1. In accordance with the discussion in §5.2 we say that *the observable possesses the value a_ℓ in state φ^s_ℓ*.

C28.1(2) When the state vector $\vec{\phi}$ is orthogonal to the eigenvector $\vec{\varphi}_\ell$ the projection of $\vec{\phi}$ onto $\vec{\varphi}_\ell$ is zero. Hence we have $\wp^A(\phi^s, a_\ell) = 0$. This means that when the state vector is orthogonal to the eigenvector $\vec{\varphi}_\ell$ the probability of a measurement of A resulting in the eigenvalue a_ℓ is zero.

C28.1(3) When the state vector $\vec{\phi}$ is neither orthogonal nor equal to the eigenvector $\vec{\varphi}_\ell$, the probability $\wp^A(\phi^s, a_\ell)$ is equal to the square of the norm of the projection of $\vec{\phi}$ onto $\vec{\varphi}_\ell$ since $\wp^A(\phi^s, a_\ell) = ||\widehat{P}^{\hat{A}}(a_\ell)\vec{\phi}\,||^2 = ||c_\ell\vec{\varphi}_\ell\,||^2 = |\,c_\ell|^2$ in accordance with Eq. (28.1).

We would expect that as the state vector $\vec{\phi}$ gets closer to the eigenvector $\vec{\varphi}_\ell$ the probability will becomes higher, i.e., the larger the projection the higher the probability. This is what Eq. (28.1) tells us. Since $\wp^A(\phi^s, a_\ell) = |c_\ell|^2$ by Eq. (28.4), we can interpret $\wp^A(\phi^s, a_\ell)$ as the absolute value square of the component of the state vector $\vec{\phi}$ on the eigenvectors $\vec{\varphi}_\ell$, or as the absolute value square of the coefficient of expansion of the state vector $\vec{\phi}$ in terms of eigenvectors $\vec{\varphi}_\ell$ since $\vec{\phi} = \sum_\ell c_\ell\vec{\varphi}_\ell$.

As an example, consider a particle in circular motion discussed in §27.8. When the state is given by the state vector $\vec{\varphi}_n(\mathcal{C}_a)$ the linear momentum, the angular momentum and the Hamiltonian all have a definite value, i.e., $p_n(\mathcal{C}_a)$, $ap_n(\mathcal{C}_a)$ and $E_n(\mathcal{C}_a)$. When the state is given by a linear combination of $\vec{\varphi}_n(\mathcal{C}_a)$, e.g.,

$$\vec{\phi}(\mathcal{C}_a) = \sum_n c_n\vec{\varphi}_n(\mathcal{C}_a), \quad \sum_n |c_n|^2 = 1, \tag{28.5}$$

none of the above observables has a definite value. Postulate 28.1(PDDO) applies to give a probability distribution of measured values for each of the observables.[2]

C28.1(4) Another interesting interpretation of $\wp^A(\phi^s, a_\ell)$ is to link the $\wp^A(\phi^s, a_\ell)$ to the proposition defined by the eigenprojector $\widehat{P}^{\hat{A}}(a_\ell)$ based on the result:

$$\wp^A(\phi^s, a_\ell) = \langle\vec{\phi} \mid \widehat{P}^{\hat{A}}(a_\ell)\vec{\phi}\,\rangle = \mathcal{E}(\widehat{P}^{\hat{A}}(a_\ell), \vec{\phi}\,). \tag{28.6}$$

Each individual probability $\wp^A(\phi^s, a_\ell)$ can be considered as an expectation value of the proposition $\widehat{P}^{\hat{A}}(a_\ell)$ associated with the observable A, i.e., *probabilities are identifiable with the expectation values of propositions.* On the other hand $\langle\vec{\phi} \mid \widehat{P}^{\hat{A}}(a_\ell)\vec{\phi}\,\rangle$ can also be interpreted as the probability of a yes answer to a yes-no experiment to measure the proposition $\widehat{P}^{\hat{A}}(a_\ell)$ in state ϕ^s.

[2] The present postulate does not apply to the linear position variable s which has a continuous spectrum. A separate postulate is required.

C28.1(5) Suppose the eigenvalue a_m is degenerate with degeneracy d, i.e., there are d orthonormal eigenvectors φ_{mj}, $j = 1, 2, \ldots, d$ corresponding to the eigenvalue a_m. Then the eigensubspace is spanned by these d eigenvectors and is therefore d-dimensional.[3] The corresponding eigenprojector is

$$\widehat{P}^{\hat{A}}(a_m) = \sum_{j=1}^{d} |\vec{\varphi}_{mj}\rangle\langle\vec{\varphi}_{mj}|. \tag{28.7}$$

The probability of a measurement resulting in the value a_m is

$$\begin{aligned}\wp^A(\phi^s, a_m) &= \langle\vec{\phi} \mid \sum_{j=1}^{d} |\vec{\varphi}_{mj}\rangle\langle\vec{\varphi}_{mj}|\,\vec{\phi}\,\rangle = \langle\vec{\phi} \mid \sum_{j=1}^{d} \langle\vec{\varphi}_{mj} \mid \phi\rangle\,\vec{\varphi}_{mj}\rangle \\ &= \sum_{j=1}^{d} \langle\vec{\varphi}_{mj} \mid \vec{\phi}\,\rangle\,\langle\vec{\phi} \mid \vec{\varphi}_{\ell j}\rangle = \sum_{j=1}^{d} |\langle\vec{\varphi}_{mj} \mid \vec{\phi}\,\rangle|^2. \end{aligned} \tag{28.8}$$

28.2 Postulate (PDCO) on Continuous Obsevables and Postulate (PD)

For continuous observables we can establish the probability distribution of a continuous set of observable values in terms spectral functions.

Postulate 28.2(PDCO) *When the system is in state ϕ^s the probability distribution of the measurable values τ of a continuous observable A is given by a probability distribution function $\mathcal{F}^A(\phi^s, \tau)$ specified by* Theorem 22.1(1).

Explicitly the postulate tells us that for $\phi^s := \vec{\phi}$ we have

$$\mathcal{F}^A(\phi^s, \tau) = \mathcal{F}^{\hat{A}}(\vec{\phi}, \tau) = \langle\vec{\phi} \mid \widehat{F}^{\hat{A}}(\tau)\,\vec{\phi}\,\rangle, \tag{28.9}$$

where $\widehat{F}^{\hat{A}}(\tau)$ is the spectral function of the selfadjoint operator $\widehat{A}$ associated with observable A. A list of comments are given below.

C28.2(1) The postulate can be stated in terms of probability measures, i.e., we can re-state the postulate as follows:

[3] *See* Fig. 30.2.

When the system is in state ϕ^s described by state vector $\vec{\phi}$ the probability distribution of the measurable values τ of a continuous observable A is given by the probability measure

$$\mathcal{M}^A(\phi^s, \Lambda) = \langle \vec{\phi} \mid \widehat{M}^{\hat{A}}(\Lambda)\vec{\phi} \rangle. \tag{28.10}$$

where $\widehat{M}^{\hat{A}}$ is the spectral measure of the selfadjoint operator $\widehat{A}$ associated with observable A.

This measure is related to the probability distribution function in accordance with Theorem 22.1(1).

C28.2(2) The probability that a measured value τ of A lies in the half-open interval $(\tau_1, \tau_2]$ is given by

$$\mathcal{M}^A\big(\phi^s, (\tau_1, \tau_2]\big) = \langle \vec{\phi} \mid \Big(\widehat{F}^{\hat{A}}(\tau_2) - \widehat{F}^{\hat{A}}(\tau_1)\Big)\vec{\phi} \rangle. \tag{28.11}$$

C28.2(3) As discussed in P4.3.2(2) a characteristic feature of any probability distribution of a continuous set of values is that the probability for a single value is zero. This feature is ensured by the spectral projector of any singleton set being zero.[4] This property corresponds to the fact that we are not able to single out a single value in a continuum due to experimental error. An example is the position of a particle which takes on a continuous set of values. It is physically impossible to determine the position of a particle to a mathematical point. We shall examine this again later.

C28.2(4) We can introduce a corresponding probability density function and calculate the expectation value in accordance with Eq. (22.6). We also have $\mathcal{E}(A, \phi^s) = \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle$.

C28.2(5) Since the probability distribution of the values of a discrete observable A can also be described by a probability distribution function generated by the spectral function $\widehat{F}^{\hat{A}}(\tau)$ of the selfadjont operator $\widehat{A}$ representing the observable, we can unify Postulates 28.1(PDDO) and 28.2(PDCO) into a single postulate in terms of probability distribution functions. In other words, we can incorporate Postulate 28.1(PDDO) into Postulate 28.2(PDCO) and rename Postulate 28.2(PDCO) as **Postulate 28.2(PD)**.[5]

[4] *See* Eq. (20.23).

[5] Postulate 28.1(PDDO) is a special case of Postulate 28.2(PD). We can also re-state the postulate in terms the probability measure shown in Eq. (28.10).

C28.2(6) Postulate 28.2(PD) provides the most important link between the theory and observational results. It also reflects the **intrinsically probabilistic** nature of quantum systems. In principle it is impossible for a state to determine a definite value of an arbitrary observable. This corresponds to the fact that a measurement of arbitrary observable does not generally yield a definite value with certainty in a given state. This differs fundamentally from the statistical character of mass phenomena, e.g., systems dealt with in statistical mechanics, where probability is introduced into the theory because of practical limitations. It is hopeless to specify all data that characterise a very large system and to deal with the host of simultaneous equations of motion. In orthodox quantum theory this postulate is inescapable. In a sense, it may be thought of as being the consequence of the interference of observation with the observed system. Yet, this by no means implies non-causality or non-objectivity of quantum physics. The meaningful questions we may ask in quantum physics are different from those in classical physics. Nevertheless, there is a clear relationship between theory and objectively ascertained observational data. The meaningful quantities are the probability distributions and the resulting expectation values.

C28.2(7) A state determines the probability distribution of the values of every observable. Conversely the probability distributions of a suitable set of observables can determine that state. The set of propositions represented by all the projectors in the state space can serve as such a set. The corresponding set of yes-no experiments to measure these propositions can determine the state. It is in this sense a state can be considered a measurable quantity.[6]

28.3 Position and Momentum

Consider a particle in one-dimensional motion along the x-axis. From Postulates 25.1(PS), 26.1(OV) and 27.2(CQ), we obtain a theory based on the Hilbert space $\vec{L}^2(I\!R)$ with states represented by

[6] Jauch pp. 93–94.

unit vectors in $\vec{L}^2(I\!R)$ and the position and momentum observables are represented by the selfadjoint operators $\hat{x}$ and $\hat{p}$ in $\vec{L}^2(I\!R)$.

Given a state φ^s described by a unit vector $\vec{\varphi} \in \vec{L}^2(I\!R)$, we desire the following:

(1) Explicit expressions for the position and momentum probability distribution and density functions.
(2) Explicit expressions for the probability for a measured value of the position and of the momentum to be in a given range $(\tau_1, \tau_2]$.
(3) Explicit expressions for the position and momentum expectation values.
(4) An interpretation of the wave function and its Fourier transform.[7]
(5) Explicit expressions for the wave function after a position or a momentum measurement.

The mathematical results in §22.3.1 and §22.3.2 provide most of the answers which are summarised in the following subsections.

28.3.1 The Position Observable

(1) The probability distribution and density functions are

$$w^x(\varphi^s, \tau) = |\varphi(\tau)|^2, \tag{28.12}$$

$$\mathcal{F}^x(\varphi^s, \tau) = \int_{-\infty}^{\tau} w^x(\varphi^s, \tau)d\tau = \int_{-\infty}^{\tau} |\varphi(\tau)|^2 d\tau. \tag{28.13}$$

(2) The probability for a measured value to be in $(\tau_1, \tau_2]$ is

$$\wp^x\big(\varphi^s, (\tau_1, \tau_2]\big) = \int_{-\infty}^{\tau} w^x(\varphi^s, \tau)d\tau = \int_{\tau_1}^{\tau_2} |\varphi(\tau)|^2 d\tau. \tag{28.14}$$

[7] In some elementary exposition of quantum theory, one may get the impression that the interpretation of the wave function and its Fourier transform is an additional postulate. Actually the interpretation is a consequence of Postulate 28.2(PDCO) as clearly shown in the next subsection.

(3) The position expectation value and uncertainty are

$$\mathcal{E}(x, \varphi^s) = \mathcal{E}(\hat{x}, \vec{\varphi}) = \langle \vec{\varphi} \mid \hat{x}\, \vec{\varphi} \rangle = \int_{-\infty}^{\infty} |\varphi(x)|^2\, x\, dx, \quad (28.15)$$

$$\Delta(x, \varphi^s) = \Delta(\hat{x}, \vec{\varphi}) = \sqrt{\langle \vec{\varphi} \mid \hat{x}^2 \vec{\varphi} \rangle - \langle \vec{\varphi} \mid \hat{x}\vec{\varphi} \rangle^2}. \quad (28.16)$$

(4) We can interpret the wave function $\varphi(x)$ is the **position probability amplitude** in the sense that its absolute value square is the position probability density function.

(5) For an explicit expression of the wave function after a position measurement is given later by Eq. (30.6).

28.3.2 The Momentum Observable

(1) The momentum probability distribution and density functions are given by Eqs. (22.22) and (22.24), i.e.,

$$\mathcal{F}^p(\varphi^s, \tau) = \mathcal{F}^{\hat{p}}(\vec{\varphi}, \tau) = \int_{-\infty}^{\tau} |\underset{\sim}{\varphi}(p)|^2\, dp, \quad (28.17)$$

$$w^p(\varphi^s, \tau) = w^{\hat{p}}(\vec{\varphi}, \tau) = |\underset{\sim}{\varphi}(\tau)|^2. \quad (28.18)$$

(2) The probability for a measured value to be in $(\tau_1, \tau_2]$ is

$$\wp^p\big(\varphi^s, (\tau_1, \tau_2]\big) = \int_{\tau_1}^{\tau_2} w^p(\varphi^s, \tau)\, d\tau = \int_{\tau_1}^{\tau_2} |\underset{\sim}{\varphi}(\tau)|^2\, d\tau. \quad (28.19)$$

(3) The momentum expectation value and uncertainty are

$$\mathcal{E}(p, \varphi^s) = \mathcal{E}(\hat{p}, \vec{\varphi}) = \langle \vec{\varphi} \mid \hat{p}\, \vec{\varphi} \rangle, \quad (28.20)$$

$$\Delta(p, \varphi^s) = \Delta(\hat{p}, \vec{\varphi}) = \sqrt{\langle \vec{\varphi} \mid \hat{p}^2 \vec{\varphi} \rangle - \langle \vec{\varphi} \mid \hat{p}\, \vec{\varphi} \rangle^2}. \quad (28.21)$$

(4) The Fourier transform $\underset{\sim}{\varphi}(p)$ of the wave function $\varphi(x)$ is the **momentum probability amplitude** in the sense that its absolute value square is the momentum probability density function.

(5) For an explicit expression of the wave function after a momentum measurement is given later by Eq. (30.7).

28.3.3 Uncertainty Relations

Consider a particle in one-dimensional motion along the x-axis. The operators $\widehat{x}$ and $\widehat{p}$ obey the canonical commutation relation

$$[\widehat{x},\ \widehat{p}\,] = i\hbar. \tag{28.22}$$

The uncertainties of the position $\Delta(x, \varphi^s)$ and momentum $\Delta(p, \varphi^s)$ in state φ^s given in terms of the operators $\widehat{x}$ and $\widehat{p}$ and state vector $\vec{\varphi}$ by Eqs. (28.16) and (28.21) satisfy the following inequality known as the *uncertainty relation* for position and momentum[8]:

$$\Delta(x, \varphi^s)\Delta(p, \varphi^s) = \Delta(\widehat{x}, \vec{\varphi}\,)\Delta(\widehat{p}, \vec{\varphi}\,) \geq \frac{1}{2}\hbar. \tag{28.23}$$

The consequence is that we cannot reduce both $\Delta(x, \varphi^s)$ and $\Delta(p, \varphi^s)$ to arbitrarily small values. If we choose a state to reduce the value of $\Delta(x, \varphi^s)$ we would automatically incur a corresponding increase in $\Delta(p, \varphi^s)$.

One is tempted to assume a similar result for the position and momentum for a particle constrained to move in a circle $\mathcal{C}_a$ on account of the similarity between Eqs. (27.119) and (28.22), i.e., one is tempted to write down

$$\Delta(\widehat{\theta}(\mathcal{C}_a), \vec{\phi}\,)\Delta\big(a\,\widehat{p}(\mathcal{C}_a), \vec{\phi}\,\big) \geq \frac{1}{2}\hbar. \tag{28.24}$$

This would be wrong when applied to a vector $\vec{\phi}$ outside the domain of operation of the commutator.[9] We can see this in two different ways[10]:

(1) For the eigenvector $\vec{\varphi}_{\lambda,n}$ in Eq. (19.38) we have

$$\Delta\big(a\,\widehat{p}(\mathcal{C}_a), \vec{\varphi}_{\lambda,n}\,\big) = 0. \tag{28.25}$$

The uncertain relation in Eq. (28.24) is violated.

(2) The reason for the violation of Eq. (28.24) is that the eigenvector $\vec{\varphi}_{\lambda,n}$ does not satisfy Eq. (27.153), i.e., $\vec{\varphi}_{\lambda,n}$ is not in the domain of the commutator, and any result arising from the commutation relation cannot be applied to $\vec{\varphi}_{\lambda,n}$.[11]

[8] Zettili pp. 94–95. This uncertainty relation is not an independent "principle."
[9] *See* §27.10.3.
[10] Fano pp. 407–408.
[11] A similar violation of the uncertainty relation also occurs for a particle confined by an infinite potential well. *See* Q22(2).

28.4 Compatible and Incompatible Observables

The intrinsic probabilistic behaviour of a quantum system has an important consequence on the compatibility of its observables, a concept introduced in §5.3 in relation to quantum properties QMP5.3(1) and QMP5.3(2). Let us examine how this concept can be formulated in mathematical terms.

28.4.1 Discrete Observables

Let A and B be two observables represented by operators $\widehat{A}$ and $\widehat{B}$, respectively. Suppose A and B both have a discrete set of values, $\{a_\ell\}$ for A and $\{b_m\}$ for B. Then:

(1) Postulate 26.1(OV) tells us that $\{a_\ell\}$ is equal to the set of eigenvalues of $\widehat{A}$ with a corresponding set of eigenvectors $\vec{\varphi}_\ell$, and $\{b_m\}$ is equal to the set of eigenvalues of the operator $\widehat{B}$ with a corresponding set of eigenvectors $\vec{\psi}_m$.

(2) Postulate 28.1(PDDO) tells us that

(a) If none of the eigenvectors of the two operators agree, i.e., $\vec{\varphi}_\ell \neq \vec{\psi}_m$ for any ℓ and m, then there is no state in which both A and B possess a value.

(b) If the two set of eigenvectors coincide, i.e., if $\widehat{A}$ and $\widehat{B}$ share a complete orthonormal set of eigenvectors, we can label their common eigenvectors as $\vec{\varphi}_{\ell,m}$ such that[12]

$$\widehat{A}\vec{\varphi}_{\ell,m} = a_\ell \vec{\varphi}_{\ell,m}, \quad \widehat{B}\vec{\varphi}_{\ell,m} = b_m \vec{\varphi}_{\ell,m}. \tag{28.26}$$

In the state represented by $\vec{\varphi}_{\ell,m}$, observable A possesses the value a_ℓ and observable B possesses the value b_ℓ. In E19.4(2) the spherical harmonics $\vec{Y}_{\ell,m_\ell}$ are the common eigenvectors of operators $\widehat{L}^2$ and $\widehat{L}_z$.[13] In a state described by $\vec{Y}_{\ell,m_\ell}$ both $\widehat{L}^2$ and $\widehat{L}_z$ possess a definite value, i.e., $\ell(\ell+1)\hbar^2$ for $\widehat{L}^2$ and $m_\ell\hbar$ for $\widehat{L}_z$.

[12] The subscripts ℓ and m may be related.

[13] *See* Eq. (19.51). Angular momentum observables are studied in details in Chapter 36.

(c) If only some of the eigenvectors agree, i.e., $\vec{\varphi}_\ell = \vec{\psi}_\ell$ for some ℓ then in a state represented by these common eigenvectors both A and B have a value. In the case of angular momentum we know that the operators $\widehat{L}_x$, $\widehat{L}_y$ and $\widehat{L}_z$ have one common eigenvector, i.e., $\vec{Y}_{\ell=0,m_\ell=0}$. In the state represented by this vector the angular momentum observables L_x, L_y and L_z possess the same value zero.

We can clarify the situation by the following definition:

Definition 28.4.1(1)

(1) *Two discrete observables are said to be* **compatible** *if a complete set of states exists in which both observables possess a value.*

(2) *Two discrete observables are said to be* **essentially incompatible** *if only an incomplete set of states exists in which both observables possess a value.*

(3) *Two discrete observables are said to be* **strictly incompatible***, or incompatible for short, if no state exists in which both observables have a definite value.*[14]

This concept of compatibility is related to simultaneous measurability. Let as clarify these notions with the following comments:

C28.4.1(1) If A and B are compatible, then they share a complete orthonormal set of eigenstates $\varphi^s_{\ell,m}$. In accordance with Postulate 28.1(PDDO) a simultaneous measurement of A and B would yield the value a_ℓ for A and b_m for B. We say that A and B are *simultaneously measurable* in state $\varphi^s_{\ell,m}$. For an arbitrary state vector $\vec{\phi}$, we have the usual expansion of $\vec{\phi}$ in terms of the eigenvectors $\vec{\varphi}_{\ell,m}$ with coefficients $c_{\ell,m}$. Then $|c_{\ell,m}|^2$ is the probability of a measurement of A resulting in the value a_ℓ and of a measurement of B resulting the value b_m. This means that A and B can also be considered **simultaneously measurable in an**

[14] The term *strictly incompatible* is introduced to contrast the term *essentially incompatible*. The term *incompatible*, rather than strictly incompatible, is used whenever there is no risk of confusion.

arbitrary state, although there is no certainty in the measured results. This enables us to make a general statement that

> *Two compatible observables A and B are simultaneously measurable in any state.*

Intuitively one would expect a simultaneous measurement of two compatible observables would involve two separate measurements performed simultaneously. However, if we can effect a simultaneous measurement of two compatible observables by the measurement of a single observable then we would be able to avoid doing two separate measurements simultaneously. An example would be a simultaneous measurement of two compatible propositions $\widehat{P}^{\hat{A}}(a_m)$ and $\widehat{P}^{\hat{B}}(b_n)$ by a single measurement of their product $\widehat{P}^{\hat{A}}(a_m)\widehat{P}^{\hat{B}}(b_n)$ which is again a proposition, i.e., the proposition that on a measurement A will be found to have the value a_m and B will be found to have the value b_n.[15] An explicit example is the position measurement of a particle in two-dimensional motion in the x-y plane. The position observables x and y represented by operators $\widehat{x}(\mathbb{R}^2)$ and $\widehat{y}(\mathbb{R}^2)$ are compatible, and so are the local position observables described by their spectral projectors $\widehat{M}^{\hat{x}}(\Lambda_x)$ and $\widehat{M}^{\hat{y}}(\Lambda_y)$ along the x- and y-axes. A simultaneous measurement of the product of two local position observables, one along the x-axis and one along y-axis corresponding to the projector $\widehat{M}^{\hat{x}}(\Lambda_x)\widehat{M}^{\hat{y}}(\Lambda_y)$ can be physically achieved in terms of a photographic plate, i.e., an image on the plate would yield a value for the position along the x-axis and the position along the y-axis.[16]

Generally Theorems 13.3.4(4) and 20.6(1) provide a way forward. According to the theorems, two commuting selfadjoint operators $\widehat{A}$ and $\widehat{B}$ are functions of a third selfadjoint operator $\widehat{C}$, i.e.,

$$\widehat{A} = f(\widehat{C}), \quad \widehat{B} = g(\widehat{C}). \tag{28.27}$$

We can measure observable C corresponding to the operator $\widehat{C}$, and then infer the values of A and B. A single measurement of C would yield a value c. We can obtain the values of A and B from $a = f(c)$

[15] Since $\widehat{P}^{\hat{A}}(a_m)$ and $\widehat{P}^{\hat{B}}(b_n)$ commute their product, $\widehat{P}^{\hat{A}}(a_\ell)\widehat{P}^{\hat{B}}(b_n)$ is a projector.
[16] Isham p. 97.

and $b = g(c)$, respectively.[17] There is no need to perform two separate measurements of A and B.

C28.4.1(2) Two incompatible observables do not have any eigenstate in common. Hence there is no state in which both observables can possess a value. In other words, *two incompatible observables are not simultaneously measurable in any state.*

C28.4.1(3) Essentially incompatible observables share some common eigenstates but not a complete set of common eigenstates. They are simultaneously measurable in these common eigenstates.

C28.4.1(4) The concept of compatibility of two observables can be formulated in terms of **successive measurements** of these two observables. The concept can also be extended to include more than two observables. We shall return to discuss this topic in §30.3.

The mathematical requirement for two observables to be compatible is contained in Theorem 13.3.4(3) and Definition 20.6(1), as explicitly stated below.

Theorem 28.4.1(1) *Two discrete observables A and B are compatible if their corresponding operators* $\widehat{A}$ *and* $\widehat{B}$ *commute.*

Corollary 28.4.1(1) *Two discrete observables A and B are compatible if all the eigenprojectors* $\widehat{P}^{\hat{A}}(a_\ell)$ *and* $\widehat{P}^{\hat{B}}(b_m)$ *of their corresponding operators* $\widehat{A}$ *and* $\widehat{B}$ *commute.*

Not all observables are compatible. The examples below illustrate various cases:

E28.4.1(1) *Compatible observables*

(1) An observable is compatible with functions of the observable, e.g., A is compatible with A^2.

(2) For motion along a circle $\mathcal{C}_a$ the linear momentum $p(\mathcal{C}_a)$ is compatible with the angular momentum $L(\mathcal{C}_a)$ and the Hamiltonian $H(\mathcal{C}_a)$. They share a complete orthonormal set of

[17] Fano p. 405. Isham p. 97.

common eigenvectors in which all three observables possess a value. The operators mutually commute.[18]

(3) The angular momentum operators $\widehat{L}^2$ and $\widehat{L}_z$ commute. Hence their corresponding angular momentum observables L^2 and L_z are compatible. In §36.1.1 we will show that they have a complete orthonormal set of common eigenvectors in which both observables possess a value.

(4) Consider a discrete observable A represented by operator $\widehat{A}$ which has a nondegenerate spectrum corresponding to a complete orthonormal set of eigenvectors $\vec{\varphi}_\ell$ together with a set of complete orthogonal set of eigenprojectors $\widehat{P}_{\vec{\varphi}_\ell}$. These eigenprojectors mutually commute and they represent compatible propositions of A. In an eigenstate φ_ℓ^s the proposition represented by $\widehat{P}_{\vec{\varphi}_\ell}$ possesses the value 1 while the proposition represented by $\widehat{P}_{\vec{\varphi}_{\ell'}}$ possesses the value 0 if $\ell \neq \ell'$.

(5) Spectral projectors are a generalisation of eigenprojectors. The fact that all the eigenprojectors of two operators commute is equivalent to the fact that all the spectral projectors of the two operators commute. It follows that for discrete observables we can establish compatibility in terms of either eigenprojectors or spectral projectors. By replacing eigenprojectors by spectral projectors Corollary 28.4.1(1) can be generalised to observables with a continuous spectrum, as seen in Definition 28.4.2(1).

E28.4.1(2) *Strictly incompatible observables*

(1) Two components of the spin angular momentum operators $\widehat{S}_x$ and $\widehat{S}_y$ do not commute. Their corresponding observables S_x and S_y are strictly incompatible (or just incompatible for short).

(2) For a particle in an infinite square potential well of width $\Lambda = [0, L]$ in the x-axis, the Hamiltonian represented by $\widehat{K}^\infty(\Lambda)$ in Eq. (19.42) is strictly incompatible with the momentum represented by $\widehat{p}_\lambda(\Lambda)$ in Eq. (17.36) for any λ. This is obvious

[18] Their corresponding operators are given by Eqs. (27.116), (19.49) and (19.49) and their common eigenvectors $\vec{\varphi}_n(\mathcal{C}_a)$ are given by Eq. (19.38) .

when we compare the eigenfunctions of $\widehat{K}^{\infty}(\Lambda)$ in Eq. (19.43) and the eigenfunctions of $\widehat{p}_{\lambda}(\Lambda)$ in Eq. (19.32).

E28.4.1(3) *Essentially incompatible observables* The angular momentum operators $\widehat{L}_x$ and $\widehat{L}_y$ do not commute. However, they do have a common eigenvector, i.e., $\vec{Y}_{\ell=0,m_\ell=0}$, corresponding to the eigenvalue 0 for both operators. Hence their corresponding observables L_x and L_y are simultaneous measurable in that particular state, but not in an arbitrary state. We say that L_x and L_y are essentially incompatible. The situation is different for spin angular momentum, i.e., S_x and S_y are strictly incompatible since there is no zero spin state.

28.4.2 Continuous Observables

Definition 28.4.1(1) does not apply to continuous observables. We can establish a definition for continuous observables using Corollary 28.4.1(1).

Definition 28.4.2(1) *Two continuous observables A and B are said to be compatible if the spectral projectors* $\widehat{M}^{\hat{A}}(\Lambda_1)$ *and* $\widehat{M}^{\hat{B}}(\Lambda_2)$ *commute for all Borel sets* Λ_1 *and* Λ_2 *of* $\mathbb{R}$*, and they are said to be incompatible otherwise.*

The commutativity of spectral projectors $\widehat{M}^{\hat{A}}(\Lambda_1)$ and $\widehat{M}^{\hat{B}}(\Lambda_2)$ implies the commutativity of the operators $\widehat{A}$ and $\widehat{B}$. It follows that Theorem 28.4.1(1) also applies to observables with a continuous spectrum.[19]

The best known incompatible observables with a continuous spectrum of a particle in one-dimensional motion along the x-axis is the position x and momentum p since their corresponding operators $\widehat{x}$ and $\widehat{p}$ do not commute.

The concept of compatibility and incompatibility can be generalised to apply to a continuous observable and a discrete observable, e.g., a continuous observable can be incompatible with a discrete observable. The momentum and the Hamiltonian of a harmonic oscillator are incompatible.

[19] Jauch p. 101. Spectral projectors correspond to propositions which are discrete observables. Theorem 28.4.1(1) applies to them.

Exercises and Problems

Q28(1) An electron spin is in state α_x^s.[20] Find the probability of a measurement of the z-component spin resulting in the value $\hbar/2$.

Q28(2) Using Eqs. (20.68) and (20.69), show that the probability distribution function and the probability measure of a proposition (as a discrete observable) represented by projector $\widehat{P}$ in state vector $\vec{\phi}$ are given by

$$\mathcal{F}^{\hat{P}}(\vec{\phi}, \tau) = \begin{cases} 0 & \tau < 0 \\ 1 - \langle \vec{\phi} \mid \widehat{P}\vec{\phi} \rangle & 0 \leq \tau < 1 \\ 1 & \tau \geq 1 \end{cases} . \tag{28.28}$$

$$\mathcal{M}^{\hat{P}}(\vec{\phi}, \Lambda) = \begin{cases} 1 - \langle \vec{\phi} \mid \widehat{P}\vec{\phi} \rangle & \text{if } \Lambda = \{0\} \\ \langle \vec{\phi} \mid \widehat{P}\vec{\phi} \rangle & \text{if } \Lambda = \{1\} \\ 0 & \text{if } \Lambda \text{ does not contain 0 or 1.} \end{cases} \tag{28.29}$$

Q28(3) For a particle in circular motion, the Hamiltonian $\widehat{H}(\mathcal{C}_a)$ is given by Eq. (27.120). Show that the eigenvalues of the Hamiltonian is degenerate with eigenvectors $\vec{\varphi}_n$ given by Eq. (19.36). Write down the spectral decomposition of the Hamiltonian in the form of Eq. (20.20).

Q28(4) A pair of annihilation and creation operators $\widehat{a}$, $\widehat{a}^\dagger$ are defined in terms of an orthonormal basis $\{\vec{\varphi}_n, n = 0, 1, 2, \ldots\}$ in the state space $\mathcal{H}$ of a quantum system.[21] The corresponding number operator is $\widehat{N} = \widehat{a}^\dagger\widehat{a}$.[22] The energy of the system is represented by the Hamiltonian operator $\widehat{H} = E_0\widehat{N}$.

(a) What are the possible energy values of the system?

(b) Let $\vec{\Phi}_z$ be the unit vector in Eq. (17.129). Find the probability mass function for the probability distribution of energy values of the system in state Φ_z^s described by the state vector $\vec{\Phi}_z$.

(c) Find the energy expectation values in state Φ_z^s.

---

[20] *See* §14.1.1 and §36.3 for the theory for electron spin.
[21] *See* Definitions 17.10(1) and 17.10(2) in §17.10.
[22] *See* Definition 19.1(5).

Q28(5) The Fourier transform $\underset{\sim}{\varphi}(p)$ of a normalised wave function $\varphi(x)$ of a particle is

$$\underset{\sim}{\varphi}(p) = \begin{cases} 0 & p \leq -p_0, \\ 1/\sqrt{2\,p_0} & p \in (-p_0,\, p_0], \\ 0 & p \geq p_0, \end{cases} \tag{28.30}$$

where $p_0 \in I\!R$. What is the probability of a momentum measurement resulting in a value in the range $(-p_0,\, p_0]$? What is the momentum expectation value and uncertainty? Write down the position probability density function in terms of the inverse Fourier transform of $\underset{\sim}{\varphi}(p)$ in Eq. (18.73).

Q28(6) Working in the momentum representation and using Eqs. (20.54) and (20.57) find the probability distribution function for the kinetic energy of a particle of half the unit mass in one-dimensional motion along the x-axis in a given state φ^s.

Chapter 29

Time Evolution

Having discussed states, observables and their relationship, we must now investigate the time evolution of quantum systems. In classical mechanics time develution is described by Newton's second law or Hamiltonian equations. These are equations of motion of the state. Since observables are functions of the state, a knowledge of the time dependence of state will also determine the time dependence of observables.

In quantum theory the situation is not so clear. We have, in addition to the state vector, operators which are independent of the state vector. A knowledge of the time dependence of state vector does not automatically imply any knowledge of the time dependence of operators. Suppose at time $t = 0$ the system is in a state $\phi^s(0)$ described by the state vector $\vec{\phi}(0)$ and its physical observables $A(0)$ are represented by operators $\widehat{A}(0)$ in the chosen state space $\vec{\mathcal{H}}$. the question is

What are the state vector $\vec{\phi}(t)$ and observables $\widehat{A}(t)$ at $t > 0$?

We would expect to have an equation of motion for the state vector and a separate equation of motion for the operators. Such a description of the time evolution of a quantum system in terms of

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

the time dependence of both the state vectors and operators is called the **Interaction picture**.[1]

It turns out that it is not necessary to have two separate and independent equations of motion for the state and observables. We can actually manage with *either* an equation of motion for the state vector *or* an equation of motion for the operators. In other words, there are two further alternative descriptions or *pictures* of quantum evolution. The **Schrödinger picture** is a description of the time evolution of a quantum system in terms of a differential equation involving the time derivative of the state alone. The **Heisenberg picture** is a description of the time evolution of a quantum system in terms of a differential equation involving only the time derivative of observables.

We shall detail these three descriptions of quantum dynamics and then show how these descriptions are physical equivalence.

29.1 The Schrödinger Picture

29.1.1 Schrödinger Equation

Traditionally the time evolution of a quantum system with a Hamiltonian $\widehat{H}$ in the Schrödinger picture is determined by a differential equation known as the *Schrödinger equation* which involves the time derivative of the state vector,[2] i.e.,

$$i\hbar \frac{d\vec{\phi}(t)}{dt} = \widehat{H}\vec{\phi}(t). \tag{29.1}$$

Observables which are not explicitly time dependent are assumed to correspond to time-independent operators. We shall confine ourselves to systems whose Hamiltonians do not depend on time explicitly. In the coordinate representation the state vector is defined by a function of position $\vec{x}$ and time t. For a Hamiltonian given in Eq. (27.93), the Schrödinger equation becomes

$$i\hbar \frac{\partial \phi(\vec{x}, t)}{\partial t} = \left(-\frac{\hbar^2}{2m} \nabla^2 + V(\vec{x}) \right) \phi(\vec{x}, t). \tag{29.2}$$

[1] Also known as the **Dirac picture**.
[2] Also known as the *time-dependent Schrödinger equation* (*see* Eq. (10.27)).

The partial derivative with respect to t is to signify the independence of $\vec{x}$ and t. Some explicit examples of quantum time evolution can be found in §35.3.1.[3]

If the Hamiltonian has a discrete spectrum $\{E_\ell\}$ together with a complete orthonormal set of eigenvectors $\vec{\varphi}_\ell$ the solutions of the Schrödinger equation can be obtained explicitly. Let $\vec{\phi}(0)$ be a given initial state vector, and let

$$\vec{\phi}(0) = \sum_\ell c_\ell \, \vec{\varphi}_\ell, \quad c_\ell = \langle \vec{\varphi}_\ell \mid \vec{\phi} \, \rangle \tag{29.3}$$

be an expansion of $\vec{\phi}(0)$ in terms of $\vec{\varphi}_\ell$. It is easily verified that

(1) If the initial state vector is an eigenvector of $\widehat{H}$, i.e., $\vec{\phi}(0) = \vec{\varphi}_\ell$ then the evolved state vector is[4]

$$\vec{\phi}(t) = e^{-\frac{i}{\hbar} E_\ell t} \, \vec{\varphi}_\ell. \tag{29.4}$$

(2) If the initial state vector is not an eigenvector of $\widehat{H}$, i.e., $\vec{\phi}(0)$ is given by Eq. (29.3), then the evolved state vector is

$$\vec{\phi}(t) = \sum_\ell c_\ell \, e^{-\frac{i}{\hbar} E_\ell t} \, \vec{\varphi}_\ell, \tag{29.5}$$

since $\vec{\phi}(t)$ satisfies the Schrödinger equation and the initial condition, i.e., at $t = 0$ the vector $\vec{\phi}(t)$ reduces to $\vec{\phi}(0)$.

(3) The norm of the state vector is preserved during evolution, i.e., $||\vec{\phi}(t)|| = ||\vec{\phi}(0)||$.

This Schrödinger equation approach to time evolution has its limitations. The Schrödinger equation is only meaningful when the state vectors $\vec{\phi}(t)$ are in the domain $\mathcal{D}(\widehat{H})$ of the Hamiltonian. For example, if the given initial state vector $\vec{\phi}(0)$ is not in $\mathcal{D}(\widehat{H})$ then $\widehat{H}\vec{\phi}(0)$ is undefined. The Schrödinger equation cannot be applied to such an initial state vector to determine the evolved state vector at $t > 0$. This limitation can be removed if we adopt a more general approach using the concept of unitary evolution operators. This

[3] It is common to call Eqs. (29.1) and (29.2) time-dependent Schrödinger equation while the eigenvalue equation for the Hamiltonian is referred to as the time-independent Schrödinger equation.

[4] Such states are called **stationary states** (*see* Eq. (35.37)).

arises from the realisation that

(1) Time evolution preserves of the norm of the state vector.

(2) The Schrödinger equation is related to a one-parameter group of unitary operators as shown in Theorem 21.2(1) of Stone.

29.1.2 Unitary Evolution

Definition 13.4.3(1) introduces the concept of a continuous one-parameter group of unitary operators $\widehat{U}(t)$. Theorems 13.4.3(2) and 21.2(1) of Stone tell us that such a group of unitary operators is generated by a single selfadjoint operator in the form of an exponential function, and conversely the derivative of $\widehat{U}(t)$ with respect to t can generate a selfadjoint operator as shown in Eq. (21.18).

Consider a system with a time-independent Hamiltonian $\widehat{H}$. This Hamiltonian generates a continuous one-parameter group of unitary operators $\widehat{U}(\widehat{H}, t)$ in accordance with Eq. (21.13) by

$$\widehat{U}(\widehat{H}, t) := e^{-\frac{i}{\hbar}\widehat{H}t}, \quad t \in \mathbb{R}. \tag{29.6}$$

Let $\vec{\phi}(0)$ be a given initial state vector in the domain $\mathcal{D}(\widehat{H})$ of $\widehat{H}$, and let

$$\vec{\phi}(t) := \widehat{U}(\widehat{H}, t)\vec{\phi}(0). \tag{29.7}$$

Equation (21.18) of Theorem 18.3(2) of Stone becomes[5]

$$i\hbar\frac{d\vec{\phi}(t)}{dt} = \widehat{H}\vec{\phi}(t). \tag{29.8}$$

This agrees with Eq. (29.1) of Schrödinger. Equation (29.7) which links $\vec{\phi}(t)$ directly to $\vec{\phi}(0)$ is well-defined for arbitrary $\vec{\phi}(0)$. This suggests that the time evolution of an arbitrary initial state vector which is not in the domain $\mathcal{D}(\widehat{H})$ can be given by Eq. (29.7). This result leads to a general postulate of quantum evolution in terms of a continuous one-parameter group of unitary operators.[6]

[5] Intuitively we can see that a formal differentiation of Eq. (29.7) would lead to the desired equation.

[6] We assume that $\widehat{H}$ is not explicitly time dependent. Otherwise the situation can become complicated (*see* Wan pp. 286–290).

Postulate 29.1.2(TESP) on quantum evolution

(1) *Observables which are not explicitly time dependent are represented by time-independent selfadjoint operators.*

(2) *The state is represented by a time-dependent state vector. The time dependence of the state vector is determined by a continuous one-parameter group of unitary operators generated by the Hamiltonian, i.e.,*

$$\widehat{U}(\widehat{H}, t) := e^{-\frac{i}{\hbar}\widehat{H}t}, \quad t \in I\!R, \tag{29.9}$$

in that an arbitrary initial state vector $\vec{\phi}(0)$ *will evolve into the state vector* $\vec{\phi}(t)$ *given by*

$$\vec{\phi}(t) := \widehat{U}(\widehat{H}, t)\vec{\phi}(0). \tag{29.10}$$

We call $\widehat{U}(\widehat{H}, t)$ *time evolution operators*, or simply **evolution operators**, generated by the Hamiltonian $\widehat{H}$. The evolved state vector is a unitary transform of the initial state vector. The norm of the state vector $\vec{\phi}(t)$ remains the same for all times as unitary transformations preserve the norm of vectors.

Definition 29.1.2(1)[7] *An observable* A *is said to be a* **constant of motion***, or to be* **conserved***, if its corresponding selfadjoint operator* $\widehat{A}$ *commutes with the Hamiltonian* $\widehat{H}$*.*

A constant of motion would have time-independent expectation values. To show this, let the state at time t be $\phi^s(t)$. The expectation value $\mathcal{E}(A, \phi^s(t))$ is equal to $\langle\vec{\phi}(t) \mid \widehat{A}\vec{\phi}(t)\rangle$ at time t. Since $\widehat{A}$ commutes with $\widehat{H}$, and hence it also commutes $\widehat{U}(\widehat{H}, t)$, we get

$$\begin{aligned}\mathcal{E}\big(A, \phi^s(t)\big) &= \langle\vec{\phi}(t) \mid \widehat{A}\,\vec{\phi}(t)\rangle = \langle\widehat{U}(\widehat{H}, t)\vec{\phi}(0) \mid \widehat{A}\,\widehat{U}(\widehat{H}, t)\vec{\phi}(0)\rangle \\ &= \langle\vec{\phi}(0) \mid \widehat{A}\,\vec{\phi}(0)\rangle.\end{aligned} \tag{29.11}$$

If we differentiate $\mathcal{E}\big(A, \phi^s(t)\big)$ with respect to t we get

$$\frac{d\mathcal{E}\big(\widehat{A}, \vec{\phi}(t)\big)}{dt} = \frac{1}{i\hbar}\langle\vec{\phi}(t) \mid [\widehat{A}, \widehat{H}\,]\vec{\phi}(t)\rangle. \tag{29.12}$$

[7] We consider observables which are not explicitly time dependent. The state is time dependent in the Schrödinger picture. As shown in §29.2 observables are time dependent in the Heisenberg picture.

Again we can see that the expectation value $\mathcal{E}\left(A, \phi^s(t)\right)$ is time-independent if $\widehat{A}$ commutes with the Hamiltonian.[8]

An example is that the momentum of a free particle is conserved. The momentum of a harmonic oscillator is not conserved. In classical mechanics an observable is conserved if its value remains unchanged in time. In quantum mechanics the requirement is weaker in that it is the expectation value, not an individual value, of the observable which is time-independent.[9]

29.2 The Heisenberg Picture

29.2.1 Unitary Evolution

In the Heisenberg picture, it is the observables which are time dependent. Following Postulate 29.1.2(TESP), it is natural to assume that time evolution of observables is again realisable as a unitary transformation of its corresponding operator. Such an assumption is stated below as a postulate.

Postulate 29.2.1(TEHP) on quantum evolution

(1) *The state is represented by a time-independent state vector.*

(2) *Observables which are not explicitly time dependent are represented by time-dependent selfadjoint operators, except for the Hamiltonian $\widehat{H}$ which remains time-independent.*[10] *Time evolution of these operators is determined by a continuous one-parameter group of unitary operators $\widehat{U}(\widehat{H}, t) = \exp(-\frac{i}{\hbar}\widehat{H}t)$ generated by the Hamiltonian, i.e., an initial operator $\widehat{A}(0)$ representing an arbitrary observable A will evolve into the operator*

$$\widehat{A}(t) := \widehat{U}^{\dagger}(\widehat{H}, t)\widehat{A}(0)\widehat{U}(\widehat{H}, t). \tag{29.13}$$

[8] *See* Eq. (29.19) in the Heisenberg picture.

[9] *See* Q29(6).

[10] The Hamiltonian $\widehat{H}$ is assumed to be not explicitly time dependent and is the same as the Hamiltonian in the Schrödinger picture in that the operator obtained by quantising the time-independent classical Hamiltonian.

Rewriting $\widehat{U}(\widehat{H}, t)$ as $\widehat{U}(t)$ for brevity Eq. (29.13) becomes

$$\widehat{A}(t) = \widehat{U}^{\dagger}(t)\widehat{A}(0)\widehat{U}(t). \tag{29.14}$$

The evolved operator $\widehat{A}(t)$ is a unitary transform of $\widehat{A}(0)$.[11] As an operator representing an observable the Hamiltonian should also evolves by Eq. (29.14), i.e.,

$$\widehat{H}(t) := \widehat{U}^{\dagger}(t)\widehat{H}\widehat{U}(t). \tag{29.16}$$

Since $\widehat{H}$ commutes with $\widehat{U}(t)$, we get $\widehat{H}(t) = \widehat{H}(0) = \widehat{H}$, i.e., the Hamiltonian remains time-independent, a result consistent with the assumption of the time-independence of the Hamiltonian of the postulate.

Postulate 29.2.1(TEHP) refers to the time evolution of observables and their associated selfadjoint operators. Non-selfgadjoint operators which do not represent any observables can also be time dependent if they are related to selfadjoint operators which do describe observables. Examples of this will be seen in Q35(13) in Problems and Exercises for Chapter 35.

29.2.2 Heisenberg Equation of Motion

In practical applications it is often more useful to describe the evolution in terms of differential equations involving the time derivatives of the operators. Differentiating Eq. (29.14) with respect to t we obtain

$$\frac{d\widehat{A}(t)}{dt} = \frac{d\widehat{U}^{\dagger}(t)}{dt}\widehat{A}(0)\widehat{U}(t) + \widehat{U}^{\dagger}(t)\widehat{A}(0)\frac{d\widehat{U}(t)}{dt}. \tag{29.17}$$

From

$$\frac{d\widehat{U}(t)}{dt} = -\imath\widehat{H}\widehat{U}(t), \quad \frac{d\widehat{U}^{\dagger}(t)}{dt} = \imath\widehat{H}\widehat{U}^{\dagger}(t) \tag{29.18}$$

we get

$$i\hbar\frac{d\widehat{A}(t)}{dt} = [\widehat{A}(t), \widehat{H}]. \tag{29.19}$$

[11] To conform to Definition 13.4.1(2), we can let $\widehat{V}(\widehat{H}, t) = \widehat{U}^{\dagger}(\widehat{H}, t)$. Then Eq. (29.14) becomes

$$\widehat{A}(t) = \widehat{V}(\widehat{H}, t)\widehat{A}(0)\widehat{V}^{\dagger}(\widehat{H}, t). \tag{29.15}$$

Since $\widehat{H} = \widehat{H}(t)$, we can rewrite the above equation as

$$i\hbar \frac{d\widehat{A}(t)}{dt} = [\widehat{A}(t), \widehat{H}(t)]. \tag{29.20}$$

This is called the **Heisenberg equation of motion**.[12] The commutator in the above equation is a unitary transform of the corresponding commutator at $t = 0$, i.e.,[13]

$$[\widehat{A}(t), \widehat{H}(t)] = \widehat{U}^{\dagger}(t)[\widehat{A}(0), \widehat{H}(0)]\widehat{U}(t), \tag{29.21}$$

since

$$\begin{aligned}
&\widehat{U}^{\dagger}(t)[\widehat{A}(0), \widehat{H}(0)]\widehat{U}(t) \\
&= \widehat{U}^{\dagger}(t)\widehat{A}(0)\widehat{H}(0)\widehat{U}(t) - \widehat{U}^{\dagger}(t)\widehat{A}(0)\widehat{H}(0)\widehat{U}(t) \\
&= \widehat{U}^{\dagger}(t)\widehat{A}(0)\widehat{U}(t)\widehat{U}^{\dagger}(t)\widehat{H}(0)\widehat{U}(t) - \widehat{U}^{\dagger}(t)\widehat{A}(0)\widehat{U}(t)\widehat{U}^{\dagger}(t)\widehat{H}(0)\widehat{U}(t) \\
&= \widehat{A}(t)\widehat{H}(t) - \widehat{H}(t)\widehat{A}(t).
\end{aligned} \tag{29.22}$$

The same applies to the commutator of two evolved operators, i.e.,

$$[\widehat{A}(t), \widehat{B}(t)] = \widehat{U}(t)^{\dagger}[\widehat{A}(0), \widehat{B}(0)]\widehat{U}(t). \tag{29.23}$$

To see how the Heisenberg equation of motion can be applied in practice, consider a particle in one-dimensional motion along the x-axis. In the Heisenberg picture, the quantised position and momentum operators $\widehat{x}$ and $\widehat{p}$ in $\vec{L}^2(I\!R)$ are taken as the position and momentum operators at $t = 0$, i.e., we have $\widehat{x}(0) = \widehat{x}$ and $\widehat{p}(0) = \widehat{p}$. These operators then evolve into $\widehat{x}(t)$ and $\widehat{p}(t)$ in accordance with Eq. (29.14). It follows that $\widehat{x}(0)^2$ and will evolve into $\widehat{x}(t)^2$ since

$$\begin{aligned}
\widehat{U}(t)^{\dagger}\widehat{x}^2\widehat{U}(t) &= \widehat{U}(t)^{\dagger}\widehat{x}\Big(\widehat{U}(t)\widehat{U}(t)^{\dagger}\Big)\widehat{x}\,\widehat{U}(t) \\
&= \big(\widehat{U}(t)^{\dagger}\widehat{x}\,\widehat{U}(t)\big)\big(\widehat{U}(t)^{\dagger}\,\widehat{x}\,\widehat{U}(t)\big).
\end{aligned} \tag{29.24}$$

Continue the process we can show that $\widehat{x}(0)^m = \widehat{x}^m$ will evolve into $\widehat{x}(t)^m$. Similarly $\widehat{p}(0)^n = \widehat{p}^n$ will evolve into $\widehat{p}(t)^n$.

[12] *See* Q35(3) for an application to non-selfadjoint operators.

[13] *See* P13.4.2(1) on the preservation of commutation relations under unitary transformation.

We can verify that results in §27.7 on various commutation relations apply to corresponding evolved operators, e.g., we have

$$[\,\widehat{x}(t),\,\widehat{p}(t)] = [\widehat{x}(0),\,\widehat{p}(0)] = i\hbar, \tag{29.25}$$

$$[\,\widehat{x}(t),\,\widehat{p}(t)^m] = i\hbar\, m\,\widehat{p}(t)^{(m-1)}, \tag{29.26}$$

$$[\,\widehat{p}(t),\,\widehat{x}(t)^n] = -i\hbar\, n\,\widehat{x}(t)^{(n-1)}. \tag{29.27}$$

Equations (27.106) to (27.107) become

$$[\,\widehat{x}(t),\,\widehat{A}\big(\widehat{x}(t),\,\widehat{p}(t)\big)] = i\hbar\frac{\partial\,\widehat{A}\big(\widehat{x}(t),\,\widehat{p}(t)\big)}{\partial\,\widehat{p}(t)}, \tag{29.28}$$

$$[\,\widehat{p}(t),\,\widehat{A}\big(\widehat{x}(t),\,\widehat{p}(t)\big)] = -i\hbar\frac{\partial\,\widehat{A}\big(\widehat{x}(t),\,\widehat{p}(t)\big)}{\partial\,\widehat{x}(t)}. \tag{29.29}$$

Now suppose the Hamiltonian operator obtained by quantising the classical Hamiltonian is of the form[14]

$$\widehat{H}\,\big(\widehat{x},\,\widehat{p}\,\big) := \sum_{m,n} c_{mn}\widehat{x}^{\,m}\widehat{p}^{\,n}. \tag{29.30}$$

This is the Hamiltonian in the Schrödinger picture and it is also the Hamiltonian is the Heisenberg picture. To apply Eq. (29.20), we need to express the Hamiltonian in terms of $\widehat{x}(t)^m$ and $\widehat{p}(t)^n$. This can be achieved by a examination of the evolved Hamiltonian in Eq. (29.16), i.e., we have

$$\begin{aligned}\widehat{H}\,(t) &= \widehat{U}^{\,\dagger}(t)\widehat{H}\,\big(\widehat{x},\,\widehat{p}\,\big)\widehat{U}\,(t) = \widehat{U}^{\,\dagger}(t)\Big(\sum_{m,n} c_{mn}\widehat{x}^{\,m}\widehat{p}^{\,n}\Big)\widehat{U}\,(t)\\ &= \sum_{m,n} c_{mn}\left(\widehat{U}^{\,\dagger}(t)\widehat{x}^{m}\widehat{U}\,(t)\right)\left(\widehat{U}^{\,\dagger}(t)\widehat{p}^{\,n}\widehat{U}\,(t)\right)\\ &= \sum_{m,n} c_{mn}\,\widehat{x}(t)^m\,\widehat{p}(t)^n = \widehat{H}\,\big(\widehat{x}(t),\,\widehat{p}(t)\big).\end{aligned} \tag{29.31}$$

On account of Eqs. (29.28) and (29.29) the Heisenberg equations for the position and momentum observables become

$$\frac{d\widehat{x}(t)}{dt} = \frac{\partial\,\widehat{H}\,(t)}{\partial\,\widehat{p}(t)}, \qquad \frac{d\,\widehat{p}(t)}{dt} = -\frac{\partial\,\widehat{H}\,(t)}{\partial\widehat{x}(t)}. \tag{29.32}$$

[14] Assuming that the position and momentum are arranged in an appropriate order, as commented in relation to Eq. (27.105).

These Heisenberg equations of motion resemble the classical Hamilton's equations of motion for $x_i(t)$ and $p_i(t)$ in Eq. (27.20). The quadratic form $\mathcal{E}\big(\widehat{A}(t), \vec{\phi}\,\big)$ at time t is equal to $\langle \vec{\phi} \mid \widehat{A}(t)\vec{\phi}\,\rangle$. Differentiating with respect to t, we get[15]

$$\frac{d\mathcal{E}\big(\widehat{A}(t), \vec{\phi}\,\big)}{dt} = \frac{1}{i\hbar}\,\langle \vec{\phi} \mid [\widehat{A}(t), \widehat{H}(t)]\,\vec{\phi}\,\rangle. \tag{29.33}$$

Definition 29.1.2(1) on constants of motion applies in the Heisenberg picture. The expectation value $\mathcal{E}\big(A(t), \phi^s\big) = \mathcal{E}\big(\widehat{A}(t), \vec{\phi}\,\big)$ is time-independent when $\widehat{A}(t)$ commutes with $\widehat{H}(t)$.[16]

In §35.3 we shall examine the motion of an oscillator in the Heisenberg picture which also serves as an example of how Heisenberg picture can be applied in practice.

29.3 Equivalence of the Schrödinger and the Heisenberg Pictures

Postulate 29.2.1(TEHP) does not really constitute a separate and independent postulate since the Heisenberg picture is unitarily related to the Schrödinger picture. For clarity let us label quantities in the Schrödinger and the Heisenberg pictures by the subscripts *Sch* and *Hei*, respectively. The unitary relationship can be established as follows:

(1) At time $t = 0$ the system is described by an initial state vector $\vec{\phi}(0)$ and operators $\widehat{A}(0)$, i.e., we have the same initial state vectors and operators in both pictures, i.e.,[17]

$$\vec{\phi}_{Sch}(0) = \vec{\phi}_{Hei}(0) = \vec{\phi}(0), \tag{29.34}$$

$$\widehat{A}_{Sch}(0) = \widehat{A}_{Hei}(0) = \widehat{A}(0). \tag{29.35}$$

[15] Using the Heisenberg equation of motion. We assume that the initial and the evolved state vectors are in the domains of relevant operators.

[16] Note that $[\widehat{A}(t), \widehat{H}(t)] = \widehat{0}$ follows from $[\widehat{A}(0), \widehat{H}(0)] = \widehat{0}$.

[17] The notation $\widehat{A}_{Sch}$ is used instead $\widehat{A}_S$ to avoid confusion with the notation for restrictions of operators to the Schwartz space shown in Eq. (17.48) and especially in Eq. (35.23) and in Q35(15).

(2) At a later time $t > 0$, these quantities will evolve as follows:
In the Schrödinger picture, we have

$$\vec{\phi}(0) = \vec{\phi}_{Sch}(0) \to \vec{\phi}_{Sch}(t) = \widehat{U}(t)\vec{\phi}_{Sch}(0) = \widehat{U}(t)\vec{\phi}(0), \quad (29.36)$$

$$\widehat{A}(0) = \widehat{A}_{Sch}(0) \to \widehat{A}_{Sch}(t) = \widehat{A}_{Sch}(0) = \widehat{A}(0). \quad (29.37)$$

In the Heisenberg picture, we have

$$\vec{\phi}(0) = \vec{\phi}_{Hei}(0) \to \vec{\phi}_{Hei}(t) = \vec{\phi}_{Hei}(0) = \vec{\phi}(0); \quad (29.38)$$

$$\widehat{A}(0) = \widehat{A}_{Hei}(0) \to \widehat{A}_{Hei}(t) = \widehat{U}(t)^{\dagger}\widehat{A}_{Hei}(0)\widehat{U}(t) = \widehat{U}(t)^{\dagger}\widehat{A}(0)\widehat{U}(t). \quad (29.39)$$

(3) Equations (29.36) to (29.39) imply that

$$\vec{\phi}_{Sch}(t) = \widehat{U}(t)\vec{\phi}_{Hei}(t), \quad (29.40)$$

$$\widehat{A}_{Hei}(t) = \widehat{U}(t)^{\dagger}\widehat{A}_{Sch}(t)\widehat{U}(t). \quad (29.41)$$

These equations can be rewritten as

$$\vec{\phi}_{Sch}(t) = \widehat{U}(t)\vec{\phi}_{Hei}(t), \quad (29.42)$$

$$\widehat{A}_{Sch}(t) = \widehat{U}(t)\widehat{A}_{Hei}(t)\widehat{U}(t)^{\dagger}, \quad (29.43)$$

or

$$\vec{\phi}_{Hei}(t) = \widehat{U}(t)^{\dagger}\vec{\phi}_{Sch}(t), \quad (29.44)$$

$$\widehat{A}_{Hei}(t) = \widehat{U}(t)^{\dagger}\widehat{A}_{Sch}(t)\widehat{U}(t). \quad (29.45)$$

(4) The above results show that at every instant of time the two pictures are related by a simultaneous unitary transformation of both the state vector and operators. It follows from the discussion in Chapter 23 that the two pictures are physically equivalent. The expectation values in both pictures are the same for all time since

$$\langle \vec{\phi}_{Sch}(t) \mid \widehat{A}_{Sch}(t)\vec{\phi}_{Sch}(t)\rangle = \langle \vec{\phi}_{Hei}(t) \mid \widehat{A}_{Hei}(t)\vec{\phi}_{Hei}(t)\rangle. \quad (29.46)$$

The essence of a quantum theory lies in its experimental measurability. By themselves the operators and the state vector are mathematical quantities. Two different mathematical descriptions

are physically indistinguishable if they give rise to the same measurable values. In §35.3 we shall investigate the time evolution of a simple harmonic oscillator in both Schrödinger and Heisenberg pictures. Questions on whether it is the wave packet of the particle which is oscillating or the position operator of the particle which is oscillating is not meaningful since there is no physical means to answer such questions. While it is intuitively pleasing to visualise the motion of a harmonic oscillator in terms of an oscillator wave packet we should not take such intuitive visualisation literally as being physical. What is physical is that the oscillation of the position and momentum expectation values.

We can go on to generate new descriptions of time evolution by unitary transformations. One such new description is known as the Interaction picture which we shall discuss in the next section.

29.4 Interacting Systems and the Interaction Picture

29.4.1 Derivation

Consider the case of a system interacting with an external potential. The Hamiltonian at time $t = 0$ is of the form

$$\widehat{H} := \widehat{H}_0 + \widehat{H}', \tag{29.47}$$

where $\widehat{H}_0$ is the Hamiltonian of the system without the external potential. It is called the *free Hamiltonian*, and $\widehat{H}'$, the potential energy term due to external interaction, is referred to as the *interaction Hamiltonian*. The resulting Hamiltonian $\widehat{H}$ is more complex. It would be desirable to divide $\widehat{H}$ into two parts, one involving $\widehat{H}_0$ and the other one involving $\widehat{H}'$. This is the reason for introducing the following Interaction picture:

(1) Let us assume that initially at time $t = 0$ the system is described by state vector $\vec{\phi}_{Sch}(0)$ and operators $\widehat{A}_{Sch}(0)$ in the Schrödinger picture. At time $t > 0$ these quantities would evolve into $\vec{\phi}_{Sch}(t)$ and $\widehat{A}_{Sch}(t)$.

(2) Introduce a one-parameter group of unitary operators generated by the free Hamiltonian $\widehat{H}_0$, i.e.,

$$\widehat{U}(\widehat{H}_0, t) := e^{-\frac{i}{\hbar}\widehat{H}_0 t}. \tag{29.48}$$

(3) Define $\vec{\phi}_I(t)$ and $\widehat{A}_I(t)$ as the unitary transforms of $\vec{\phi}_{Sch}(t)$ and $\widehat{A}_{Sch}(t)$ by the unitary operators $\widehat{U}(\widehat{H}_0, t)$, i.e.,

$$\vec{\phi}_I(t) := \widehat{U}(\widehat{H}_0, t)^\dagger \vec{\phi}_{Sch}(t), \tag{29.49}$$

$$\widehat{A}_I(t) := \widehat{U}(\widehat{H}_0, t)^\dagger \widehat{A}_{Sch}(t)\widehat{U}(\widehat{H}_0, t), \tag{29.50}$$

or

$$\vec{\phi}_{Sch}(t) := \widehat{U}(\widehat{H}_0, t)\vec{\phi}_I(t), \tag{29.51}$$

$$\widehat{A}_{Sch}(t) := \widehat{U}(\widehat{H}_0, t)\widehat{A}_I(t)\widehat{U}(\widehat{H}_0, t)^\dagger. \tag{29.52}$$

What we have obtained is a new description of the time evolution of the system in terms of $\vec{\phi}_I(t)$ and $\widehat{A}_I(t)$. This description is called the **Interaction picture**. This new picture is mathematically different from the Heisenberg picture since the unitary operators $\widehat{U}(\widehat{H}_0, t)$ are generated by the free Hamiltonian $\widehat{H}_0$, not the total Hamiltonian $\widehat{H}$. As a result both the state vector $\vec{\phi}_I(t)$ and the operators $\widehat{A}_I(t)$ are time dependent as shown in the table below:

initial description at $t = 0$		later description at $t > 0$
$\vec{\phi}(0) = \vec{\phi}_{Sch}(0)$	$\longrightarrow$	$\vec{\phi}_{Sch}(t) \neq \vec{\phi}_{Sch}(0) = \vec{\phi}(0)$,
$\widehat{A}(0) = \widehat{A}_{Sch}(0)$	$\longrightarrow$	$\widehat{A}_S(t) = \widehat{A}_{Sch}(0) = \widehat{A}(0)$.
$\vec{\phi}(0) = \vec{\phi}_{Hei}(0)$	$\longrightarrow$	$\vec{\phi}_{Hei}(t) = \vec{\phi}_{Hei}(0) = \vec{\phi}(0)$,
$\widehat{A}(0) = \widehat{A}_{Hei}(0)$	$\longrightarrow$	$\widehat{A}_{Hei}(t) \neq \widehat{A}_{Hei}(0) = \widehat{A}(0)$.
$\vec{\phi}(0) = \vec{\phi}_I(0)$	$\longrightarrow$	$\vec{\phi}_I(t) \neq \vec{\phi}_I(0) = \vec{\phi}(0)$,
$\widehat{A}(0) = \widehat{A}_I(0)$	$\longrightarrow$	$\widehat{A}_I(t) \neq \widehat{A}_I(0) = \widehat{A}(0)$.

This new picture is physically equivalent to the Schrödinger and Heisenberg pictures.

29.4.2 Equations of Motion

Let us derive the equations of motion for the state and observables. To start with the free Hamiltonian $\widehat{H}_0$ is assumed to be time-independent as before. We can also check that $\widehat{H}_{0I}(t)$ given in accordance with Eq. (29.52) is time-independent since $\widehat{H}_0$ commutes with $\widehat{U}(\widehat{H}_0, t)$. In other words, the free Hamiltonian remains the same in both the Schrödinger and the Interaction pictures. Since the unitary transformation is generated by the free Hamiltonian, the evolution of observables is the same as the evolution of observables in the Heisenberg picture for the free system, e.g., their operators $\widehat{A}_I(t)$ satisfies the Heisenberg equation for the free Hamiltonian:

$$i\hbar\frac{d}{dt}\,\widehat{A}_I(t) = [\widehat{A}_I(t),\, \widehat{H}_{0I}(t)] = [\widehat{A}_I(t),\, \widehat{H}_0]. \tag{29.53}$$

The interaction Hamiltonian is generally time dependent since it may not commute with $\widehat{U}(\widehat{H}_0, t)$. Its evolution is given by

$$\widehat{H}'_I(t) = \widehat{U}^\dagger(\widehat{H}_0, t)\widehat{H}'(0)\widehat{U}(\widehat{H}_0, t). \tag{29.54}$$

Here $\widehat{H}'(0)$ is the interaction Hamiltonian at $t = 0$ and it is the same as the interaction Hamiltonian in the Schrödinger picture. The time dependence of the interaction Hamiltonian can often be obtained by solving its equation of motion

$$i\hbar\frac{d}{dt}\,\widehat{H}'_I(t) = [\widehat{H}'_I(t),\, \widehat{H}_{0I}(t)] = [\widehat{H}'_I(t),\, \widehat{H}_0]. \tag{29.55}$$

An example will be given in §35.3.3.

For the state we have the state vector satisfy the following Schrödinger-type equation with the interaction Hamiltonian:

$$i\hbar\frac{d}{dt}\,\vec{\phi}_I(t) = \widehat{H}'_I(t)\,\vec{\phi}_I(t). \tag{29.56}$$

This is arrived at by differentiating $\vec{\phi}_I(t)$ in Eq. (29.49):

$$\begin{aligned} i\hbar\frac{d}{dt}\vec{\phi}_I(t) &= i\hbar\frac{d}{dt}\left(\widehat{U}^\dagger(\widehat{H}_0, t)\vec{\phi}_{Sch}(t)\right) \\ &= i\hbar\left(\frac{d\widehat{U}^\dagger(\widehat{H}_0, t)}{dt}\,\vec{\phi}_{Sch}(t) + \widehat{U}^\dagger(\widehat{H}_0, t)\frac{d\vec{\phi}_{Sch}(t)}{dt}\right). \end{aligned} \tag{29.57}$$

The term $d\vec{\phi}_{Sch}(t)/dt$ is given by the Schrödinger equation

$$i\hbar \frac{d\vec{\phi}_{Sch}(t)}{dt} = (\widehat{H}_0 + \widehat{H}')\vec{\phi}_{Sch}(t). \tag{29.58}$$

It follows that the left-hand-side of Eq. (29.57) is equal to

$$\begin{aligned} & i\hbar \left(-\frac{\widehat{H}_0}{i\hbar}\widehat{U}^\dagger(\widehat{H}_0, t)\vec{\phi}_{Sch}(t) + \widehat{U}^\dagger(\widehat{H}_0, t)\frac{1}{i\hbar}(\widehat{H}_0 + \widehat{H}')\vec{\phi}_{Sch}(t) \right) \\ & = -\widehat{H}_0\widehat{U}^\dagger(\widehat{H}_0, t)\vec{\phi}_{Sch}(t) + \widehat{U}^\dagger(\widehat{H}_0, t)\widehat{H}_0\vec{\phi}_{Sch}(t) + \widehat{U}^\dagger(\widehat{H}_0, t)\widehat{H}'\vec{\phi}_{Sch}(t) \\ & = \widehat{U}^\dagger(\widehat{H}_0, t)\widehat{H}'\widehat{U}(\widehat{H}_0, t)\vec{\phi}_I(t) \\ & = \widehat{H}'_I(t)\,\vec{\phi}_I(t). \end{aligned} \tag{29.59}$$

We have used $\vec{\phi}_{Sch}(t) = \widehat{U}(\widehat{H}_0, t)\vec{\phi}_I(t)$. The evolution of the new state vector is seen to be governed by the interaction Hamiltonian.

In the Interaction picture we have succeeded in separating the total Hamiltonian into a free part and an interaction part with the former governing the evolution of observables and the latter governing the evolution of state. This new description is particularly useful when the evolution of the free system in the Heisenberg picture is already known. An example in §35.3.3 will illustrate how the calculations are carried out.

Exercises and Problems

Q29(1) Using the expression for the spectral decomposition of unitary operators in Eq. (21.6) and the result in Eq. (21.19), show that Eq. (29.5) can be obtained from the unitary evolution Eq. (29.10).

Q29(2) Verify Eq. (29.12) in the Schrödinger picture and Eq. (29.33) in the Heisenberg picture.

Q29(3) The Hamiltonian of a system at time $t = 0$ is given in terms of a pair of annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$ by

$$\widehat{H} = \left(\widehat{a}^\dagger\widehat{a} + \frac{1}{2}\right)\hbar\omega. \tag{29.60}$$

(a) Using the method of induction and the commutation relation of $\widehat{a}$ and $\widehat{a}^\dagger$, prove

$$\widehat{H}^n\,\widehat{a} = \widehat{a}\,(\widehat{H} - \hbar\omega)^n, \quad n = 0, 1, 2, 3, \cdots. \tag{29.61}$$

(b) Assuming that the time dependence of the annihilation and creation operators in the Heisenberg picture are given by the Heisenberg equation of motion in Eq. (29.20), show that

$$\widehat{a}_{Hei}(t) = \widehat{a}_{Hei}(0)\,e^{-i\omega t}, \quad \widehat{a}^\dagger_{Hei}(t) = \widehat{a}^\dagger_{Hei}(0)\,e^{i\omega t}. \tag{29.62}$$

(c) The annihilation operator in the Schrödinger picture, denoted by $\widehat{a}_{Sch}(t)$, is related to $\widehat{a}_{Hei}(t)$ by

$$\widehat{a}_{Sch}t) = e^{\widehat{H}t/i\hbar}\,\widehat{a}_{Hei}(t)\,e^{-\widehat{H}t/i\hbar}. \tag{29.63}$$

By expanding the exponential in term of a series, i.e.,[18]

$$e^{\widehat{H}t/i\hbar} = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{t}{i\hbar}\right)^n \widehat{H}^n, \tag{29.64}$$

and using Eq. (29.61) show that

$$e^{\widehat{H}t/i\hbar}\,\widehat{a}_{Hei}(t) = \widehat{a}_{Hei}(t)\,e^{i\omega t}\,e^{\widehat{H}t/i\hbar}. \tag{29.65}$$

Hence verify explicitly that $\widehat{a}_{Sch}(t)$ is time-independent.

Q29(4) Let $\widehat{A}$ be the selfadjoint operator representing an observable A in the Schrödinger picture and let $\widehat{F}^{\widehat{A}}(\tau)$ be its spectral function. Let the corresponding operator at time t in the Heisenberg picture be denoted by $\widehat{A}_{Hei}(t)$ and let $\widehat{F}^{\widehat{A}_{Hei}(t)}(\tau)$ be its spectral function. How are these two spectral functions related and how are the probability distribution functions generated by these spectral functions in a given state related?

Q29(5) The Hamiltonian of a quantum particle of mass m in one-dimensional motion along the x-axis with potential energy $V(\widehat{x})$ is given by

$$\widehat{H} = \frac{1}{2m}\,\widehat{p}^{\,2} + V(\widehat{x}). \tag{29.66}$$

[18] Assuming an appropriate domain on which the expansion is valid.

Assuming $V(\widehat{x})$ to be a polynomial function of $\widehat{x}$ show that in the Heisenberg picture, we have[19]

$$m\frac{d}{dt}\langle\vec{\phi} \mid \widehat{x}\vec{\phi}\rangle = \langle\vec{\phi} \mid \widehat{p}\,\vec{\phi}\rangle, \tag{29.67}$$

$$\frac{d}{dt}\langle\vec{\phi} \mid \widehat{p}\vec{\phi}\rangle = -\langle\vec{\phi} \mid \frac{dV(\widehat{x})}{d\widehat{x}}\vec{\phi}\rangle. \tag{29.68}$$

Establish the following **Ehrenfest's theorem**:

$$m\frac{d^2}{dt^2}\langle\vec{\phi} \mid \widehat{x}\vec{\phi}\rangle = \langle\vec{\phi} \mid \widehat{F}\vec{\phi}\rangle, \quad \text{where } \widehat{F} = -\frac{dV(\widehat{x})}{d\widehat{x}}. \tag{29.69}$$

Discuss the physical significance of this result,

Q29(6) Discuss the fundamental differences in the concept of constants of motion in classical and quantum mechanics.

[19] The subscript *Hei* for Heisenberg picture quantities is omitted for brevity. We also assume that $\vec{\phi}$ is in an appropriate domain of all the operators involved.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 30

State after Measurement

30.1 Discrete Observables

30.1.1 Postulate (PPDO)

The disturbance during a measurement on a classical system can be made arbitrarily small. One may then assume that a measurement does not change the state, e.g., the radar used by police to measure the speed of a car has a negligible effect on the speed of the car. Classical dynamics also applies to the measuring process as well. For quantum systems Postulate 26.1(OV) tells us that a measurement of a discrete observable A will yield an eigenvalue of its corresponding operator $\widehat{A}$. Let the initial state vector be $\vec{\phi}_i$ at time $t = 0$ when the measurement begins and let the final state vector be $\vec{\phi}_f$ at time $t = \Delta t$ when the measurement ends. A natural question to ask is

> *Can disturbance during a measurement on the system be made arbitrarily small so that the state can remain the same, and if not, does the initial state evolve into the final state in a unitary manner describable in terms of* Postulate 29.1.2(TESP)*?*

The answers to these questions are complicated since there are different types of quantum systems and many different kinds of measurements. We shall confine ourselves to ideal measurements

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

introduced in relation to QMP5.3(3). For ideal measurements the answer to the above question is given by what is generally known as the *von Neumann's projection postulate* or simply the **projection postulate** on the state after measurement.[1] We shall consider discrete and continuous observables separately in terms of two separate postulates.[2]

Postulate 30.1.1(PPDO) *A measurement of a discrete observable A projects the initial state vector $\vec{\phi}_i$ onto the eigensubspace $\vec{\mathcal{S}}^{\hat{A}}(a_m)$ of the measured eigenvalue a_m of $\widehat{A}$, i.e., the final state vector is given by*

$$\vec{\phi}_f = c\, \widehat{P}^{\hat{A}}(a_m)\vec{\phi}_i, \tag{30.1}$$

where $\widehat{P}^{\hat{A}}(a_m)$ is the eigenprojector onto the eigensubspace $\vec{\mathcal{S}}^{\hat{A}}(a_m)$ and c is a normalisation constant.[3]

The transition from $\vec{\phi}_i$ to $\vec{\phi}_f$ is known as the **reduction of the state vector** or **collapse of the wave packet**.[4] There are two cases to consider: (1) when the measured eigenvalue is nondegenerate and (2) when the measured eigenvalue is degenerate.

30.1.2 Nondegenerate Eigenvalues

Suppose the measured eigenvalue a_ℓ is nondegenerate, and it corresponds to the unit eigenvector $\vec{\varphi}_\ell$. The initial state vector is expressible as a linear combination of the eigenvectors, i.e., $\vec{\phi} = \sum_\ell c_\ell \vec{\varphi}_\ell$. The transition to the final state in a measurement amounts to the projection of the initial state $\vec{\phi}_i$ onto $\vec{\varphi}_\ell$ in the linear combination. This can be visualised in Fig. 30.1.[5]

Such a reduction of the state vector has a number of physical implications:

C30.1.2(1) An initial state remains unchanged during the measurement if it is an eigenstate of the observable being measured, e.g.,

---

[1] von Neumann Chapters IV, V and VI.

[2] These postulates can be regarded as a definition of ideal measurements.

[3] Here $\widehat{P}^{\hat{A}}(a_m)$ is the eigenprojector given by Eq. (20.18). The normalisation constant is given by $c = ||\widehat{P}^{\hat{A}}(a_m)\vec{\phi}_i||^{-1}$. It is assumed that $||\widehat{P}^{\hat{A}}(a_m)\vec{\phi}_i|| \neq 0$. *See* C30.2.1(4) for a discussion of this point.

[4] This "collapse" is demonstrated by a position measurement shown in Eq. (30.6).

[5] For nondegenerate eigenvalues the projector in Eq. (30.1) is given by Eq. (20.16).

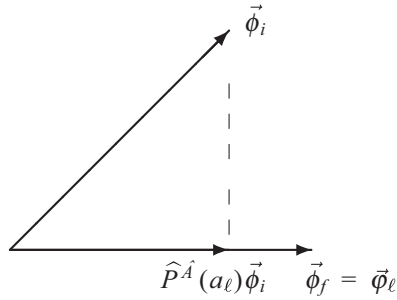


Figure 30.1 Measurement resulting in a change of $\vec{\phi}_i$ to $\vec{\phi}_f = \vec{\varphi}_\ell$.

if $\vec{\phi}_i = \vec{\varphi}_\ell$ then a measurement of A would yield the value a_ℓ and the state immediately after the measurement will be $\vec{\phi}_f = \vec{\varphi}_\ell = \vec{\phi}_i$ in accordance with Postulate 30.1.1(PPDO). This is consistent with the notion that the observable *possesses* a value in an eigenstate, a concept first discussed in §5.2.

C30.1.2(2) An arbitrary state changes during a measurement. The state of the system will become an eigenstate of A given by the eigenvector $\vec{\varphi}_\ell$ of $\widehat{A}$ belonging to the measured eigenvalue a_ℓ immediately after the measurement. This result is independent of what the initial state is.[6] This also means that a *final state does not determine the initial state*.

C30.1.2(3) A measurement can be a method to bring a system into an eigenstate from an unknown initial state, i.e., the measured eigenvalue a_ℓ tells us that the final state must be $\vec{\varphi}_\ell$. Hence,

> *an ideal measurement can be regarded as a method of preparing a system in a specified state.*

This is valid only for a measured eigenvalue which is not degenerate. When the measured eigenvalue is degenerate, the final state as given by Eq. (30.1) would depend on the initial state, as shown graphically

[6] This is the original von Neumann projection postulate. Equation (30.1) is valid for degenerate cases. It was Lüders who extends the postulate to include degenerate cases. Hence Eq. (30.1) is also referred to as *Lüder's rule*.

in Fig. 30.2. Generally we have to separate a measurement process from a state preparation process.[7]

C30.1.2(4) Since the probability of a measurement resulting in the value a_ℓ is equal to $|c_\ell|^2$, we can interpret $|c_\ell|^2$ as

> *the probability of finding the system initially in state $\vec{\phi}_i$ to be in the eigenstate $\vec{\varphi}_\ell$ after a measurement of A resulting in the value a_ℓ.*[8]

In line with the notion of position and momentum probability amplitudes introduced in §28.3 we also call c_ℓ the corresponding *probability amplitude*.

C30.1.2(5) The projection postulate together with Postulate 28.1(PDDO) implies the *reproducibility* of measurement. Suppose we make a measurement of A which results in the value a_ℓ. The state vector immediately after the measurement is $\vec{\varphi}_n$. Then an immediate repetition of the measurement would yield the same value. This is consistent with ideal measurements.

C30.1.2(6) *The projection postulate is consistent with the quantum property* QMP5.3(1) *that not all physical observables can be measured simultaneously.* Suppose a measurement projects the initial state vector onto an eigenvector of the observed quantity. Then the simultaneous measurement of another observable is possible only if this measurement will also project the initial state onto the same vector as the first measurement does. For two observables to be generally measurable simultaneously to yield a pair of eigenvalues they must have a complete set of common eigenvectors. As we know this is the case if and only if the two operators corresponding to the two observables commute with each other. Thus the necessary and sufficient condition for the *general simultaneous measurability* of two or more observables of a quantum system is that the corresponding operators commute. Such observables

[7] Wan pp. 300–329. Isham p. 134.

[8] Here c_n are the coefficients of expansion of the initial state in terms of the eigenvectors $\vec{\varphi}_\ell$ of $\widehat{A}$, i.e., $\vec{\phi}_i = \sum_\ell c_\ell \vec{\varphi}_\ell$.

are compatible in the sense of Definition 28.4.1(1).[9] For example, the z-components $\widehat{L}_z$ of the orbital angular momentum and the square of the total orbital angular momentum $\widehat{L}^2$ can be measured simultaneously while the x and y components of an electron spin cannot be measured simultaneously. All this is also consistent with Postulate 28.2(PD) in that not all observables are compatible.

The projection postulate is highly controversial.[10] It tells us what happens to the initial state on a measurement without any reference to the interaction between the quantum system and the measuring device. The postulate has nothing to say about the nature of such interaction and there is no equation of motion for the evolution of the state during measurement.[11] Since the same initial state can result in many different final states, depending on the measured values, the transition from the initial state to all these possible final states cannot be described as a usual quantum evolution in terms of the Schrödinger picture.[12] The projection postulate is the source of what is known as the *measurement problem* and the *non-locality problem* in quantum mechanics. A great deal of research has gone into the study of these problems. Some of these problems are discussed in §34.5, §34.6 and §34.7.

30.1.3 Degenerate Eigenvalues and Propositions

Suppose the measured eigenvalue a_1 of $\widehat{A}$ is degenerate corresponding to two orthonormal eigenvectors $\vec{\varphi}_{11}$ and $\vec{\varphi}_{12}$ of $\widehat{A}$. The final state $c\widehat{P}^{\hat{A}}(a_1)\vec{\phi}_i$ would depend on the initial state, e.g., $\vec{\phi}_i = \vec{\varphi}_{11}$ would lead to $\vec{\phi}_f = \vec{\varphi}_{11}$ and $\vec{\phi}_i = \vec{\varphi}_{12}$ would lead to $\vec{\phi}_f = \vec{\varphi}_{12}$. In other words, the projection postulate cannot be used to produce a desired final state from an unknown initial state in such cases. An illustration is shown in Fig. 30.2. The figure shows that an initial

[9] In the intermediate case, e.g., $\widehat{L}_x$ and $\widehat{L}_y$, the observables are not regarded as generally simultaneously measurable, since they possess simultaneous values only in a limited number of eigenstates.

[10] *See* §34.7.1 for more discussion.

[11] There is no reference to the duration of the interaction so that the reduction of state is often taken to be instantaneous. This would make it impossible to formulate any equation of motion for the interaction.

[12] *See* §31.1, §31.3 and §34.7 for more discussion on this. In the Schrödinger picture an initial state will lead to a single final state.

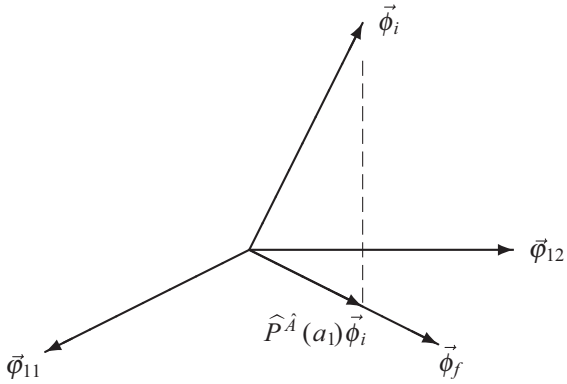


Figure 30.2 Measurement resulting in a change of state.

state is projected onto the eigensubspace of the eigenvalue a_1 which is spanned by $\vec{\varphi}_{11}$ and $\vec{\varphi}_{12}$.

An example relevant to the discussion on continuous observables in the next section is a proposition P and its corresponding projector $\widehat{P}$. If $\widehat{P}$ is 1-dimensional then the eigenvalue 1 is nondegenerate but the eigenvalue 0 is degenerate. If $\widehat{P}$ is N-dimensional, where N can range from 2 upwards, then both eigenvalues 1 and 0 are degenerate. A yes answer to the yes-no experiment to measure the proposition will cause an initial state $\vec{\phi}_i$ to change to the final state $\vec{\phi}_f = c\,\widehat{P}\vec{\phi}_i$ which is an eigenvector of $\widehat{P}$ corresponding to the eigenvalue 1.

30.2 Continuous Observables

30.2.1 Postulate (PPCO) and Postulate (PP)

For a continuous observable A represented by the operator $\widehat{A}$, a measurement will generally result in a value τ lying in a Borel set Λ. The change of the initial state $\vec{\phi}_i$ to the final state $\vec{\phi}_i$ is given by the following postulate:

Postulate 30.2.1(PPCO) *A measurement of a continuous observable A which results in a value τ lying in a Borel set Λ projects the initial state vector $\vec{\phi}_i$ onto the spectral subspace $\vec{S}^{\hat{A}}(\Lambda)$ of $\widehat{A}$, i.e., the*

final state, is given by

$$\vec{\phi}_f = c\, \widehat{M}^{\hat{A}}(\Lambda)\vec{\phi}_i, \tag{30.2}$$

where $\widehat{M}^{\hat{A}}(\Lambda)$ *is the corresponding spectral projector of* $\widehat{A}$ *and c is the normalisation constant given by*[13]

$$c = ||\widehat{M}^{\hat{A}}(\Lambda)\vec{\phi}_i||^{-1}. \tag{30.3}$$

The following comments aim to clarify the postulate:

C30.2.1(1) As in the case of propositions discussed earlier, the final state vector is an eigenvector of the spectral projector corresponding to the eigenvalue 1. In other words, a measurement of a continuous observable represented by a selfadjoint operator $\widehat{A}$ which yields a value in a Borel Λ will carry the initial state into an eigenvector of the spectral projector $\widehat{M}^{\hat{A}}(\Lambda)$ corresponding to the eigenvalue 1. For practical applications we are more interested in a measured value lying in an interval $\Lambda = (\tau_1, \tau_2]$. Then the spectral projector is expressible in terms of the spectral function by

$$\widehat{M}^{\hat{A}}(\Lambda) = \widehat{F}^{\hat{A}}(\tau_2) - \widehat{F}^{\hat{A}}(\tau_1). \tag{30.4}$$

C30.2.1(2) Postulate 30.2.1(PPCO) is applicable to discrete observables. From the discussion in §20.3, we know that for a discrete observable A we have[14]

$$\widehat{M}^{\hat{A}}\big((\tau_1, \tau_2]\big) = \widehat{P}^{\hat{A}}(a_m), \tag{30.5}$$

when Λ contains a single eigenvalue a_m of $\widehat{A}$. Then Eq. (30.2) reduces to Eq. (30.1). It follows that we may consider Postulate 30.1.1(PPDO) as a special case of Postulate 30.2.1(PDCO). In other words, we can incorporate Postulate 30.1.1(PPDO) into Postulate 30.2.1(PPCO) and rename Postulate 30.2.1(PDCO) simply as

Postulate 30.2.1(PP).

---

[13] As for Postulate 30.1.1(PPDO) it is assumed that $||\widehat{M}^{\hat{A}}(\Lambda)\vec{\phi}_i|| \neq 0$ here. *See* also Blank, Exner and Havlíček p. 257.

[14] *See* Eq. (20.17) in particular.

C30.2.1(3) When we discuss the reduction of state vectors in a measurement we assume that the measurement produces a *positive result* in the sense that $\widehat{M}^{\hat{A}}(\Lambda)\vec{\phi}_i \neq 0$. Equations (30.2) and (30.3) do not apply when $\widehat{M}^{\hat{A}}(\Lambda)\vec{\phi}_i = 0$ since the application of Eq. (30.2) would mean the annihilation of the initial state vector.[15] This corresponds to a zero probability of the measurement getting a value in the Borel set Λ. We shall say that the measurement produces a *negative result*. There are two cases:

(1) When $\vec{\phi}_i$ lies in the orthogonal complement of the subspace $\vec{\mathcal{S}}^{\hat{A}}(\Lambda)$ a measurement of $\widehat{M}^{\hat{A}}(\Lambda)$ would produce a negative result. For a positive result, we would measure $\widehat{M}^{\hat{A}}(\Lambda^{\perp})$. This is similar to the situation for a discrete observable represented by operator $\widehat{A}$. If the initial state vector is an eigenvector φ_ℓ of $\widehat{A}$ a measurement of the proposition represented by the eigenprojector $\widehat{P}(a_m)$ would have a negative result if $a_m \neq a_\ell$.

(2) When Λ lies outside the spectrum of the observable a measurement of A would produce a negative result.

30.2.2 Position Measurement

Let us examine position measurements of a particle in one-dimensional motion along the x-axis. Physically a position measurement is usually made using a *detector*, such as a Geiger counter. In §34.7 we shall discuss how a detector can be modelled and how a local position measuring process can be accomplished. Here we will confine ourselves to the discussion on the desired properties of detectors and the application of Postulate 30.2.1(PP). Using Geiger counters as a model we would endow a detector with the following properties:

P30.2.2(1) A detector is designed to ascertain the presence or otherwise of a particle. When a particle's presence is detected the device would *fire* or *click*, and otherwise it would not. In other words,

[15] Blank, Exner and Havliček p. 257. A similar comment applies to Eq. (30.1) for discrete observables.

a detector yields only two results, a positive one and a negative one, during a run of a measurement. When it clicks the particle is detected. We would quantify this result with the value 1. When it does not click the particle is not detected. We would quantify this result with the value 0.

P30.2.2(2) A detector has a *size* in the form of a finite interval Λ. When it clicks the particle is detected within the interval. An immediate repetition of the detection process would produce the same result.

P30.2.2(3) The detector is ineffective if the state vector is given by a wave function lying outside the interval Λ, i.e., if $\phi(x) = 0, \ \forall \in \Lambda$, in the sense that it would return a negative result for every run of the measurement.

P30.2.2(4) A detector of finite size cannot measure the position observable represented by the position operator $\widehat{x}$ in $\vec{L}^2(I\!R)$. An infinite array of detectors is needed to measure the position observable. This is practically impossible to achieve.[16] The question then arises as to what observable an individual detector of size Λ measures, if not the position observable $\widehat{x}$. The answer is

> *A detector of size Λ measures the position proposition (local position observable) described by the spectral projector $\widehat{M}^{\hat{x}}(\Lambda)$ and the converse is also true.*[17]

In other words, a detector of size Λ performs a *yes-no experiment* of the local position observable represented by the spectral projector $\widehat{M}^{\hat{x}}\big((\tau_1, \tau_2]\big)$. The statement that the yes-no experiment yields the value 1 is equivalent to saying that a position measurement results in a value in the interval $(\tau_1, \tau_2]$. After such a measurement the state vector would correspond to a wave function obtained by truncating the initial wave function outside the interval $(\tau_1, \tau_2]$, i.e.,

[16] It can be argued that the measurement is impossible in principle since the construction of an infinite array of detectors would require an infinite amount of resources in terms of materials and man-hours.

[17] Local position observables are introduced in §26.2.2.

the measurement causes the transition from an initial wave function $\phi_i(x)$ to the final wave function $\phi_f(x)$ given by[18]

$$\begin{aligned}\phi_i(x) \quad \rightarrow \quad \phi_f(x) &= c\,\chi_{(\tau_1,\tau_2]}(x)\phi_i(x) \\ &= c \begin{cases} \phi_i(x), & x \in (\tau_1, \tau_2] \\ 0, & x \notin (\tau_1, \tau_2] \end{cases}, \end{aligned} \tag{30.6}$$

where c is a normalisation constant. A further position measurement would have a probability 1 of finding the particle in $(\tau_1, \tau_2]$ again.

P30.2.2(5) A similar analysis applies to the momentum operator. The parallel is most apparent when working in the momentum representation. The wave function in the momentum representation after a momentum measurement resulting in a value in the $(\tau_1, \tau_2]$ in the momentum space is given in accordance with Eq. (20.30) by

$$c\,\underset{\sim}{\widehat{\chi}}_{(\tau_1,\tau_2]}(p)\,\underset{\sim}{\phi}(p). \tag{30.7}$$

where $\underset{\sim}{\phi}(p)$ is the Fourier transform of $\phi(x)$ and c is a normalisation constant.

30.3 Complete Sets of Compatible Observables

30.3.1 Discrete Observables

The concept of compatibility of two observables in Definition 28.4.1(1) is related to the simultaneous measurability of the two compatible discrete observables. This concept is also linked to the **successive measurements** of the two observables. By definition two compatible discrete observables A, B would have a complete set of states in which both observables would have a value. Let us label these states by $\varphi^s_{\ell,m}$. As shown in Eq. (28.26), observable A has the value a_ℓ and observable B has the value b_m in state $\varphi^s_{\ell,m}$. These values can be obtained by successive measurements of A and B irrespective of the order of the measurements:

(1) When the system is in state $\varphi^s_{\ell,m}$ a measurement of A would yield the value a_ℓ while leaving the state $\varphi^s_{\ell,m}$ unchanged

[18] $\widehat{M}^x(\Lambda)$ is defined by the characteristic function $\chi_\Lambda(x)$ by Eq. (20.28).

in accordance with the projection postulate. A subsequent measurement of B would yield the value b_m.

(2) When the system is in state $\varphi^s_{\ell,m}$ a measurement of B would yield the value b_n leaving the state $\varphi^s_{\ell,m}$ unchanged. A subsequent measurement of A would yield the value a_ℓ.

The conclusion is that for two compatible discrete observables there exists a complete set of states in which successive measurement of the two observables would yield the same results irrespective of the order of the measurements are carried out.[19]

Starting with a given observable A we may be able to find another observable B which is independent and compatible with A. We can go to try to obtain a third observable C which is compatible of both A and B. The process can be repeated to obtain a set of such mutually compatible and independent observables. For a given system, there may exists a maximal set of such observables.[20] By the definition of compatibility these observables correspond to commuting selfadjoint operators. Following Definition 20.6(2), we can define a complete set of compatible observables.

Definition 30.3.1(1) *A set of discrete observables $A, B, C, \ldots$ associated with a complete set of commuting selfadjoint operators $\widehat{A}, \widehat{B}, \widehat{C} \ldots$ of a discrete spectrum is called a complete set of discrete observables*.

The physical significance of such a set of observables is seen in the following comments:

C30.3.1(1) A simultaneous measurement of this set of observables can be made. A set of values $a_i, b_j, c_k \ldots$ will be obtained. These are eigenvalues of $\widehat{A}, \widehat{B}, \widehat{C} \ldots$.

C30.3.1(2) These eigenvalues determine a single unit eigenvector $\vec{\varphi}_{ijk\ldots}$ as shown in Eq. (20.62).

C30.3.1(3) By the projection postulate this eigenvector will be the state vector after the measurement.

---

[19] Zettili p. 168. Blank, Exner and Havlicek p. 274. Isham pp. 139–141.
[20] This is the same argument as in QMP5.3(1).

C30.3.1(4) We can conclude that a complete set of compatible discrete observables can determine a state.[21]

A complete set can consist of a single observable, i.e., an observable described by a nondegenerate selfadjoint operator constitutes a complete set by itself. Following the examples in terms operators serve illustrate the concept:

E30.3.1(1) In $\vec{L}^2(\Lambda)$ the operator $\widehat{p}_\lambda(\Lambda)$ in Eq. (17.34) is nondegenerate. Hence $\widehat{p}_\lambda(\Lambda)$ for each value of λ forms a complete set in $\vec{L}^2(\Lambda)$. An eigenvalue of $\widehat{p}_\lambda(\Lambda)$ determines a single unit eigenvector. The same is true for the Hamiltonian $\widehat{K}^\infty(\Lambda)$ in Eq. (19.42) is nondegenerate. Complete sets are clearly not unique.

E30.3.1(2) In $\vec{L}^2(\Lambda)$ the operator $\widehat{K}_\lambda(\Lambda)$ in Eq. (19.46) is degenerate and does not form a complete set.

E30.3.1(3) In $\vec{L}^2(I\!R)$ The harmonic oscillator Hamiltonian $\widehat{H}_{ho}$ in Eq. (19.54) is nondegenerate. Hence it forms a complete set on its own.

E30.3.1(4) In $\vec{L}^2(\mathcal{S}_u)$ both the operators $\widehat{L}_z$ and $\widehat{L}^2$ are degenerate. Hence, neither constitutes a complete set. The two operators together do form a complete set.

30.3.2 Continuous Observables

As pointed out earlier a complete set of compatible observables in $\vec{L}^2(I\!R)$ can consist of a single observable, e.g., $\widehat{H}_{ho}$. It is natural to ask if the momentum also constitutes a complete set on its own in $\vec{L}^2(I\!R)$. Intuitively this should be the case since the spectrum of the momentum operator may be considered nondegenerate in that for each eigenvalues $p \in I\!R$ there corresponds to a single generalised eigenfunction $f_p(x)$ given by Eq. (18.12). A similar argument applies to the position observable. However, these generalised eigenfunctions do not give rise to well-defined vectors of finite norm in $\vec{L}^2(I\!R)$. It follows that Definition 20.6(2) cannot be applied. Instead we have to use Definition 20.6(3) to define a complete set

[21] This concept of a complete set of compatible observables and its role in the characterisation of states are first discussed in §5.3.

of commuting selfadjoint operators with a continuous spectrum. This enables us to state a definition without reference to whether the observables are discrete or continuous, i.e., it is valid for both discrete and continuous observables.

Definition 30.3.2(1) *A set of observables* $A,\ B,\ C,\ \ldots$ *associated with a complete set of commuting selfadjoint operators* $\widehat{A},\ \widehat{B},\ \widehat{C}\ \ldots$ *is called a complete set of observables*.

In $\vec{L}^2(I\!R)$ the position observable forms a complete set on its own and so does the momentum observable. In $\vec{L}^2(I\!R^2)$ a complete set would consist two observables, e.g., the position along the x and y axes or the momentum $p_x,\ p_y$ along the x and y axes.

A measurement of a complete set of continuous observables does not prepare a state in the same way it does for discrete observables. In Eq. (30.6) the final state vector $\vec{\phi}_f$ after a measurement of A resulting in a value in a given interval $(\tau_1, \tau_2]$ depends on the initial state vector $\vec{\phi}_i$, e.g., different initial state vectors may leads to linearly independent final state vectors.

Exercises and Problems

Q30(1) An electron spin is in state α_x^s.[22] What is the state and the state vector immediately after a measurement of the z-component spin resulting in the value $\hbar/2$?

Q30(2) Find the subspace associated with the projector $\widehat{F}^{\hat{x}}(x_2) - \widehat{F}^{\hat{x}}(x_1)$ on $\vec{L}^2(I\!R)$.

Q30(3) Consider a particle in one-dimensional motion along the x-axis. According to Postulate 30.2.1(PPCO), if a momentum measurement yields a value in the interval $(p_1,\ p_2]$ the state vector $\vec{\varphi}$ in the momentum representation right after the measurement must satisfy

$$\left(\widehat{\underset{\sim}{F}}^{\hat{p}}(p_2) - \widehat{\underset{\sim}{F}}^{\hat{p}}(p_1)\right)\vec{\underset{\sim}{\varphi}} = \vec{\underset{\sim}{\varphi}}, \tag{30.8}$$

where $\widehat{\underset{\sim}{F}}^{\hat{p}}(p)$ is the spectral function of the momentum operator given in Theorem 20.4.2(1), i.e., $\widehat{\underset{\sim}{F}}^{\hat{p}}(p)$ is defined

[22] *See* §14.1.1 for the theory for electron spin.

by the characteristic function $\underset{\sim}{\chi}_{(-\infty,\tau]}(p)$ of the interval $(-\infty, \tau]$ on the momentum space $\underset{\sim}{I\!R}$.

Express Eq. (30.8) in terms of $\underset{\sim}{\chi}_{(-\infty,\tau]}(p)$ and functions $\underset{\sim}{\varphi}(p)$ in $L^2(\underset{\sim}{I\!R})$. Give an example of such a function in the momentum space and its corresponding function in the coordinate space.

Chapter 31

Pure and Mixed States

31.1 Classical Mixtures of States

So far we have only considered states in Postulate 25.1(PS) which correspond to a maximal amount of information of the system. These are pure states φ^s describable by unit vectors $\vec{\varphi}$. Often we may only have a partial knowledge of a system. As an example, consider a less than ideal state preparation process which is unable to prepare the system in a desired pure state φ^s. Instead the state preparation process can only determine the system to within a set of possible pure states φ^s_ℓ, $\ell = 1, 2, \ldots$, namely the system may end up to be in a pure state φ^s_1, or φ^s_2 and so on. Suppose the state preparation process can also tell us the probabilities w_ℓ of the system ending up in these pure states. In other words, we do not know for certain which pure state the system is actually in. We only know that the system has a probability w_ℓ to be in pure state φ^s_ℓ. The system is then said to be in a **classical mixture of states**. In more abstract terms, the concept can be stated as

> *A classical mixture of states is a characterisation of a given quantum system in terms of a set of possible pure states $\vec{\varphi}_\ell$ together with a corresponding set of probabilities w_ℓ of their occurrence.*

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

To see how a classical mixture can arise, let us consider the measurement of a discrete observable A having a nondegenerate spectrum $\{a_\ell\}$ corresponding to eigenstates φ_ℓ^s of A. In accordance of Postulates 28.1(PDDO) and 30.1.1(PPDO) a measurement of A would result in different final eigenstates probabilistically, i.e., we have the following transition:

$$\phi_i^s \quad \rightarrow \quad \begin{cases} \varphi_1^s, \text{ with probability } w_1 = \wp^A(\phi_i^s, a_1) \\ \varphi_2^s, \text{ with probability } w_2 = \wp^A(\phi_i^s, a_2) \\ \varphi_3^s, \text{ with probability } w_3 = \wp^A(\phi_i^s, a_3) \\ \vdots \end{cases} . \tag{31.1}$$

To pursue the analysis more explicitly, suppose we are given a collection of N copies of a given system all in the same pure state ϕ_i^s, N being a large number. If we then perform a measurement of A on each of the systems in the collection we will end up with a collection of systems in which a number $N_\ell \leq N$ of the systems is in state φ_ℓ^s. The fraction N_ℓ/N is identifiable with the probability $\wp^A(\phi_i^s, a_\ell)$. We will say that a system in the collection after the measurement is in a *classical mixture* of states in the sense that we do not know which pure state the system is actually in.[1]

Let us consider what happens if we wish to calculate the expectation value of a bounded observable B represented by bounded operator $\widehat{B}$ in such a classical mixture of states.[2] The calculation would consist of two averaging processes:

(1) The expectation value of B in pure state φ_ℓ^s is[3]

$$\mathcal{E}(B, \varphi_\ell^s) = \mathcal{E}(\widehat{B}, \vec{\varphi}_\ell) = \langle \vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell \rangle. \tag{31.2}$$

(2) The average of the expectation values $\mathcal{E}(B, \varphi_\ell^s)$ over the set of pure states φ_ℓ^s with probabilities w_ℓ is

$$\sum_\ell w_\ell \, \mathcal{E}(B, \varphi_\ell^s) = \sum_\ell w_\ell \, \langle \vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell \rangle. \tag{31.3}$$

[1] *See* Isham pp. 74, 89, Merzbacher p. 365 for other interpretation of the term.

[2] Bounded observables are represented by bounded selfadjoint operators. There are no domain problems from bounded operators to complicate things.

[3] Here $\vec{\varphi}_\ell$ are the state vectors corresponding to states φ_ℓ^s. Since the observable is nondegenerate, the vectors $\vec{\varphi}_\ell$ form a complete orthonormal set.

The above sum can be arrived at in a single step in terms of a density operator. Let $\widehat{D}$ be the density operator obtained by the convex combination of the projectors $\widehat{P}_{\vec{\varphi}_\ell} = |\vec{\varphi}_\ell\rangle\langle\vec{\varphi}_\ell|$, with weights w_ℓ as described by Eq. (18.24), i.e.,[4]

$$\widehat{D} = \sum_\ell w_\ell \, \widehat{P}_{\vec{\varphi}_\ell}. \tag{31.4}$$

The properties of density operators shown in Eqs. (18.25) and (18.26) imply that the sum in Eq. (31.3) is equal to the trace of the operator $\widehat{B}\widehat{D}$, i.e.,

$$\mathrm{tr}\left(\widehat{B}\widehat{D}\right) = \sum_\ell w_\ell \, \langle\vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell\rangle. \tag{31.5}$$

This result shows that density operators can embody the properties of a pure state as well as a classical mixture of states:

(1) When a density operator is a projector it is a one-dimensional projector of the form $\widehat{P}_{\vec{\phi}}$ for some unit vector $\vec{\phi}$. Equation (31.5) reduces to

$$\mathrm{tr}\left(\widehat{B}\widehat{D}\right) = \mathrm{tr}\left(\widehat{B}\widehat{P}_{\vec{\phi}}\right) = \langle\vec{\phi} \mid \widehat{B}\vec{\phi}\,\rangle = \mathcal{E}\left(\widehat{B}, \vec{\phi}\,\right). \tag{31.6}$$

The probability distribution function in Eq. (28.9) can also be obtained directly from the density operator, i.e.,

$$\mathcal{F}^B(\phi^s, \tau) = \mathrm{tr}\left(\widehat{F}^{\widehat{B}}(\tau)\widehat{D}\right) = \mathrm{tr}\left(\widehat{F}^{\widehat{B}}(\tau)\widehat{P}_{\vec{\phi}}\right). \tag{31.7}$$

(2) When a density operator is not a projector it can describe a classical mixture of states. The density operator in Eq. (31.4) embodies all the information contained in the classical mixture, leading to the expectation value given by $\mathrm{tr}\left(\widehat{B}\widehat{D}\right)$.

All this suggests that we can incorporate mixtures into our description of states if we represent states generally by density operators. The transition in Eq. (31.1) can be interpreted as[5]

[4] A one-dimensional projector is a density operator.
[5] Isham pp. 133–138.

a transition from the pure state $\widehat{D}_i = \widehat{P}_{\vec{\phi}_i}$ *to the classical mixture described by a density operator* $\widehat{D}_f$*, i.e.,*

$$\widehat{D}_i = \widehat{P}_{\vec{\phi}_i} \quad \rightarrow \quad \widehat{D}_f := \sum_\ell \wp^{\hat{B}}(\phi_i^s, b_\ell)\, \widehat{P}_{\vec{\varphi}_\ell} \tag{31.8}$$

Alternatively we can say that

the projection postulate causes the transition of an initial pure state into a classical mixture of states.

31.2 Quantum Mixtures of States

It is tempting to assume that there are only two kinds of states, pure states described by unit vectors and classical mixtures of states described by density operators. This would be wrong. While a given classical mixture of states determines a density operator *the converse is not true*. Our previous discussion in §18.2 tells us that the decomposition of a density operator in terms of a convex combination of projectors is not unique.

Consider an example on electron spin.[6] Our present discussion tells us that the density operator $\widehat{D}_z$ in Eq. (18.29) can be used to describe a classical mixture of pure states $\vec{\alpha}_z$ and $\vec{\beta}_z$ with equal weights, i.e., a spin described by $\widehat{D}_z$ has an equal probability of being in the pure state α_z^s or β_z^s.[7] On the other hand the density operator $\widehat{D}_x$ in Eq. (18.30) will describe a classical mixture of pure states $\vec{\alpha}_x$ and $\vec{\beta}_x$ with equal weights. While the two classical mixtures are physically different the two density operators $\widehat{D}_z$ and $\widehat{D}_x$ are identical.[8]

If we are to extend Postulate 25.1(PS) to assume generally that a quantum state is described by a density operator $\widehat{D}$ with the expectation value of an observable A given by $\mathrm{tr}\left(\widehat{A}\widehat{D}\right)$, then we cannot insist that a state so described must be in any particular classical mixture. This is illustrated by the following examples:

[6] Isham pp. 91–94.

[7] This is the situation after a spin measurement along the z-direction when the initial spin state is α_x^s. *See* Q28(1) and Q29(1).

[8] The two mixtures are composed of different collections of spins.

E31.2(1) For electron spin with its state described by the density operator $\widehat{D}_z$, we cannot assert that the spin must be in either pure state α_z^s or pure state β_z^s since it may well be in pure state α_x^s or pure state β_x^s.

E31.2(2) For a general system with state described by the density operator in Eq. (31.4), we cannot assert that the system must be in one of the pure states $\vec{\varphi}_\ell$.

The above discussion shows that a quantum state described by a density operator is quite distinct from the corresponding situation in classical statistical mechanics where, despite our ignorance, the system is really in a certain definite pure state.[9] In other words, the interpretation of classical statistical states, known as the **ignorance interpretation**, does not generally apply to states described by density operators in quantum theory.[10]

It follows that if we are to describe quantum states by density operators we would include three classes of states:

Pure states These are states described by density operators which are projectors. The resulting projectors are one-dimensional. These states coincide with the states in Postulate 25.1(PS), i.e., they are pure states.

Classical mixture of states A density operator can also be used to describe a classical mixture.[11]

Quantum mixture of states These are states described by density operators which are not projectors. These states should generally be regarded as states in their own right and they should not be considered simply as a classical mixture of a certain set of pure states. To highlight this we shall generally call a state represented by a density operator which is not a projector a **quantum mixture of states** or a *quantum mixture*, or simply a **mixed state** for short.

---

[9] Pure states are states embodying a maximal possible knowledge of the system. In the case of a classical particle a pure state corresponds to a unique set of its position and momentum values.

[10] Beltrametti and Cassinelli pp. 6–13. D'Espagnat has a more detailed conceptual discussion.

[11] *See* P32.1(4) in Chapter 32 on superselection rules.

31.3 Changes of States

A state changes due to time evolution or due to measurement. We wish to know how these changes are described in terms of density operators.

The time evolution in terms of density operators can be formulated in the Schrödinger picture as follows.[12] Consider a pure state at time $t = 0$ described by unit vector $\vec{\phi}(0)$. The corresponding density operator is

$$\widehat{D}(0) = |\vec{\phi}(0)\rangle\langle\vec{\phi}(0)|. \tag{31.9}$$

The initial state vector $\vec{\phi}(0)$ will evolve into a new state vector $\vec{\phi}(t) = \widehat{U}(t)\vec{\phi}(0)$ given by Eq. (29.10). This evolved state vector defines a new density operator

$$\widehat{D}(t) = |\vec{\phi}(t)\rangle\langle\vec{\phi}(t)| = |\widehat{U}(t)\vec{\phi}(0)\rangle\langle\widehat{U}(t)\vec{\phi}(0)|. \tag{31.10}$$

This is the evolved density operator. In accordance with Eqs. (18.35) and (18.36), we can write down an explicit expression linking the initial to the evolved density operator as

$$\widehat{D}(t) = \widehat{U}(t)\widehat{D}(0)\widehat{U}^{\dagger}(t). \tag{31.11}$$

Differentiating the expression with respect to t we get[13]

$$i\hbar\frac{d\widehat{D}(t)}{dt} = [\widehat{H}, \widehat{D}(t)]. \tag{31.12}$$

It is natural that we extend the above results to the evolution of arbitrary mixed states. A formal statement to this effect will be given in the next section.

Apart from time evolution a state is generally changed during a measurement. The change of an initial state vector $\vec{\phi}_i$ to the final state vector $\vec{\phi}_f$ after a single measurement of a discrete

[12] *See* Jauch pp. 155–157 for a formulation in the Heisenberg picture.

[13] This equation is not the same as Eq. (29.20) in the Heisenberg picture for the evolution of observables.

observable is given by Postulate 30.1.1(PPDO), i.e., by Eq. (30.1). The corresponding the initial operator $\widehat{D}_i = |\vec{\phi}_i\rangle\langle\vec{\phi}_i|$ and the final operator $\widehat{D}_f = |\vec{\phi}_f\rangle\langle\vec{\phi}_f|$ are related by

$$\widehat{D}_f = c^2\, \widehat{P}^{\hat{A}}(a_\ell)\widehat{D}_i\, \widehat{P}^{\hat{A}}(a_\ell), \quad c = ||\widehat{P}^{\hat{A}}(a_\ell)\vec{\phi}_i||^{-1}, \tag{31.13}$$

since, by Eq. (18.42), we have[14]

$$\begin{aligned} |\vec{\phi}_f\rangle\langle\vec{\phi}_f| &= |c\, \widehat{P}^{\hat{A}}(a_\ell)\vec{\phi}_i\rangle\langle c\, \widehat{P}^{\hat{A}}(a_\ell)\vec{\phi}_i| \\ &= c^2\, \widehat{P}^{\hat{A}}(a_\ell)\, |\vec{\phi}_i\rangle\langle\vec{\phi}_i|\, \widehat{P}^{\hat{A}}(a_\ell). \end{aligned} \tag{31.14}$$

31.4 Postulates Incorporating Mixed States

Previous postulates formulated for pure states need to be modified to take account of mixed states. Postulates 26.1(OV) and 27.2(CQ) which involve only observables remains unchanged. A list of modified postulates, with prefix M for inclusion of mixed states, is given below where B denotes a bounded observable, $\widehat{B}$ denotes the bounded selfadjoint operator representing B, $\widehat{F}^{\hat{B}}(\tau)$ and $\widehat{M}^{\hat{B}}(\Lambda)$ denote the spectral function and spectral measure of $\widehat{B}$.[15]

Postulate 31.4(MS) *The state space of a quantum system is a complex Hilbert space $\vec{\mathcal{H}}$. A state of the system is describable by a density operator on $\vec{\mathcal{H}}$ and the set of all states corresponds one-to-one to the set of density operators on $\vec{\mathcal{H}}$.*[16]

The state described by a density operator $\widehat{D}$ is denoted by D^s. The state is pure if $\widehat{D}$ is a projector,[17] otherwise it is called a **quantum mixture of states** or a **mixed state**.[18]

Postulate 31.4(MPD) *When the system is in state D^s described by a density operator $\widehat{D}$ the probability distribution for the measured*

[14] In terms of $\widehat{D}_i$ we have $c^2 = \left(\mathrm{tr}\,(\widehat{P}^{\hat{A}}(a_\ell)\widehat{D}_i\,)\right)^{-1}$.

[15] Isham pp. 89–94.

[16] For orthodox quantum systems without superselection rules.

[17] This is true in the absence of superselectioin rule. *See* P32.1(4) in §32.1 on the effect superselection rule on the interpretation of a state described by a one-dimensional projector.

[18] We should stress again that a quantum mixture of $\vec{\varphi}_\ell$ does not imply an *ignorance interpretation* which is applicable to classical mixtures.

values of an observable B is determined by the probability distribution function $\mathcal{F}^B(D^s, \tau)$ given by

$$\mathcal{F}^B(D^s, \tau) := \mathrm{tr}\left(\widehat{F}^{\hat{B}}(\tau)\widehat{D}\right). \tag{31.15}$$

The corresponding probability measure is given by

$$\mathcal{M}^B(D^s, \Lambda) := \mathrm{tr}\left(\widehat{M}^{\hat{B}}(\Lambda)\widehat{D}\right). \tag{31.16}$$

As a selfadjoint operator a density operator $\widehat{D}$ is decomposable as a linear combination of its eigenprojectors by the spectral theorem. We have $\widehat{D} = \sum_\ell \omega_\ell \widehat{P}_{\vec{\varphi}_\ell}$ in accordance with Eq. (18.24), when the eigenvalues are all nondegenerate corresponding to a complete orthonormal set of eigenvectors $\vec{\varphi}_\ell$. Since the trace operation is linear as shown in Eqs. (18.25) and (18.26), we can rewrite Eqs. (31.15) and (31.16) as[19]

$$\mathcal{F}^B(D^s, \tau) = \sum_\ell \omega_\ell \mathcal{F}^{\hat{B}}(\vec{\varphi}_\ell, \tau), \tag{31.17}$$

$$\mathcal{M}^B(D^s, \Lambda) = \sum_\ell \omega_\ell \mathcal{M}^{\hat{B}}(\vec{\varphi}_\ell, \Lambda). \tag{31.18}$$

The expectation value of B in state D^s is then given by

$$\mathcal{E}(\widehat{B}, \widehat{D}) := \int_{-\infty}^{\infty} \tau \, d_\tau\left(\mathrm{tr}\left(\widehat{F}^{\hat{B}}(\tau)\widehat{D}\right)\right). \tag{31.19}$$

The above expectation value is equal to the trace of $\widehat{B}\widehat{D}$, i.e., we have

$$\mathcal{E}(\widehat{B}, \widehat{D}) = \mathrm{tr}\,(\widehat{B}\widehat{D}) = \sum_\ell \omega_\ell \langle \vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell \rangle. \tag{31.20}$$

Noting that $\langle \vec{\varphi}_\ell \mid \widehat{B}\vec{\varphi}_\ell \rangle$ is the expectation value of A in the pure state φ^s_ℓ we can rewrite the above equation as

[19] *See* Postulate 28.2(PDCO) and Eq. (28.9) for comparison and notation. On account of Eq. (18.16) we have

$$\mathrm{tr}\left(\widehat{F}^{\hat{B}}(\tau)\widehat{P}_{\vec{\varphi}_\ell}\right) = \langle \vec{\varphi}_\ell \mid \widehat{F}^{\hat{B}}(\tau)\vec{\varphi}_\ell \rangle = \mathcal{F}^{\hat{B}}(\vec{\varphi}_\ell, \tau).$$

$$\mathcal{E}(\widehat{B},\,\widehat{D}) = \sum_{\ell} \omega_\ell\, \mathcal{E}(\widehat{B},\,\widehat{P}_{\vec{\varphi}_\ell}). \tag{31.21}$$

These results are consistent with the concept of mixed states introduced by Eq. (31.1) and the discussion of their expectation values in Eqs. (31.2) and (31.3).

The time evolution of mixed state in the Schrödinger picture is given by the following postulate:

Postulate 31.4(MTESP)[20] *When the system is in a state described by a density operator* $\widehat{D}(0)$ *at time* $t = 0$ *its state at time t is given by the density operator*

$$\widehat{D}(t) = \widehat{U}\,(t)\widehat{D}(0)\widehat{U}\,(t)^\dagger, \tag{31.22}$$

where $\widehat{U}\,(t)$ *are evolution operators given by* Eq. (29.9)*. Observables which are not explicitly time dependent in their definition remain unchanged in time.*

Finally the projection postulate also needs to be modified.

Postulate 31.4(MPPCO) *A measurement of an observable B which results in a value* τ *lying in a Borel set* Λ *reduces the initial density operator* $\widehat{D}_i$ *to the final density operator* $\widehat{D}_f$ *given by*

$$\widehat{D}_f = \frac{1}{\mathrm{tr}\,(\widehat{M}^{\hat{B}}(\Lambda)\widehat{D}_i\,)}\, \widehat{M}^{\hat{B}}(\Lambda)\widehat{D}_i\widehat{M}^{\hat{B}}(\Lambda), \tag{31.23}$$

where $\widehat{M}^{\hat{A}}$ *is the spectral measure of* $\widehat{B}$.

31.5 Correlations and the Superposition Principle

The mixed state described by the density operator $\widehat{D}$ in Eq. (31.4) is fundamentally different from a pure state represented by the unit vector $\vec{\phi}$ given by a linear combination of $\vec{\varphi}_\ell$, i.e.,

$$\vec{\phi} = \sum_{\ell} c_\ell\, \vec{\varphi}_\ell, \quad c_\ell = \sqrt{\omega_\ell}. \tag{31.24}$$

[20] This is time evolution in the Schrödinger picture. In the Heisenberg picture, the density operator is time-independent while the operators are time dependent (*see* Jauch pp. 155–157).

We can distinguish the mixed state from the pure state since they lead to different expectation values for some observables. For the pure state we have $\mathcal{E}(\widehat{B}, \vec{\phi}\,) = \langle \vec{\phi} \mid \widehat{B}\,\vec{\phi}\,\rangle$. More explicitly we have

$$\mathcal{E}(\widehat{B}, \vec{\phi}\,) = \sum_{\ell} w_\ell \,\langle \vec{\varphi}_\ell \mid \widehat{B}\,\vec{\varphi}_\ell\rangle + \sum_{\ell \neq \ell'} c_\ell^* \, c_{\ell'} \,\langle \vec{\varphi}_\ell \mid \widehat{B}\,\vec{\varphi}_{\ell'}\rangle. \tag{31.25}$$

For the mixed state we have

$$\mathcal{E}(\widehat{B}, \widehat{D}) = \mathrm{tr}(\widehat{B}\widehat{D}) = \sum_{\ell} w_\ell \,\langle \vec{\varphi}_\ell \mid \widehat{B}\,\vec{\varphi}_\ell\rangle. \tag{31.26}$$

The expectation value of the pure state $\mathcal{E}(\widehat{B}, \vec{\phi}\,)$ contains additional terms which sum up to

$$C = \sum_{\ell \neq \ell'} c_\ell^* \, c_{\ell'} \,\langle \vec{\varphi}_\ell \mid \widehat{B}\,\vec{\varphi}_{\ell'}\rangle. \tag{31.27}$$

Some comments are warranted to clarify the situation[21]:

C31.5(1) The quantity C in Eq. (31.27) is called the **correlation term** or the **interference term** generated by operator $\widehat{B}$. We also call the individual terms $\langle \vec{\varphi}_\ell \mid \widehat{B}\,\vec{\varphi}_{\ell'}\rangle$ correlation terms or interference terms generated by operator $\widehat{B}$.[22] These terms represent the correlations between different constituent states in the linear combination generated by observable B. It is these correlation terms which distinguish the pure state ϕ^s in Eq. (31.24) from the quantum mixture D^s in Eq. (31.4).

C31.5(2) The pure state ϕ^s in Eq. (31.24) is said to be a **coherent superposition** of pure states φ_ℓ^s because of the existence of correlation terms in the expectation value.[23] Pure states are also said to be **coherent** and satisfy the **superposition principle** in that a linear combination of them would form a coherent superposition. A superposition is preserved in time since a superposed state in

[21] Some of these results and terminology have to be amended when we consider systems having a superselection rule. This will be discussed in §32.

[22] These are first introduced in Eq. (10.21).

[23] A linear combination is *not* a coherent superposition in the absence of correlation terms. This occurs for systems with superselection rules discussed in Chapter 32.

Eq. (29.3) will evolve into a similarly superposed state in Eq. (29.5). Generally if at time $t = 0$ we have

$$\vec{\phi}(0) = \sum_{\ell} c_{\ell}\ \vec{\varphi}_{\ell}(0), \tag{31.28}$$

then the evolved state vector in the Schrödinger picture is also a linear combination with coefficients c_{ℓ}, i.e.,

$$\vec{\phi}(t) = \widehat{U}(t)\vec{\phi}(0) = \sum_{\ell} c_{\ell}\,\vec{\varphi}_{\ell}(t), \quad \vec{\varphi}_{\ell}(t) = \widehat{U}(t)\vec{\varphi}_{\ell}(0). \tag{31.29}$$

C31.5(3) To generalise the concept of correlations further, let $\vec{\phi}$ and $\vec{\psi}$ be two arbitrary unit vectors and let $\widehat{P}_{\vec{\phi}} = |\vec{\phi}\,\rangle\langle\vec{\phi}\,|$ and $\widehat{P}_{\vec{\psi}} = |\vec{\psi}\,\rangle\langle\vec{\psi}\,|$ be the projectors generated by them. Following Eq. (31.27) we call $\langle\vec{\phi}\ |\ \widehat{P}_{\vec{\psi}}\ \vec{\psi}\,\rangle$ the correlation term between $\vec{\phi}$ and $\vec{\psi}$ generated by $\widehat{P}_{\vec{\psi}}$. Since $\langle\vec{\phi}\ |\ \widehat{P}_{\vec{\psi}}\ \vec{\psi}\,\rangle = \langle\vec{\phi}\ |\ \vec{\psi}\,\rangle$ we shall also call $\langle\vec{\phi}\ |\ \vec{\psi}\,\rangle$ the *correlation term* or simply the *correlation* between $\vec{\phi}$ and $\vec{\psi}$.[24] This terminology reflects the following interpretation. If the two vectors are not orthogonal, we can express one in terms of the other plus a remainder, i.e., we have

$$\vec{\phi} = c_1\vec{\psi} + c_2\vec{\psi}^{\perp}, \quad \text{where } \langle\vec{\psi}\ |\ \vec{\psi}^{\perp}\rangle = 0. \tag{31.30}$$

As mentioned in the discussion on Eq. (28.6), the probability of a measurement of the proposition $\widehat{P}_{\vec{\psi}}$ in state ϕ^s resulting in a yes answer is given by $\langle\vec{\phi}\ |\ \widehat{P}_{\vec{\psi}}\,\vec{\phi}\,\rangle = \big|\langle\vec{\phi}\ |\ \vec{\psi}\,\rangle\big|^2$. Since the projector $\widehat{P}_{\vec{\psi}}$ is the proposition that *the system is in state* ψ^s, we can interpret $\langle\vec{\phi}\ |\ \widehat{P}_{\vec{\psi}}\,\vec{\phi}\,\rangle$ as the probability of finding the system initially in state ϕ^s to be in state ψ^s on a measurement of the proposition.[25]

Alternatively, we can interpret $|\langle\vec{\phi}\ |\ \vec{\psi}\,\rangle|^2 = |\langle\vec{\psi}\ |\ \vec{\phi}\,\rangle|^2$ as the *transition probability* between states $\vec{\phi}$ and ψ,[26] or the probability that the ψ "occurs" in state $\vec{\phi}$.[27]

In a scattering experiment in which an initial state $\phi^s(0)$ is scattered into a new state $\phi^s(t)$ at time t which is a superposition of

[24] The real part Re $\langle\vec{\phi}\ |\ \vec{\psi}\,\rangle$ of $\langle\vec{\phi}\ |\ \vec{\psi}\,\rangle$ is also called the correlation term.
[25] *See* §26.2.3.
[26] Beltrametti and Cassinelli (1981) p. 12. We may call $\langle\vec{\phi}\ |\ \vec{\psi}\,\rangle$ *transition amplitude*.
[27] Roman (1965) p. 26.

various states $\varphi_\ell^s(t)$, the probability of finding the scattered system in state $\varphi_\ell^s(t)$ at time t is given by $|\langle\vec{\varphi}_\ell(t) \mid \vec{\phi}(t)\rangle|^2$.[28]

Exercises and Problems

Q31(1) Explain why the transition shown in Eq. (31.1) due to measurement cannot be generated by a unitary transformation.[29]

Q31(2) Prove Eqs. (31.6) and (31.7).

Q31(3) Prove Eqs. (31.17) and (31.18).

Q31(4) Prove Eq. (31.20) for the expectation value.

[28] Greiner p. 188 and Eq. (11.54) on p. 210. Zettili pp. 554–559.

[29] This means that the transition as a time evolution does not satisfy Postulate 29.1.2(TESP) which applies to quantum evolution in the Schrödinger picture.

Chapter 32

Superselection Rules

32.1 Superselection Rules

For orthodox quantum systems, pure states correspond one-to-one to all one-dimensional subspaces and observables correspond one-to-one to all selfadjoint operators in the state space. A relaxation of this one-to-one correspondence will bring about superselection rules. A general formulation of superselection rules is technically complicated.[1] We shall adopt an intuitive approach here and limit our discussion to some special cases to see how superselection rules can be incorporated in a Hilbert space structure of quantum theory.

Definition 32.1(1) *A quantum system is said to possess a superselection rule if its state space $\vec{\mathcal{H}}$ and observables possess the following properties:*

(1) *There exists a preferred complete orthogonal family of subspaces $\vec{\mathcal{S}}^{(n)}$ of $\vec{\mathcal{H}}$ such that pure states correspond one-to-one to one-dimensional subspaces $\vec{\mathcal{S}}_{\vec{\varphi}^{(n)}}$ spanned by unit vectors $\vec{\varphi}^{(n)}$ each lying within a subspace $\vec{\mathcal{S}}^{(n)}$, i.e.,*

$$\textit{pure states} \quad \leftrightarrow \quad \vec{\mathcal{S}}_{\vec{\varphi}^{(n)}} \ \textit{ where } \ \vec{\varphi}^{(n)} \in \vec{\mathcal{S}}^{(n)}. \tag{32.1}$$

[1] Wan pp. 337–391 for a general formulation of superselection rules.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

(2) *Bounded observables B correspond one-to-one to bounded self-adjoint operators $\widehat{B}_{re}$ which are reducible by every subspace $\vec{\mathcal{S}}^{(n)}$.*[2]

(3) *The expectation value of a bounded observable B in a pure state corresponding to the subspace $\vec{\mathcal{S}}_{\vec{\varphi}^{(n)}}$ is given by* Eq. (28.2)*, i.e.,*

$$\mathcal{E}(\widehat{B}_{re}, \vec{\varphi}^{\,(n)}) = \langle \vec{\varphi}^{\,(n)} \mid \widehat{B}_{re}\vec{\varphi}^{\,(n)} \rangle. \tag{32.2}$$

The superselection rule is said to be generated by the complete orthogonal family of subspaces $\vec{\mathcal{S}}^{(n)}$.

Note that while $\vec{\mathcal{S}}_{\vec{\varphi}^{(n)}}$ is one-dimensional the subspaces $\vec{\mathcal{S}}^{(n)}$ may not be one-dimensional. Definition 32.1(1) can be conveniently stated in terms of one-dimensional projectors $\widehat{P}_{\vec{\varphi}^{(n)}}$ instead of one-dimensional subspaces $\vec{\mathcal{S}}_{\vec{\varphi}^{(n)}}$.

The following properties would distinguish quantum systems with a superselection rule from previous orthodox quantum systems:

P32.1(1) The one-to-one correspondence between pure states and *all* one-dimensional subspaces of the state space $\vec{\mathcal{H}}$ for orthodox quantum systems is broken. A unit vector $\vec{\varphi}$ not lying in one of the subspaces, e.g., $\vec{\varphi} = (\vec{\varphi}^{\,(1)} + \vec{\varphi}^{\,(2)})/\sqrt{2}$, where $\vec{\varphi}^{\,(1)} \in \vec{\mathcal{S}}^{(1)}$ and $\vec{\varphi}^{\,(2)} \in \vec{\mathcal{S}}^{(2)}$, does not describe a pure state.

P32.1(2) The one-to-one correspondence between observables and *all* selfadjoint operators for orthodox quantum systems is broken since only reducible selfadjoint operators can represent observables.[3] Being reducible the operator $\widehat{B}_{re}$ leaves every subspace $\vec{\mathcal{S}}^{(n)}$ invariant, i.e.,

$$\widehat{B}_{re}\vec{\varphi}^{\,(n)} \in \vec{\mathcal{S}}^{(n)} \quad \forall \vec{\varphi}^{\,(n)} \in \vec{\mathcal{S}}^{(n)}. \tag{32.3}$$

The operator $\widehat{B}_{re}$ can be written in the form

$$\widehat{B}_{re} = \sum_n \widehat{B}^{(n)}, \tag{32.4}$$

[2] The subscript re indicates the reducibility of the operator $\widehat{B}_{re}$ by every subspace $\vec{\mathcal{S}}^{(n)}$. We confine ourselves to bounded operators for simplicity.

[3] The properties of orthodox systems hold only within each subspace $\vec{\mathcal{S}}^{(n)}$.

where $\widehat{B}^{(n)}$ is the part of $\widehat{B}_{re}$ in $\vec{\mathcal{S}}^{(n)}$ given by Eq. (17.111), i.e.,

$$\widehat{B}^{(n)} = \widehat{P}^{(n)}\widehat{B}_{re}\widehat{P}^{(n)}, \tag{32.5}$$

where $\widehat{P}^{(n)}$ is the projector onto $\vec{\mathcal{S}}^{(n)}$.

P32.1(3) An operator of the form of $\widehat{B}_{re}$ is unable to relate vectors in different subspaces $\vec{S}^{(n)}$, i.e., we have

$$\langle \vec{\varphi}^{(m)} \mid \widehat{B}_{re}\vec{\varphi}^{(n)} \rangle = 0 \quad \text{if } m \neq n, \tag{32.6}$$

on account of Eq. (32.3).

P32.1(4) Let $\vec{\phi}$ be a unit vector in $\mathcal{H}$ with projections in different subspaces $\vec{\mathcal{S}}^{(n)}$, i.e.,

$$\vec{\phi} = \sum_n c_n \vec{\varphi}^{(n)}, \quad \sum_n c_n^* c_n = 1. \tag{32.7}$$

where $\vec{\varphi}^{(n)}$ is a unit vector in $\vec{\mathcal{S}}^{(n)}$. Then we have the following interpretation:

(1) The vector $\vec{\phi}$ does not represent a pure state. This is because that the quadratic form $\mathcal{E}(\widehat{B}_{re}, \vec{\phi})$ reduces to a single sum, similar to Eq. (31.3), on account of Eq. (32.6), i.e.,

$$\mathcal{E}(\widehat{B}_{re}, \vec{\phi}) = \langle \vec{\phi} \mid \widehat{B}_{re}\vec{\phi} \rangle = \sum_n w_n \langle \vec{\varphi}^{(n)} \mid \widehat{B}_{re}\vec{\varphi}^{(n)} \rangle, \tag{32.8}$$

where $w_n = c_n^* c_n$. Compared with Eq. (31.25) we can see the absence of correlation terms.

(2) The vector $\vec{\phi}$ would represent a *mixed state* since the quadratic form $\mathcal{E}(\widehat{B}_{re}, \vec{\phi})$ in Eq. (32.8) taken as the expectation value is seen to embody two averaging processes similar to Eq. (31.3). First we have an average over each pure state $\varphi^{s(n)}$ described by the state vector $\vec{\varphi}^{(n)}$, i.e., $\langle \vec{\varphi}^{(n)} \mid \widehat{B}^{(n)}\vec{\varphi}^{(n)} \rangle$. These values are averaged again over all $\varphi^{s(n)}$ with probabilities w_n.

(3) The vector $\vec{\phi}$ represents a *classical mixture of pure states* $\varphi^{s(n)}$, since no other linear combination of vectors, one in each $\vec{\mathcal{S}}^{(n)}$, would lead to $\vec{\phi}$ in Eq. (32.7).[4] In other words, *a unit vector with projections in different subspaces* $\vec{\mathcal{S}}^{(n)}$ *represents a classical mixture of pure states.*

[4] There does not exist a different expansion of $\vec{\phi}$ in terms of a different set of vectors $\vec{\varphi}'^{(n)} \in \vec{\mathcal{S}}^{(n)}$.

P32.1(5) *The principle of superposition breaks down*. Only pure states within the same subspace $\vec{\mathcal{S}}^{(n)}$ remains coherent, i.e., a linear combination of pure states within the same $\vec{\mathcal{S}}^{(n)}$ remains a coherent superposition. A linear combination of pure states from different $\vec{\mathcal{S}}^{(n)}$ does not constitute a coherent superposition.

For a quantum system possessing a superselection rule generated by a complete orthogonal family of subspaces $\vec{\mathcal{S}}^{(n)}$, we have the following definitions:

Definition 32.1(2)

(1) *Each subspace* $\vec{\mathcal{S}}^{(n)}$ *is called a* **coherent subspace**, *since the principle of superposition holds within such a subspace. These subspaces are also known as* **superselection sectors** *or* **supersectors** *for short.*[5]

(2) *A selfadjoint operator of the form*

$$\widehat{C} = \sum_n c_n \widehat{P}^{(n)}, \quad c_n \in I\!R, \tag{32.9}$$

where $\widehat{P}^{(n)}$ *is the projector onto the subspace* $\vec{\mathcal{S}}^{(n)}$, *is called a* **superselection operator**.

An observable of a classical mechanical system possesses a definite value in every state. We can use this property to formulate a general definition of *classical observables* of any system.

Definition 32.1(3) *An observable A is called a* **classical observable** *if A possesses a definite value in every pure state. A system having classical observables is said to possess* **classical properties**.[6]

An orthodox quantum system does not have any classical observables.[7] A quantum system with a superselection rule possesses classical observables described by superselection operators since an observable represented by a superselection operator $\widehat{C}$ in Eq. (32.9) possesses a definite value in every pure state of the system, e.g., it possesses the value c_n in the pure state corresponding to any unit vector $\vec{\varphi}_n$ lying in the supersector $\vec{\mathcal{S}}^{(n)}$. We can formalise this by the following definition:

[5] The properties of orthodox quantum systems hold within each supersector.
[6] Wan p. 346.
[7] We exclude fixed quantities like the mass of the particle.

Definition 32.1(4) *Quantum systems possessing a superselection rule are referred to as* **mixed quantum systems**.[8]

Mixed quantum systems possess both quantum and classical properties.

32.2 Supersectors and Direct Sum Decomposition

Superselection rules can be formulated in terms of a direct sum decomposition of Hilbert spaces and operators by treating each subspace $\vec{\mathcal{S}}^{(n)}$ as a Hilbert space in its own right.[9] First we can identify the original state space $\vec{\mathcal{H}}$ with the direct sum of $\vec{\mathcal{S}}^{(n)}$, i.e., we can write $\mathcal{H} = \mathcal{H}^{\oplus} = \oplus_n \vec{\mathcal{S}}^{(n)}$. Secondly we can identify $\widehat{B}_{re}$ which are reducible by all those subspaces as a direct sum of operators $\widehat{B}^{(n)}$ defined on $\vec{\mathcal{S}}^{(n)}$, i.e., $\widehat{B}_{re} = \oplus_n \widehat{B}^{(n)}$. In other words, $\widehat{B}_{re}$ is a decomposable operator.[10] Definition 32.1(1) can be stated in terms of direct sum decomposition of the state space as seen in the following definition:

Definition 32.1(5) *A quantum system is said to possess a superselection rule if its state space $\vec{\mathcal{H}}$ and observables possess the following properties:*

(1) *There exists a preferred decomposition of the state space as a direct sum of a complete orthogonal family of subspaces $\vec{\mathcal{S}}^{(n)}$ of $\vec{\mathcal{H}}$, i.e.,*[11]

$$\vec{\mathcal{H}} = \vec{\mathcal{H}}^{\oplus} = \oplus \vec{\mathcal{S}}^{(n)}. \tag{32.10}$$

Pure states correspond one-to-one to one-dimensional subspaces each lying within a subspace $\vec{\mathcal{S}}^{(n)}$.[12]

[8] Wan pp. 378–383. These systems possess a mix of quantum and classical properties. In contrast, orthodox quantum system in Definition 26.1(1) which does not have classical properties.

[9] Beltrametti and Gassinelli pp. 45–48. Blank, Exner and Havliček pp. 268–273. Wan pp. 337–391. Van Fraasen for conceptual discussion. *See* Chapter 24 for direct sum decomposition of Hilbert spaces.

[10] *See* Definition 24.1.2(2).

[11] *See* §24.1.1 for notation, especially Eqs. (24.14) to (24.18).

[12] *See* Eq. (24.24).

(2) *Observables B correspond one-to-one to selfadjoint decomposable operators, i.e., $\widehat{B}^{\oplus} = \oplus_n \widehat{B}^{(n)}$.*

(3) *The expectation value of a bounded observable B in a pure state described by a unit vector $\vec{\varphi}^{(n)}$ in $\vec{\mathcal{S}}^{(n)}$ is given by $\langle \vec{\varphi}_n \mid \widehat{B}^{(n)} \vec{\varphi}^{(n)} \rangle$.*

The superselection rule is said to be generated by the direct sum decomposition.

The subspaces $\vec{\mathcal{S}}^{(n)}$ are referred to as *superesectors*, and selfadjoint diagonalisable operators on $\mathcal{H}^{\oplus}$ are called *superselection operators*.[13] These agree with Definition 32.1(2).

There are many reasons why a superselection rule may emerge. However, we shall not pursue an investigation of this.[14]

32.3 An Example

Consider the model physical system introduced in Q24(1) in Exercises and Problems for Chapter 24. The system has the following properties:

P32.3(1) The state space is a three-dimensional Hilbert space $\vec{\mathcal{H}}^{(m)}$ with a preferred decomposition as a direct sum of a complete orthogonal family of one-dimensional subspaces $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$, i.e., we have[15]

$$\vec{\mathcal{H}}^{(m)} = \vec{\mathcal{S}}^{(-)} \oplus \vec{\mathcal{S}}^{(0)} \oplus \vec{\mathcal{S}}^{(+)}. \tag{32.11}$$

Let $\vec{\eta}^{(-)}$, $\vec{\eta}^{(0)}$ and $\vec{\eta}^{(+)}$ be unit vectors in $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$, respectively. A vector in $\mathcal{H}^{(m)}$ is of the form

$$\begin{aligned} \vec{\eta}^{\oplus} &= c_- \vec{\eta}^{(-)} \oplus c_0 \vec{\eta}^{(0)} \oplus c_+ \vec{\eta}^{(+)} \\ &= c_- \vec{\eta}^{(-)\oplus} + c_0 \vec{\eta}^{(0)\oplus} + c_+ \vec{\eta}^{(+)\oplus}, \end{aligned} \tag{32.12}$$

[13] *See* Definition 24.1.2(2) for decomposable and diagonalisable operators in a direct sum space.

[14] Those interested in the origins of superselection rules are referred to Wan pp. 345–346, 558–561.

[15] This system is used in §34.7 to model a measuring device, hence the superscript (m).

where[16]

$$c_-, \quad c_0, \quad c_+ \in \mathbb{C}, \tag{32.13}$$

$$\vec{\eta}^{(-)\oplus} = \vec{\eta}^{(-)} \oplus \vec{0}^{(0)} \oplus \vec{0}^{(+)}, \tag{32.14}$$

$$\vec{\eta}^{(0)\oplus} = \vec{0}^{(-)} \oplus \vec{\eta}^{(0)} \oplus \vec{0}^{(+)}, \tag{32.15}$$

$$\vec{\eta}^{(+)\oplus} = \vec{0}^{(-)} \oplus \vec{0}^{(0)} \oplus \vec{\eta}^{(+)}. \tag{32.16}$$

Since pure states correspond to unit vectors lying within a supersector, there are only three distinctive pure states, i.e., $\eta^{s(-)}$, $\eta^{s(0)}$ and $\eta^{s(+)}$ which are described by $\vec{\eta}^{(-)\oplus}$, $\vec{\eta}^{(0)\oplus}$ and $\vec{\eta}^{(+)\oplus}$, respectively.

P32.3(2) Observables $B^{(m)}$ correspond one-to-one to selfadjoint decomposable operators, i.e., they are of the form

$$\widehat{B}^{(m)} = \widehat{B}^{(-)} \oplus \widehat{B}^{(0)} \oplus \widehat{B}^{(+)}, \tag{32.17}$$

where $\widehat{B}^{(-)}$, $\widehat{B}^{(0)}$, and $\widehat{B}^{(+)}$ are selfadjoint operators on $\vec{\mathcal{S}}^{(-)}, \vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$, respectively. Since the subspaces are all one-dimensional, we have

$$\widehat{B}^{(-)} = b_- \, \widehat{\mathit{I\!I}}^{(-)}, \quad \widehat{B}^{(0)} = b_0 \, \widehat{\mathit{I\!I}}^{(0)}, \quad \widehat{B}^{(+)} = b_+ \, \widehat{\mathit{I\!I}}^{(+)}, \tag{32.18}$$

where $b_-, b_0, b_+ \in \mathbb{R}$. These operators clearly leave every supersector invariant. Selfadjoint decomposable operators in Eq. (32.17) are also diagonalisable, i.e.,

$$\widehat{B}^{(m)} = b_- \, \widehat{\mathit{I\!I}}^{(-)} \oplus b_0 \, \widehat{\mathit{I\!I}}^{(0)} \oplus b_+ \, \widehat{\mathit{I\!I}}^{(+)}. \tag{32.19}$$

This model system clearly possesses a superselection rule generated by the family of one-dimensional subspaces $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$. The following comments highlight the consequences of the superselection rule:

C32.3(1) Observables which are represented by selfadjoint diagonalisable operators $\widehat{B}^{(m)}$ in Eq. (32.19). have a definite value in every

[16] We employ $\vec{0}^{(-)}$, $\vec{0}^{(0)}$, $\vec{0}^{(+)}$ to denote the zero vectors on $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$, $\vec{\mathcal{S}}^{(+)}$. The corresponding identity operators are denoted by $\widehat{\mathit{I\!I}}^{(-)}$, $\widehat{\mathit{I\!I}}^{(0)}$, $\widehat{\mathit{I\!I}}^{(+)}$.

pure state, i.e.,

$$B^{(m)} \text{ possesses the value } b_- \text{ in state } \eta^{s(-)}, \tag{32.20}$$

$$B^{(m)} \text{ possesses the value } b_0 \text{ in state } \eta^{s(0)}, \tag{32.21}$$

$$B^{(m)} \text{ possesses the value } b_+ \text{ in state } \eta^{s(+)}. \tag{32.22}$$

These are superselection operators which can be used to describe classical properties of the system.[17]

C32.3(2) A linear combination of vectors from different supersectors, e.g., $\vec{\eta}^{\,\oplus}$ in Eq. (32.12), does not lead to a coherent superposition of states, since all the interference terms vanish, e.g.,

$$\langle \vec{\eta}^{\,(0)\oplus} \mid \widehat{B}^{(m)} \vec{\eta}^{\,(+)\oplus} \rangle = 0. \tag{32.23}$$

Following discussion in P32.1(4), we conclude that the vector $\vec{\eta}^{\,\oplus}$ describes a classical mixture of pure states $\eta^{s(-)}$, $\eta^{s(0)}$ and $\eta^{s(+)}$.

C32.3(3) Generally the Hamiltonian of a system is an observable. For our present system, this means that the Hamiltonian is a decomposable operator. Let us assume that the time evolution of our present system can be described by the Schrödinger picture, e.g., a pure state will evolve in accordance with Eq. (29.7). Then a pure state will involve within the same supersector, since the Hamiltonian in the form of a decomposable operator would leave a supersector invariant. There will be no transition from one supersector to another. *For such a transition to occur the system has to couple itself to an external system in such a way that its time evolution is no longer generated by a Hamiltonian which is an observable of the system.*[18] An illustration of such a situation will be presented in §34.7 where a discussion will be given on how this system can model a measuring device.

C32.3(4) Selfadjoint operators which are not decomposable do not represent observables. As examples the operators $\widehat{L}_+^{(m)}$, $\widehat{L}_-^{(m)}$ and $\widehat{L}^{(m)}$ defined on $\vec{\mathcal{H}}^{\oplus}$ by Eqs. (24.78), (24.79) and (24.80) are selfadjoint but not decomposable as they relate different supersectors. In §34.7.2 such operators are used to model measuring processes.

[17] *See* §34.7.2 for application to model measuring devices.

[18] Such a Hamiltonian would have to incorporate the interaction with an external system.

Exercises and Problems

Q32(1) Prove Eq. (32.4).

Q32(2) For the example in §32.3 discuss how the initial state $\eta^{(0)\oplus}$ given by Eq. (32.15) would evolve in the Schrödinger picture under a selfadjoint and decomposable Hamiltonian $\widehat{H}^{\oplus} = \widehat{B}^{(-)} \oplus \widehat{B}^{(0)} \oplus \widehat{B}^{(+)}$.

Q32(3) A system has a two-dimensional state space $\vec{\mathcal{H}}$. A superselection rule operates with the state space $\vec{\mathcal{H}}$ having the following preferred direct sum decomposition:

$$\vec{\mathcal{H}} = \vec{\mathcal{H}}^{\oplus} = \vec{\mathcal{H}}^{(1)} \oplus \vec{\mathcal{H}}^{(2)}, \tag{32.24}$$

where $\vec{\mathcal{H}}^{(1)}$ is spanned by a unit vector $\vec{\eta}^{\,(1)} \in \vec{\mathcal{H}}$ and $\vec{\mathcal{H}}^{(2)}$ is spanned by a unit vector $\vec{\eta}^{\,(2)} \in \vec{\mathcal{H}}$ which is orthogonal to $\vec{\eta}^{\,(1)}$. Let $\widehat{L}$ be an operator on $\vec{\mathcal{H}}$ defined by

$$\widehat{L}\vec{\eta}^{\,(1)} = \vec{\eta}^{\,(2)}, \quad \widehat{L}\vec{\eta}^{\,(2)} = \vec{\eta}^{\,(1)}. \tag{32.25}$$

(a) Show that $\widehat{L}$ is selfadjoint and explain why $\widehat{L}$ cannot represent an observable.

(b) Suppose the Hamiltonian of the system is $\lambda\widehat{L}$, where $\lambda \in \mathbb{R}$. Show that in the Schrödinger picture the initial state vector $\vec{\eta}(0) = i\vec{\eta}^{\,(2)}$ at time $t = 0$ evolves in time to the following state vector at time t:

$$\vec{\eta}(t) = \sin(\lambda t/\hbar)\vec{\eta}^{\,(1)} + i\cos(\lambda t/\hbar)\vec{\eta}^{\,(2)}. \tag{32.26}$$

(c) Discuss how the evolution may cause a transition from a pure state to a mixture.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 33

Many-Particle Systems

33.1 Identity and Distinguishability

A general physical system is a compound system consisting of many constituents. A constituent may be made up of more than one particle, e.g., a molecule is made of atoms which themselves are made up of many particles. In a superconductor at low temperatures the electrons form pairs.[1] Each pair can be treated as a constituent of a superconducting current. For convenience we also call these constituents *particles*. A general physical system is referred to as a many-particle system. The theory for many-particle systems can be built up from one-particle theory. The situation can be complicated since there are different types of many-particle systems:

(1) A compound system may consists of a definite number of particles, i.e., the number of particles in the system remains the same for all time. Such a system may consist of

   (a) distinguishable particles, or
   (b) identical, i.e., indistinguishable, particles.

[1] Feynman, Leighton and Sands §21-7. Wan Chapter 7.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

(2) A compound system may consists of an indefinite number of particles, i.e., the number of particles in the system may change in time due to interactions. The particles involved may or may not be identical.

The mathematics of tensor products presented in §24.2 is required to build up a theory for the compound system. Since different types of systems behave very differently, they require different theories for their description.

Before proceeding let us explain what is meant by identical and distinguishable particles. A particle has properties which are independent of its state, e.g., its mass and charge. These are referred to as *intrinsic properties*. A particle also has properties which dependent on the state, e.g., its position and momentum. These are referred to as *extrinsic properties*,[2] Two particles are considered *identical* if they have the same intrinsic properties, e.g., two electrons are identical while an electron and a positron are not identical. In classical mechanics two identical particles can be distinguished by their extrinsic properties. For example, two classical particles with different initial position and momentum can be identified and tracked at all times since they move in different trajectories in the state space. In quantum mechanics identical particles are not distinguishable in such a way since they do not have well-defined trajectories which can be tracked. The wave functions of the particles may in time develop an overlap such that we cannot tell the particles apart. In other words, identical quantum particles are indistinguishable.

33.2 Distinguishable Particles

Consider a compound system composed of two distinguishable particles. Let $\vec{\mathcal{H}}^{(1)}$ be the state space of particle 1 and $\vec{\mathcal{H}}^{(2)}$ be the state space of particle 2. In the absence of superselection rules a theory for the two-particle system is can be established based on the following assumptions:

[2] Following Jauch pp. 275–278.

(1) *The state space of a compound system, denoted by* $\vec{\mathcal{H}}^{(c)}$*, is the tensor product of the state spaces of its constituents, i.e.,* $\vec{\mathcal{H}}^{(c)} = \vec{\mathcal{H}}^{(1)} \otimes \vec{\mathcal{H}}^{(2)}$*. Pure states correspond one-to-one to one-dimensional subspaces on* $\vec{\mathcal{H}}^{(c)}$*.*

(2) *Observables of the systems correspond one-to-one to selfadjoint operators* $\widehat{A}^{(c)}$ *in the state space* $\vec{\mathcal{H}}^{(c)}$*.*

(3) *The expectation value of an observable in a pure state is given by* Eq. (28.2), i.e.,

$$\mathcal{E}(\widehat{A}^{(c)}, \vec{\Phi}^{(c)}) = \langle \vec{\Phi}^{(c)} \mid \widehat{A}^{(c)} \vec{\Phi}^{(c)} \rangle^{\otimes}, \tag{33.1}$$

where $\vec{\Phi}^{(c)}$ is a unit vector in $\vec{\mathcal{H}}^{(c)}$.

Let $\{\vec{\phi}_j^{\,(1)}\}$ be an orthonormal basis for $\vec{\mathcal{H}}^{(1)}$ and let $\{\vec{\phi}_k^{\,(2)}\}$ be an orthonormal basis for $\vec{\mathcal{H}}^{(2)}$. From Eq. (24.47) we know that $\{\vec{\phi}_j^{\,(1)} \otimes \vec{\phi}_k^{\,(2)}\}$ is an orthonormal basis of the tensor product space, i.e., every vector $\vec{\Phi}^{(c)} \in \vec{\mathcal{H}}^{\otimes}$ has the following expansion[3]:

$$\vec{\Phi}^{(c)} = \sum_{j,k} c_{j,k}\, \vec{\phi}_j^{\,(1)} \otimes \vec{\phi}_k^{\,(2)}, \quad c_{j,k} = \langle \vec{\phi}_j^{\,(1)} \otimes \vec{\phi}_k^{\,(2)} \mid \vec{\Phi}^{(c)} \rangle^{\otimes}. \tag{33.2}$$

We can divide the states of the compound system into two types:

(1) **Product states** These are described by the tensor products of two one-particle state vectors, i.e., $\vec{\phi}^{\,(1)} \otimes \vec{\phi}^{\,(2)}$.

(2) **Entangled states** These are described by linear combinations of product state vectors, i.e., $\vec{\Phi}^{(c)}$ in Eq. (33.2), which cannot be factorised into a single tensor product of two one-particle state vectors, e.g.,[4]

$$\vec{\Phi}^{(c)} = \frac{1}{\sqrt{2}} \left(\vec{\phi}^{\,(1)} \otimes \vec{\phi}^{\,(2)} + \vec{\psi}^{\,(1)} \otimes \vec{\psi}^{\,(2)} \right), \tag{33.3}$$

where $\vec{\phi}^{\,(1)}$ is said to be *entangled* with $\vec{\phi}^{\,(2)}$ and $\vec{\psi}^{\,(1)}$ *entangled* with $\vec{\psi}^{\,(2)}$. The physical meaning of this entanglement will become apparent in the next section.

[3] Following the notation in §24.2.

[4] The one-particle state vectors are assumed to be orthogonal, e.g., $\langle \vec{\phi}^{\,(1)} \mid \vec{\psi}^{\,(1)} \rangle = 0$.

We can also divide observables of the compound system into two types:

(1) **One-particle observables** Observables corresponding to selfadjoint operators in $\vec{\mathcal{H}}^{(c)}$ of the form

$$\widehat{A}^{(c1)} = \widehat{A}^{(1)} \otimes \widehat{\mathbb{1}}^{(2)}, \quad \widehat{A}^{(c2)} = \widehat{\mathbb{1}}^{(1)} \otimes \widehat{A}^{(2)} \tag{33.4}$$

are called *one-particle observables* as they relate to only one of the two particles in the compound system and their expectation value can be obtained from the relevant one-particle states,[5] e.g., for the state vector $\vec{\Phi}^{(c)}$ in Eq. (33.2) the quadratic form $\mathcal{E}\left(\widehat{A}^{(c1)}, \vec{\Phi}^{(c)}\right)$ is given by[6]

$$\begin{aligned} &\sum_{j,k,m,n} c^*_{j,k} c_{m,n} \langle \vec{\phi}^{(1)}_j \otimes \vec{\phi}^{(2)}_k \mid \left(\widehat{A}^{(1)} \otimes \widehat{\mathbb{1}}^{(2)}\right) \vec{\phi}^{(1)}_m \otimes \vec{\phi}^{(2)}_n \rangle^{\otimes} \\ &= \sum_{j,k,m} c^*_{j,k} c_{m,k} \langle \vec{\phi}^{(1)}_j \mid \widehat{A}^{(1)} \vec{\phi}^{(1)}_m \rangle^{(1)}. \end{aligned} \tag{33.5}$$

The calculation is seen to involve quantities of particle 1 only.

(2) **Two-particle observables** These are observables involving both particles. An example is an observable corresponding to selfadjoint operators the form

$$\widehat{A}^{(c)} = \widehat{A}^{(1)} \otimes \widehat{A}^{(2)}, \tag{33.6}$$

which is equal to the product of $\widehat{A}^{(c1)}$ and $\widehat{A}^{(c2)}$ in Eq. (33.4). For the state vector $\vec{\Phi}^{(c)}$ in Eq. (33.2) the quadratic form $\mathcal{E}\left(\widehat{A}^{(c)}, \vec{\Phi}^{(c)}\right)$ of $\widehat{A}^{(c)}$ is given by

$$\sum_{j,k,m,n} c^*_{j,k} c_{m,n} \langle \vec{\phi}^{(1)}_j \otimes \vec{\phi}^{(2)}_k \mid \widehat{A}^{(1)} \otimes \widehat{A}^{(2)} \vec{\phi}^{(1)}_m \otimes \vec{\phi}^{(2)}_n \rangle^{\otimes}, \tag{33.7}$$

which is equal to

$$\sum_{j,k,m,n} c^*_{j,k} c_{m,n} \langle \vec{\phi}^{(1)}_j \mid \widehat{A}^{(1)} \vec{\phi}^{(1)}_m \rangle^{(1)} \langle \vec{\phi}^{(2)}_k \mid \widehat{A}^{(2)} \vec{\phi}^{(2)}_n \rangle^{(2)}. \tag{33.8}$$

[5] Jauch pp. 179–182. Peres Chapter 5.

[6] Using Eq. (24.57) and the orthonormality of the basis vectors $\vec{\phi}^{(2)}_k$ of $\vec{\mathcal{H}}^{(2)}$.

The expression involves quantities of both particles. Since $\widehat{A}^{(c1)}$ and $\widehat{A}^{(c2)}$ commute, the observables they represent are simultaneously measurable. The measurement of these two observables would constitute a measurement of $\widehat{A}^{(1)} \otimes \widehat{A}^{(2)}$. This is also seen in the above expression which contains terms $\langle \vec{\phi}_j^{(1)} \mid \widehat{A}^{(1)} \vec{\phi}_m^{(1)} \rangle^{(1)}$ and $\langle \vec{\phi}_k^{(2)} \mid \widehat{A}^{(2)} \vec{\phi}_n^{(2)} \rangle^{(2)}$.[7]
Another example of two-particle observables is one described by the following operator:

$$\widehat{A}^{(c1)} + \widehat{A}^{(c2)} = \widehat{A}^{(1)} \otimes \widehat{I\!I}^{(2)} + \widehat{I\!I}^{(1)} \otimes \widehat{A}^{(2)}. \tag{33.9}$$

If $\widehat{A}^{(1)}$ is the Hamiltonian of particle 1 and $\widehat{A}^{(2)}$ is the Hamiltonian of particle 2 then $\widehat{A}^{(c1)} + \widehat{A}^{(c2)}$ is the total Hamiltonian of the two-particle in the absence of any interaction.[8]

The above formulation can also be applied to construct a theory of one-particle systems which have two very distinctive and independent properties. Let us illustrate this with the example of a spin-$\frac{1}{2}$ particle.

Spin-$\frac{1}{2}$ particles A spin-$\frac{1}{2}$ particle has two very distinctive and independent properties due to its motion in physical space and its spin motion. Properties due to its spatial motion are described by selfajoint operators $\widehat{A}$ in the Hilbert space $\vec{L}^2(I\!R^3)$ and properties due to its spin motion are described by selfajoint operators in the Hilbert space $\vec{\mathbb{W}}^2$, e.g., $\widehat{S}_x$, $\widehat{S}_y$ and $\widehat{S}_z$ in §14.1.1. These two properties can be combined and described in the tensor product space[9]

$$\vec{\mathcal{H}}^{(s)}(I\!R^3) := \vec{L}^2(I\!R^3) \otimes \vec{\mathbb{W}}^2. \tag{33.10}$$

For states we have

(1) State vectors for spin-up states are of the form $\vec{\phi} \otimes \vec{\alpha}_z$.[10]

[7] Note that $\langle \vec{\phi}_j^{(1)} \mid \widehat{A}^{(1)} \vec{\phi}_m^{(1)} \rangle^{(1)} = \langle \vec{\phi}_j^{(1)} \otimes \vec{\phi}_k^{(2)} \mid \widehat{A}^{(c1)} \vec{\phi}_m^{(1)} \otimes \vec{\phi}_k^{(2)} \rangle^{\otimes}$.

[8] Isham pp. 143–147 which contains a discussion on probability distributions of measured values.

[9] *See* Eq. (24.64). The reason that a small symbol $\otimes$ is used for the tensor product will becomes apparent in Eq. (34.8).

[10] Here $\vec{\phi}$ is in $\vec{L}^2(I\!R^3)$, and $\vec{\alpha}_z$ and $\vec{\beta}_z$ are defined in §14.1.1 as two orthonormal members of $\vec{\mathbb{W}}^2$.

(2) State vectors for spin-down states are of the form $\vec{\phi} \otimes \vec{\beta}_z$.
(3) A general state is of the form $\vec{\phi} \otimes \vec{\alpha}_z + \vec{\phi} \otimes \vec{\beta}_z$.

For observables we have[11]:

(1) Spin operators take the form

$$\widehat{I\!I}(I\!R^3) \otimes \widehat{S}_x, \quad \widehat{I\!I}(I\!R^3) \otimes \widehat{S}_y \quad \text{and} \quad \widehat{I\!I}(I\!R^3) \otimes \widehat{S}_z. \tag{33.11}$$

(2) Spatial operators take the form $\widehat{A} \otimes \widehat{I\!I}(\vec{\mathbb{W}}^2)$.

An evaluation of various expectation values demonstrate that the above description makes sense. For example, we have[12]

$$\mathcal{E}\big(\widehat{I\!I}(I\!R^2) \otimes \widehat{S}_z, \vec{\phi} \otimes \vec{\alpha}_z\big) = \langle \vec{\alpha}_z \mid \widehat{S}_z \vec{\alpha}_z \rangle = \frac{1}{2}\hbar. \tag{33.12}$$

$$\mathcal{E}\big(\widehat{A} \otimes \widehat{I\!I}(\vec{\mathbb{W}}^2), \vec{\phi} \otimes \vec{\alpha}_z\big) = \langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle. \tag{33.13}$$

33.3 Identical Particles

The description of a system of distinguishable particles in the preceding section does not apply to a system of identical particles. The fact that the particles in a compound system are identical and hence indistinguishable would impose some restrictions on the vectors and operators used to describe the states and observables of the system. A further complication is that there are two main types of particles in nature, i.e., bosons and fermions[13]:

(1) Bosons are particles of zero or integer spin ($0, \hbar, 2\hbar, \ldots$), such as photons, pions and Cooper pairs.[14]

---

[11] Here $\widehat{I\!I}(I\!R^3)$ is the identity operator on $\vec{L}^2(I\!R^3)$, and $\widehat{I\!I}(\vec{\mathbb{W}}^2)$ is the identity operator on $\vec{\mathbb{W}}^2$.

[12] Here $\langle \vec{\alpha}_z \mid \widehat{S}_z \vec{\alpha}_z \rangle$ is the scalar product in $\vec{\mathbb{W}}^2$ and $\langle \vec{\phi} \mid \widehat{A}\vec{\phi} \rangle$ is the scalar product in $\vec{L}^2(I\!R^3)$.

[13] Bose (1894–1974) was an Indian mathematical physicist whose work laid the foundation for Bose-Einstein statistics and Bose-Einstein condensate. Fermi (1901–1954) was an Italian physicist whose work led to a statistical theory now known as Fermi-Dirac statistics.

[14] Feynman, Leighton and Sands §21-7 and Wan Chapter 7 for a discussion on Cooper pairs and their relation to superconductivity.

(2) Fermions are particles of half integer spin ($\hbar/2$, $3\hbar/2$, ...), such as electrons, protons, neutrons.

Fermions obey the **Pauli exclusion principle** which says that no two identical fermions can be simultaneously in the same state. Bosons are not restricted by this requirement so that any number of them can be simultaneously in the same state. It follows that the description of a system of identical fermions are different from that of a system of bosons.

Let us consider two identical particles. Separately each particle has its own state space, i.e., $\vec{\mathcal{H}}^{(1)}$ for particle 1 and $\vec{\mathcal{H}}^{(2)}$ for particle 2. The particles being identical means that the two state spaces are identical, i.e., $\vec{\mathcal{H}}^{(1)} = \vec{\mathcal{H}}^{(2)} = \vec{\mathcal{H}}$. Next consider these two particles forming a compound system. The state space of the compound system is again assumed to be $\vec{\mathcal{H}}^{(c)} = \vec{\mathcal{H}} \otimes \vec{\mathcal{H}}$. Let $\{\vec{\varphi}_m\}$ be an orthonormal basis for $\vec{\mathcal{H}}$. Then $\{\vec{\varphi}_m \otimes \vec{\varphi}_n\}$ is an orthonormal basis in $\vec{\mathcal{H}}^{(c)}$.[15] The particles being identical means that when we interchange the particles observables should remain the same and the state may change at most by a phase factor, e.g., by a multiplicative constant –1, so that the interchange causes no observable effect.

Let us examine how to describe the changes in vectors and operators when we interchange the particles:

(1) **Vectors** The product vector $\vec{\varphi}_m \otimes \vec{\varphi}_n$ indicates that particle 1 is in state $\vec{\varphi}_m$ while particle 2 is in state $\vec{\varphi}_n$. An interchange of the particles would change the vector $\vec{\varphi}_m \otimes \vec{\varphi}_n$ to $\vec{\varphi}_n \otimes \vec{\varphi}_m$. Such an exchange of particles can be described mathematically in terms of a *permutation operator* $\widehat{U}_p$ defined on $\vec{\mathcal{H}}^{(c)}$ by[16]

$$\widehat{U}_p\Big(\sum_{m,n} c_{mn}\vec{\varphi}_m \otimes \vec{\varphi}_n\Big) := \sum_{m,n} c_{mn}\,\vec{\varphi}_n \otimes \vec{\varphi}_m. \tag{33.14}$$

This is a bounded operator with $\vec{\mathcal{H}}^{(c)}$ as its domain. This permutation operator is selfadjoint, unitary with its square equal to the identity operator.[17]

[15] The superscripts (1) and (2) used for non-identical particles are omitted. We will also drop these superscripts for operators.
[16] These operators are introduced first in Eq. (24.81).
[17] *See* Eqs. (24.82).

(2) **Operators** On an interchange of particles the operator $\widehat{A} \otimes \widehat{B}$ will go to $\widehat{B} \otimes \widehat{A}$. This change can be obtained by the following unitary transformation[18]:

$$\widehat{U}_p \left(\widehat{A} \otimes \widehat{B} \right) \widehat{U}_p^\dagger = \widehat{B} \otimes \widehat{A}. \tag{33.15}$$

Definition 33.3.1(1)

(1) *A vector in* $\vec{\mathcal{H}}^{(c)}$ *is said to be*

(a) **symmetrical**, *to be denoted by* $\vec{\Phi}^{(cs)}$, *if it is unchanged by the permutation operator, i.e.,*

$$\widehat{U}_p \vec{\Phi}^{(cs)} = \vec{\Phi}^{(cs)}, \tag{33.16}$$

(b) **antisymmetrical**, *to be denoted by* $\vec{\Phi}^{(ca)}$, *if it is changed by a phase factor of* -1 *by the permutation operator, i.e.,*

$$\widehat{U}_p \vec{\Phi}^{(ca)} = -\vec{\Phi}^{(ca)}. \tag{33.17}$$

(2) *An operator in* $\vec{\mathcal{H}}^{(c)}$ *is said to be* **symmetrical**, *to be denoted by* $\widehat{C}^{(cs)}$, *if it is unchanged by the permutation operator, i.e.,*[19]

$$\widehat{C}^{(cs)} = \widehat{U}_p \widehat{C}^{(cs)} \widehat{U}_p^\dagger. \tag{33.18}$$

The following examples serve to illustrate the definition[20]:

$$\vec{\Phi}^{(cs)} = \frac{1}{\sqrt{2}} \left(\vec{\varphi}_m \otimes \vec{\varphi}_n + \vec{\varphi}_n \otimes \vec{\varphi}_m \right) \text{ is symmetrical,} \tag{33.19}$$

$$\vec{\Phi}^{(cs)} = \frac{1}{\sqrt{2}} \left(\vec{\varphi}_m \otimes \vec{\varphi}_n - \vec{\varphi}_n \otimes \vec{\varphi}_m \right) \text{ is antisymmetrical,} \tag{33.20}$$

$$\widehat{A} \otimes \widehat{I\!I} + \widehat{I\!I} \otimes \widehat{A} \text{ and } \widehat{A} \otimes \widehat{B} + \widehat{B} \otimes \widehat{A} \text{ are symmetrical.} \tag{33.21}$$

Symmetrical vectors in $\vec{\mathcal{H}}^{(c)}$ form a subspace $\vec{\mathcal{S}}^{(cs)}$, to be called the **symmetrical subspace** of $\vec{\mathcal{H}}^{(c)}$. Antisymmetrical vectors in $\vec{\mathcal{H}}^{(c)}$ also form a subspace $\vec{\mathcal{S}}^{(ca)}$, to be called the **antisymmetrical subspace**

[18] Jauch pp. 277–278. Blank, Exner and Havliček pp. 389–392. *See* also Eq. (24.83). To avoid technical complications, we only consider bounded operators.
[19] Not to be confused with symmetric operators in Definition 19.1(1).
[20] Here we have $m \neq n$.

of $\vec{\mathcal{H}}^{(c)}$. We can define two projectors $\widehat{P}^{(cs)}$ and $\widehat{P}^{(ca)}$ associated with these two subspaces by[21]

$$\widehat{P}^{(cs)} := \frac{1}{2}\Big(\widehat{\mathbb{I}} + \widehat{U}_p\Big), \quad \widehat{P}^{(ca)} := \frac{1}{2}\Big(\widehat{\mathbb{I}} - \widehat{U}_p\Big). \tag{33.22}$$

These operators will project a vector in $\vec{\mathcal{H}}^{(c)}$ onto $\vec{\mathcal{S}}^{(cs)}$ and $\vec{\mathcal{S}}^{(ca)}$, respectively, i.e.,

$$\widehat{P}^{(cs)}\Big(\sum_{m,n} c_{mn}\vec{\varphi}_m \otimes \vec{\varphi}_n\Big) = \sum_{m,n} c_{mn}\,\frac{1}{2}\Big(\vec{\varphi}_m \otimes \vec{\varphi}_n + \vec{\varphi}_n \otimes \vec{\varphi}_m\Big),$$
$$\widehat{P}^{(ca)}\Big(\sum_{m,n} c_{mn}\vec{\varphi}_m \otimes \vec{\varphi}_n\Big) = \sum_{m,n} c_{mn}\,\frac{1}{2}\Big(\vec{\varphi}_m \otimes \vec{\varphi}_n - \vec{\varphi}_n \otimes \vec{\varphi}_m\Big).$$

These two projectors are orthogonal to each other. We also have $\widehat{P}^{(cs)} + \widehat{P}^{(ca)} = \widehat{\mathbb{I}}$.

The description of a system of two identical particles can be formulated as follows.

(1) For the state space we have a choice:
   (a) We can choose the state space to be the symmetrical subspace $\vec{\mathcal{S}}^{(cs)}$.
   (b) Alternatively we can choose the state space to be the antisymmetrical subspace $\vec{\mathcal{S}}^{(ca)}$.

(2) Observables should be represented by selfadjoint operators which are symmetrical, i.e., $\widehat{C}^{(cs)}$, so that the interchange of the two particles would have no effect on the observables. Since these operators commute with $\widehat{U}_p$, they also commute with $\widehat{P}^{(cs)}$ and $\widehat{P}^{(ca)}$. It follows from Theorem 17.9(1) that these operators are reduced by the symmetrical subspace. An operator $\widehat{C}^{(cs)}$ in $\vec{\mathcal{H}}^{(c)}$ has a part in the subspace $\vec{\mathcal{S}}^{(cs)}$. If we treat $\vec{\mathcal{S}}^{(cs)}$ as a Hilbert space in its own right the part of $\widehat{C}^{(cs)}$ in $\vec{\mathcal{S}}^{(cs)}$ can be regarded as an operator in $\vec{\mathcal{S}}^{(cs)}$. We can conveniently denote both $\widehat{C}^{(cs)}$ in $\vec{\mathcal{H}}^{(c)}$ and its part in $\vec{\mathcal{S}}^{(cs)}$ by the same notation. A similar notation will denote the part of the operator in the antisymmetrical subspace $\vec{\mathcal{S}}^{(ca)}$ which also reduces the operator.

[21] These operators are introduced in Eq. (24.84).

33.3.1 Bosons and Fermions

Consider a two-particle system having the symmetrical subspace $\vec{\mathcal{S}}^{(cs)}$ of $\vec{\mathcal{H}}^{(c)}$ as its state space and observables described by selfadjoint operators $\widehat{C}^{(cs)}$ in $\vec{\mathcal{S}}^{(cs)}$. The indistinguishability of the two particles is accounted for by the fact that an interchange of the two particles does not change the state vector or the operators representing observables. It follows that the expectation values are invariant with respect to the exchange of the two particles. The system does not obey the Pauli exclusion principle since the two particles can be in the same state, e.g., the product vector $\vec{\varphi}_m \otimes \vec{\varphi}_m$ is symmetrical. All these vectors are in $\vec{\mathcal{S}}^{(cs)}$ and hence it can serve as a state vector. Particles described above are therefore **bosons**.

A two-particle system having the antisymmetrical subspace $\vec{\mathcal{S}}^{(ca)}$ of $\vec{\mathcal{H}}^{(c)}$ as its state space would obey the Pauli exclusion principle. This is because the antisymmetrical nature of the state vectors precludes the two particles being in the same state. For example, if we put $n = m$ in $\vec{\Phi}^{(ca)}$ in Eq. (33.20) we would obtain the zero vector. In other words, we cannot create a state for the compound system with both of its constituent particles in the same state. Observables are represented by selfadjoint operators $\widehat{C}^{(cs)}$ which are symmetrical. Generally an interchange of the two particles change the state vector only by a phase factor of -1 while the operators representing observables remain unchanged. The expectation values are invariant with respect to the exchange of the two particles. Particles described above are therefore **fermions**.

33.3.2 The Pauli Exclusion Principle

Consider a system of two identical fermions in a state at time t described by the state vector[22]

$$\vec{\Phi}^{(ca)}(t) := \frac{1}{\sqrt{2}} \left(\vec{\varphi}_m(t) \otimes \vec{\varphi}_n(t) - \vec{\varphi}_n(t) \otimes \vec{\varphi}_m(t) \right). \tag{33.23}$$

The exclusion principle entails a correlation between the particles in that identical fermions would seem to be “aware of” the existence of each other so as to “arrange” the state of the system as a whole to

[22] We use the Schrödinger picture for time evolution here.

be antisymmetrical with $\vec{\varphi}_m \neq \vec{\varphi}_n$ to satisfy the exclusion principle. Pauli realised this and he qualified the principle with the following statement[23]:

> *From a superficial consideration of the exclusion principle, it might be thought that a sort of action-at-a-distance is being postulated, as a result of which even two widely separated particles are aware of one another ("sign of a contract"). However, this is not so, because the exclusion principle is only valid as long as the wave packets of the two particles overlap.*

We can understand Pauli's statement in an asymptotic sense. Let us examine how the exclusion principle can become invalid. The state vector $\vec{\Phi}^{(ca)}(t)$ in Eq. (33.23) is a linear combination of $\vec{\varphi}_m(t) \otimes \vec{\varphi}_n(t)$ and $\vec{\varphi}_n(t) \otimes \vec{\varphi}_m(t)$. The expectation value of an observable represented by the operator $\widehat{A} \otimes \widehat{B} + \widehat{B} \otimes \widehat{A}$ in the state described by $\vec{\Phi}^{(ca)}(t)$ at time t can be calculated, e.g.,

$$\begin{aligned}
&\langle \vec{\Phi}^{(ca)}(t) \mid \widehat{A} \otimes \widehat{B}\, \vec{\Phi}^{(ca)}(t) \rangle \\
&= \frac{1}{2} \Big(\langle \vec{\varphi}_m(t) \mid \widehat{A}\, \vec{\varphi}_m(t) \rangle \langle \vec{\varphi}_n(t) \mid \widehat{B}\, \vec{\varphi}_n(t) \rangle \\
&\quad + \langle \vec{\varphi}_n(t) \mid \widehat{A}\, \vec{\varphi}_n(t) \rangle \langle \vec{\varphi}_m(t) \mid \widehat{B}\, \vec{\varphi}_m(t) \rangle \Big) + C, \qquad (33.24)
\end{aligned}$$

where

$$\begin{aligned}
C = -\frac{1}{2} \Big(&\langle \vec{\varphi}_m(t) \mid \widehat{A}\, \vec{\varphi}_n(t) \rangle \langle \vec{\varphi}_n(t) \mid \widehat{B}\, \vec{\varphi}_m(t) \rangle \\
&- \langle \vec{\varphi}_n(t) \mid \widehat{A}\, \vec{\varphi}_m(t) \rangle \langle \vec{\varphi}_m(t) \mid \widehat{B}\, \vec{\varphi}_n(t) \rangle \Big). \qquad (33.25)
\end{aligned}$$

The term C due to the correlations between $\vec{\varphi}_m(t)$ and $\vec{\varphi}_n(t)$ generated by $\widehat{A}$ and $\widehat{B}$ will vanish asymptotically as $\vec{\varphi}_m(t)$ and $\vec{\varphi}_n(t)$ moves spatially apart independently in time so that at large times T the value given by Eq. (33.24) tends to the sum of[24]

$$\frac{1}{2} \langle \vec{\varphi}_m(T) \mid \widehat{A}\, \vec{\varphi}_m(T) \rangle \langle \vec{\varphi}_n(T) \mid \widehat{B}\, \vec{\varphi}_n(T) \rangle \qquad (33.26)$$

[23] Pauli §36 p. 168.

[24] Intuitively we can appreciate that if $\vec{\varphi}_m(t)$ and $\vec{\varphi}_n(t)$ correspond to two different and disjoint ranges of momentum values then the two corresponding wave functions will move apart in time so that $\langle \vec{\varphi}_n(t) \mid \widehat{A}\, \vec{\varphi}_m(t) \rangle$ and $\langle \vec{\varphi}_m(t) \mid \widehat{B}\, \vec{\varphi}_n(t) \rangle$

and

$$\frac{1}{2}\,\langle\vec{\varphi}_m(T)\mid\widehat{B}\,\vec{\varphi}_m(t)\rangle\langle\vec{\varphi}_n(T)\mid\widehat{A}\,\vec{\varphi}_n(T)\rangle. \tag{33.27}$$

As a result the antisymmetrical state vector $\vec{\Phi}^{(ca)}(T)$ becomes effectively the same as the product state vector $\vec{\varphi}_m(T)\otimes\vec{\varphi}_n(T)$, i.e., we have

$$\begin{aligned}&\langle\vec{\Phi}^{(ca)}(T)\mid\big(\widehat{A}\otimes\widehat{B}+\widehat{B}\otimes\widehat{A}\big)\vec{\Phi}^{(ca)}(T)\rangle\\ &\approx\ \langle\vec{\varphi}_m(T)\otimes\vec{\varphi}_n(T)\mid\big(\widehat{A}\otimes\widehat{B}+\widehat{B}\otimes\widehat{A}\big)\vec{\varphi}_m(T)\otimes\vec{\varphi}_n(T)\rangle.\end{aligned} \tag{33.28}$$

Consequently, Pauli exclusion principle loses its constraining power on the state. It can be disregarded for all practical purposes, e.g., there is no need to formally writing down an antisymmetrical state vector for an electron here on earth and one on a galaxy on the other side of the universe when we are dealing with observables localised here on earth.[25]

33.4 Indefinite Number of Particles

The theories presented in §33.3.2 and §33.3.2 can be extended to systems of many particles. For an n-particle system, we would start with the tensor product of the n constituent state spaces. Let us denote the tensor product space by $\vec{\mathcal{H}}^{(c)}(n)$.[26] All symmetrical vectors in $\vec{\mathcal{H}}^{(c)}(n)$ form a symmetric subspace $\vec{\mathcal{S}}^{(cs)}(n)$ and all antisymmetrical vectors in $\vec{\mathcal{H}}^{(c)}(n)$ also form an antisymmetrical subspace $\vec{\mathcal{S}}^{(ca)}(n)$.[27] Then

(1) A compound system of two identical bosons would have a state space $\vec{\mathcal{S}}^{(cs)}(2)$ which is the symmetrical subspace of the tensor product of two constituent Hilbert spaces.

would tend to zero. This can be formulated rigorously by introducing the concepts of *asymptotic localisation and separation*. Details are available in Wan p. 299, pp. 317–317. *See* Wan pp. 620–622 for a formulation of *asymptotically separable quantum theory*.

[25] Gasiorowicz pp. 204–206.

[26] Jauch pp. 280–281, 285. Zettili pp. 447–449.

[27] Symmetrical or antisymmetrical with respect to the interchange any two particles.

(2) A compound system of three identical bosons would have a state space $\vec{\mathcal{S}}^{(cs)}(3)$ which is the symmetrical subspace of the tensor product of three constituent Hilbert spaces

(3) A compound system of n identical bosons has a state space $\vec{\mathcal{S}}^{(cs)}(n)$ which is the symmetrical subspace of the tensor product of n constituent Hilbert spaces.

There are in nature systems composed an indefinite number of particles. For example, the number of particles can change. New particles can be created and existing ones can be annihilated by interaction. When dealing with such a system having an indefinite number of particles the state space cannot be a rigid tensor product of a definite number of constituent state spaces.[28] Let us consider a system of bosons. To include a general situation, we shall start with what is known as the **vacuum**, a state with no particle at all. This will be formally described by a one-dimensional Hilbert space to be denoted by $\vec{\mathcal{S}}(0)$. Next we have the one-particle state space $\vec{\mathcal{S}}(1) = \vec{\mathcal{H}}(1)$. The two-particle state space would be the symmetrical subspace $\vec{\mathcal{S}}^{(cs)}(2)$ and similarly the n-particle state space is $\vec{\mathcal{S}}^{(cs)}(n)$. The state space $\mathcal{H}_F^{(cs)}$ for a system of indefinite number of bosons is taken to be the direct sum of all these state spaces, i.e.,

$$\mathcal{H}_F^{(cs)} := \vec{\mathcal{S}}(0) \oplus \vec{\mathcal{S}}(1) \oplus \vec{\mathcal{S}}^{(cs)}(2) \oplus \vec{\mathcal{S}}^{(cs)}(3) \oplus \cdots . \tag{33.29}$$

Such a space is known as a **Fock space**. The transition from $\vec{\mathcal{S}}^{(cs)}(m)$ to $\vec{\mathcal{S}}^{(cs)}(n)$ can be carried out by operators defined on $\vec{\mathcal{H}}_F^{(cs)}$ having the properties of creation and annihilation operators described in §27.9.[29]

For a system of indefinite number of fermions, we can construct a Fock space $\vec{\mathcal{H}}_F^{(ca)}$ in terms of antisymmertric subspaces $\vec{\mathcal{S}}^{(ca)}(n)$, i.e.,

$$\vec{\mathcal{H}}_F^{(ca)} := \vec{\mathcal{S}}(0) \oplus \vec{\mathcal{S}}(1) \oplus \vec{\mathcal{S}}^{(ca)}(2) \oplus \vec{\mathcal{S}}^{(ca)}(3) \oplus \cdots . \tag{33.30}$$

[28] Jauch pp. 280–287. A system of photons serves to illustrate the situation.
[29] Jauch pp. 280–287. Reed and Simon Vol. 1 p. 53.

Exercises and Problems

Q33(1) Calculate the expectation value of the two-particle observable represented by the operator in Eq. (33.9) in the state described by the vector $\vec{\Phi}^{(c)}$ in Eq. (33.2).

Q33(2) Show that symmetrical vectors in Eq. (33.19) form a subspace of $\vec{\mathcal{H}}^{(c)}$.

Chapter 34

Conceptual Issues

Many quantum predictions seem to defy "common sense" and appear beyond the comprehension of some of the greatest human minds. Richard Feynman states that "*I think I can safely say that nobody understands quantum mechanics*" in his book entitled *The Character of Physical Laws*.[1] Many pioneers of quantum theories have constructed examples to highlight this situation. These examples are now known as *quantum paradoxes*. Moreover, the common understanding that only microscopic systems behave quantum mechanically is incorrect. Macroscopic systems, e.g., superconducting systems and Bose-Einstein condensate, also exhibit quantum behaviour. These systems do not satisfy the properties of orthodox quantum systems.[2]

In this chapter we shall discuss a few well-known examples to illustrate the conceptual issues confronting quantum theory.[3]

[1] Feynman (1918–1988) was an American theoretical physicist and a Nobel Prize winner. The statement is on p. 129 of his book.

[2] *See* Definition 26.1(1). Wan §7.

[3] We shall not consider other less well-known cases such as Zeno paradox and delayed-choice experiments (*see* Auletta, Fortunato and Parisi pp. 320–324).

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

34.1 Understanding Quantum Behaviour

The probabilistic behaviour of quantum systems arising from Postulate 28.2(PD) tells us that an observable may not possess a definite value in a given state. Two examples come to mind immediately:

E34.1(1) A classical particle has a definite position in any given state. As the state evolves in time the particle will move along a definite path in the physical space. In contrast a quantum particle does not possess a definite position value in any state. It follows that a quantum particle does not move along a definite path in the physical space. A similar situation applies to the momentum.

E34.1(2) The lack of a definite value is not just for the position and momentum. For a particle in circular motion, its Hamiltonian does not possess a value in the state given by Eq. (28.5).

The fact that *a quantum observable may have no definite value in a given state* can be understood in many cases in terms of the wave nature of quantum systems arising from the description of quantum states in Postulate 25.1(PS). For example, the state of a quantum particle in one-dimensional motion along the x-axis is describable by a wave function $\phi(x, t)$. This is a solution of the time-dependent Schrödinger equation.[4] It has already been pointed out in §2.2 that some observables of a classical wave may not have a definite value. It is then not so surprising that observables of a quantum system whose state is described in terms of a wave function may not have a definite value in a general state. An observable has a definite value only in an eigenstate of the observable. Of course, quantum systems are not the same as classical wave systems. For example, we can distinguish a quantum system from a classical wave system by the way measured values are related to the state, i.e., for quantum systems the relationship is probabilistic as stated in Postulate 28.2(PD). A quantum wave function is generally complex and it has the interpretation of a position probability amplitude.

[4] *See* §10.3.

While we can understand why an observable of a quantum system may not possess a definite value in a given state, there are other quantum behaviours which are more difficult to understand. Intuitively it seems that all the "troubles" are due to the wave nature of quantum systems, the superposition principle in particular. In the following sections, we shall discuss some explicit examples to illustrate the conceptual challenges confronting quantum mechanics.

34.2 Particle-Wave Duality

In 1801 Thomas Young proved that light was a wave.[5] In his famous double-slit interference experiment he sent a beam of light through a double-slit to produce an interference pattern on a screen behind the slits. It is this interference pattern which proved that light was a wave, not a beam of classical particles as proposed by Newton. A beam of classical particles will not produce the interference pattern on the screen.

Similar experiments have been done with a beam of electrons.[6] Consider an experiment which sends a beam of electrons through a double-slit. Having gone through the double-slit the electrons will strike a screen placed behind the double-slit. This results in a typical double-slit diffraction pattern on the screen. The conclusion is that electrons possess wave properties. Experiments have been done by sending one electron at a time through the double-slit.[7] Initially the experiment produces what seems to be a random distribution of dots on the screen. However, the double-slit interference pattern is again obtained after a sufficiently large number of electrons have gone through the slits. A similar experiment with photons produces similar results.[8] This shows that the interference pattern is not due to the interaction between electrons in the beam.

As a wave it can split up and go through the two slits simultaneously and then recombine on the screen to produce a

[5] Young (1773–1829) was a British physicist.
[6] Davidson and Germer in New York in 1927 and Thomson and Reed in 1927 in Aberdeen, Scotland were the first to perform such experiments.
[7] Halliday, Resnick and Walker §3.8.5, §3.8.6, Feynman, Leighton and Sands §1 and §2.
[8] Hey and Walters Chapter 1. Halliday, Resnick and Walker §3.8.5.

double-slit interference pattern. As a particle the electron is a single indivisible object which cannot be split into two halves.[9] If each individual electron in the beam does not slit up, then half of the electrons in the beam would have to go through the upper slit and the other half would go though the lower slit. This would then give rise to two single-slit patterns on the screen since there would be no interference of the particles emerging from the upper and low slits. But the observed double-slit interference pattern is quite distinct from two single-slit patterns. So, we are confronted by the following question[10]:

> *How does each electron physically go through the double-slit to produce the interference pattern?*

34.3 Classical and Quantum Divide

Classical systems, e.g., everyday objects, are physically made up of quantum systems, electrons, protons, neutrons, atoms and so on. One would expect classical mechanics to be derivable from quantum mechanics. All is needed would be to construct a quantum theory for many-particle systems and then take the limit as the number of particles gets larger and larger, in much the same way as Newtonian mechanics is recovered from Special Relativity for objects whose the speed is much smaller than the speed of light. Indeed it is possible to obtain the classical limits from a quantum theory in some specific cases. However, no general theory has been found to derive classical mechanics from quantum mechanics. This is not just a matter of technical complexity of the problem. The crux of the matter is that quantum mechanics cannot be entirely independent of classical physics, since the mathematical formalism of quantum mechanics is understandable only in relation to measurements which use measuring devices with classical properties, e.g., a measuring device produces definite values in an experiment. The

[9] Feynman, Leighton and Sands §1-6 has a detailed discussion on the indivisibility of electrons.
[10] Gasiorowicz pp. 318–320.

situation is summarised by Landau and Lifshitz in their book *Quantum Mechanics* as follows[11]:

> *Thus quantum mechanics occupies a very unusual place among physical theories: it contains classical mechanics as a limiting case, yet at the same time it requires this limiting case for its own formulation*.

We can see a "chicken and egg" situation here. A simple-minded application of quantum formalism to classical objects would lead to paradoxical conclusions. A well-known example is given by Schrödinger in the form of a cat, as seen in the next section.

34.4 Schrödinger's Cat Paradox

As early as 1935 Schrödinger was aware of some of the fundamental difficulties in interpreting quantum formalism, especially when macroscopic systems are involved. He illustrates the difficulty with an example which becomes known as the Schrödinger's cat paradox[12]:

> *A cat is placed in a steel chamber, together with the following hellish contraption (which must be protected against direct interference by the cat): In a Geiger counter there is a tiny amount of radiative substance, so tiny that maybe within an hour one of the atoms decays, but equally probably none of them decays. If one decays then the counter triggers and via a relay activates a little hammer which breaks a container of cyanide. If one has left this entire system for an hour, then one would say that the cat is still living if no atom has decayed. The first decay would have poisoned it. The wave function of the entire system would express this by containing equal part of the living and the dead cat.*

[11] Landau and Lifshitz p. 3. Landau (1908–1968) and Lifshitz (1915–1985) were Soviet theoretical physicists. Landau won the 1962 Nobel Prize.
[12] The quotation is taken from Jauch p. 185.

Quantum mechanically we would describe the atom as being in a superposition of two states, the undecayed state and the decayed state. This probabilistic state of affair of the atom is transmitted to the cat. If we apply orthodox quantum formalism directly to the cat we would represent the state of a living cat by a state vector $\vec{\Phi}_l$ and the state of a dead cat by another state vector $\vec{\Phi}_d$ in the state space of the cat. Then the state vector of the cat in the steel chamber would be a superposition of $\vec{\Phi}_l$ and $\vec{\Phi}_d$, i.e.,

$$\vec{\Phi} = \frac{1}{\sqrt{2}}\left(\vec{\Phi}_l + \vec{\Phi}_d\right). \tag{34.1}$$

In the absence of a superselection rule separating the live state of the cat from the dead state of the cat such a superposition would represent the cat to be in a state of "living dead." But we do not have such zombi cats in real life, or do we?[13]

34.5 De Broglie Paradox and Non-Locality

In 1951 de Broglie proposed a thought experiment which becomes known as the de Broglie paradox.[14] The experimental apparatus consists of a box with impenetrable walls which can be divided into two halves, B_1 and B_2, by a sliding wall in the middle of the box. Suppose that there is an electron inside the box. The box is then divided into B_1 and B_2 by the sliding wall and B_2 is moved to Tokyo and B_1 to Paris. Suppose one then opens box B_1 in Paris and finds the electron inside. The question is whether the electron was inside B_1 before one opened the box and found the electron inside. There are two cases to consider:

Case 1. Suppose that the electron was initially localised on one half of the box, i.e., the half denoted by B_1, and that the insertion of the sliding wall does not disturb this localisation. In other words, the

[13] There has been a great deal of research theoretically and experimentally on this paradox, e.g., in formulating systems which resemble the Schrödinger's cat. *See* Auletta, G, Fortunato, M and Parisi pp. 533–540. Wan pp. 589–619. Wan and Menzies.

[14] de Broglie (1892–1987) was a French physicist.

electron's wave function was initially zero on the other half of the box and remains so after the partition. Then, according to quantum mechanics the electron is in B_1 all the time, i.e., before and after the box is opened. No paradoxical situation arises.

Case 2. Suppose that at time $t = 0$ the electron is not initially localised on any half of box, i.e., the wave function $\phi(x, 0)$ at $t = 0$ was a superposition of two parts with one part $\phi_1(x, 0)$ localised in B_1 and the other part $\phi_2(x, 0)$ localised in B_2, e.g., we have

$$\phi(x, 0) = \frac{1}{\sqrt{2}}\big(\phi_1(x, 0) + \phi_2(x, 0)\big), \tag{34.2}$$

where

$$\phi_1(x, 0) = 0 \ \ \forall x \in B_2, \quad \phi_2(x, 0) = 0 \ \ \forall x \in B_1 \tag{34.3}$$

Suppose the partition of the box into B_1, B_2 does not disturb this situation. Now move B_1 to location Λ_1 and B_2 to location Λ_2 where Λ_1 and Λ_2 are far apart so that the wave function after the separation at time T becomes[15]

$$\phi(x, T) = \frac{1}{\sqrt{2}}\big(\phi_1(x, T) + \phi_2(x, T)\big), \tag{34.4}$$

where $\phi_1(x, T)$ and $\phi_2(x, T)$ are localised in Λ_1 and Λ_2, respectively. Then a paradoxical situation arises when B_1 is opened and the particle is found inside. Before box B_1 was opened there was a probability of the electron being in B_2 since the wave function was not zero in B_2. In other words, the particle was not actually in B_1. As discussed in §30.2.2 the detection of the particle in B_1 is a measurement of the local position observable $\widehat{M}^{\hat{x}}(\Lambda_1)$. A positive result projects the wave function $\phi(x, T)$ onto $\phi_1(x, T)$ in accordance with Eq. (30.6), i.e., the wave function $\phi(x, T)$ would collapse onto $\phi_1(x, T)$ due to the measurement. It follows that after finding the particle in B_1, the wave function in B_2 must vanish and there is then no probability of finding the electron in B_2.

It is difficult to see how the opening of box B_1 can instantly affect whatever that is in B_2 which is very far away. This apparent

[15] Wan pp. 586–589. Instead of a thought experiment with a box and a sliding partition wall it is possible to formulate mathematical models with two parts of the a wave function in Eq. (34.4) moving apart in time to achieve a large spatial separation.

instant influence of what happens in B_1 on B_2 appears to come into conflict with the spirit of Special Relativity on the finite speed of transmission of influences, i.e., the instant influence of a disjoint and distant part of the wave function by a local measurement act. Generally the effect of the projection postulate which causes the collapse of the wave packet appears to be *non-local*, i.e., *action-at-a-distance*, in nature. Moreover, quantum mechanics does not provide any mechanism for this transmission of influence. The situation, commonly referred to as **quantum non-locality**, is therefore paradoxical.[16]

34.6 Entanglement and EPR Paradox

There has been a great deal of interest in many-particle systems exhibiting what has become known as **entanglement** ever since the problem was first raised in 1935 by Einstein, Podolsky and Rosen (EPR).[17] There are experiments showing strong evidence of entanglement even for macroscopic quantum systems, e.g., experiments on the entanglement of two superconducting rings with a capacitive junction.[18] Here we shall present a brief introduction of the subject.

Consider a system composed of two distinguishable particles.[19] Let us suppose that the two particles interact and as a result of the interaction the two-particle state cannot be factorised into a single tensor product of two one-particle states. For example, at time t we could have an entangled state vector of the form of Eq. (33.3), i.e.,

$$\vec{\Phi}^{(c)}(t) = \frac{1}{\sqrt{2}}\left(\vec{\phi}^{(1)}(t) \otimes \vec{\phi}^{(2)}(t) + \vec{\psi}^{(1)}(t) \otimes \vec{\psi}^{(2)}(t)\right). \qquad (34.5)$$

[16] One may try to resolve the paradox in terms of superselection rules similar to the situation when the Pauli exclusion principle becomes ineffective, i.e., at large separation $\phi(x, T)$ in Eq. (34.4) becomes a classical mixture.

[17] Einstein, Podolsky and Rosen (1935). *See* Isham pp. 179–185 for a discussion on EPR paradox and Bell inequalities. For a book popularising entanglement, see Aczel (2002).

[18] Berkley et al. (2003), Johnson et al. (2003).

[19] This is to avoid the complications which may arise when the particles are indistinguishable.

Suppose the two particles move freely apart after the initial interaction with the state vector maintaining the entangled form, i.e., at time T the state vector becomes[20]

$$\vec{\Phi}^{(c)}(T) = \frac{1}{\sqrt{2}}\left(\vec{\phi}^{(1)}(T) \otimes \vec{\phi}^{(2)}(T) + \vec{\psi}^{(1)}(T) \otimes \vec{\psi}^{(2)}(T)\right), \quad (34.6)$$

where $\vec{\phi}^{(1)}(T)$ and $\vec{\psi}^{(1)}(T)$ are localised in a region far away from the region of localisation of $\vec{\phi}^{(2)}(T)$ and $\vec{\psi}^{(2)}(T)$. Now suppose a measurement of an appropriate observable of particle 2 is carried out which projects the state of particle 2 onto the vector $\vec{\psi}^{(2)}(T)$. As a result, the two-particle state will change

$$\text{from} \quad \vec{\Phi}^{(c)}(T) \quad \text{to} \quad \vec{\psi}^{(1)}(T) \otimes \vec{\psi}^{(2)}(T). \quad (34.7)$$

Let us illustrate the situation more vividly with an example of two distinguishable spin-$\frac{1}{2}$ particles with the following entangled state vector which is a coherent superposition of two state vectors:

$$\vec{\Phi}^{(c)}(T) = \frac{1}{\sqrt{2}}\left(\left(\vec{\phi}^{(1)}(T) \otimes \vec{\alpha}_z^{(1)}\right) \otimes \left(\vec{\phi}^{(2)}(T) \otimes \vec{\beta}_z^{(2)}\right) + \left(\vec{\psi}^{(1)}(T) \otimes \vec{\beta}_z^{(1)}\right) \otimes \left(\vec{\psi}^{(2)}(T) \otimes \vec{\alpha}_z^{(2)}\right)\right). \quad (34.8)$$

A measurement of z-component spin of particle 2 resulting in the value $-\hbar/2$ will project the above state from

$$\vec{\Phi}^{(c)}(T) \quad \text{to} \quad \left(\vec{\phi}^{(1)}(T) \otimes \vec{\alpha}_z^{(1)}\right) \otimes \left(\vec{\phi}^{(2)}(T) \otimes \vec{\beta}_z^{(2)}\right). \quad (34.9)$$

This means that the state of particle 1 is also forced to change, e.g., its spin is now in state described by state vector $\vec{\alpha}_z^{(1)}$, instead of a combination of $\vec{\alpha}_z^{(1)}$ and $\vec{\beta}_z^{(1)}$ before the measurement. The situation is similar to that of the de Broglie paradox in that a non-local effect takes place with the action on particle 2 affecting particle 1 at a distance away. This distant correlations between two or more particles are known as the **EPR paradox** or **entanglement**. Opposite to entanglement is a process of **decoherence** which may provide an understanding to some of the paradoxes discussed so far.[21]

[20] Wan pp. 620–628.

[21] Wan pp. 581–633 for attempts to understand distant correlations in terms of asymptotically separable quantum mechanics.

34.7 Quantum Measurement

34.7.1 The Measurement Problem

The interpretation of the formalism of quantum theory is based on the results from measurements. Unfortunately the generally accepted formalism of quantum mechanics is lacking in details on the subject of measurement. The projection postulate takes a "black box" approach and it does not make any reference to any measuring process nor the duration of the process. It is highly controversial for the following reasons:

1. *Mathematical problem* The transition from the initial state to a final state in a single measurement act is implicitly assumed to be instantaneous in the projection postulate. It follows that there can be no dynamical evolution process for the transition from the initial state to a final state and no dynamical equation for such measuring processes. As von Neumann pointed out the projection postulate involved discontinuous and non-causal changes while quantum evolution given by Postulates 29.1.2(SP) and 29.2.1(HP) is continuous and causal.[22] The transition from the initial to the final state as shown in Eq. (31.1) is probabilistic and irreversible. This contrasts sharply with the deterministic and reversible quantum evolution given by Postulates 29.1.2(SP) and 29.2.1(HP). The transition *from an initial pure state to a classical mixture of state* shown in Eq. (31.1) precludes a unitary evolution process which would produce a final pure state from any given initial pure state.

2. *Conceptual Problems* The instantaneous change of the initial state $\vec{\phi}_i$ to a final state $\vec{\phi}_f$ causes the collapse or reduction of the wave packet and various non-local effects, e.g., the de Broglie and EPR paradoxes.[23]

3. *Non-universal applicability* Not all measurements obey this postulate. By detecting a photon (a photon position measurement) we may also destroy the photon, i.e., there is no projection of an

[22] von Neumann p. 349.

[23] For more discussion see the book edited by Cini, M and Levy-Labond, J. M which contains a number of review articles.

initial state of the photon to a final state of the photon. Instead we have an annihilation of the photon. The position obtained by the measurement is not the position of the photon after the measurement since the photon no longer exists.

4. *Measuring devices and measuring processes* A measuring device should have some quantum properties so that it can be coupled to a quantum system to perform a measurement. It should also have some classical properties in that it must produce results in the form of definite values. This means that a measuring device cannot be an orthodox quantum system.[24] We need to establish a theory for the kind of systems which would embody these properties of measuring devices.

A great deal of effects have been made to tackle the measurement problem, with numerous articles and monographs devoted on this subject.[25] Instead of the "black box" approach of the projection postulate one should examine some simple examples to illustrate the problems involved. This should include an investigation into some actual physical measurement processes. We shall demonstrate such an approach in the following section.

34.7.2 Measuring Devices and Processes

34.7.2.1 Introduction

Consider an orthodox quantum system with a two-dimensional state space $\vec{\mathcal{H}}^{(q)}$. Let $\widehat{A}^{(q)}$ be an observable of such a system which has two orthonormal unit eigenvectors $\vec{\varphi}_-$ and $\vec{\varphi}_+$ corresponding to eigenvalues a_- and a_+, respectively. Suppose we want to measure this observable when the system is state ϕ^s given by the following state vector:

$$\vec{\phi} = c_- \vec{\varphi}_- + c_+ \vec{\varphi}_+, \quad |c_-|^2 + |c_+|^2 = 1. \tag{34.10}$$

This is a pure state, being a coherent superposition of states φ^s_- and φ^s_+. In this state $\widehat{A}^{(q)}$ does not possess a value. We need to

[24] An orthodox quantum system is described by Definition 26.1(1). A quantum system with a superselection rule is not an orthodox quantum system.

[25] Jauch pp. 160–185. Busch, Lahti and Mittelstaedt. Auletta, Fortunato and Parisi Chapter 9. Wan §3.5 and §3.6.

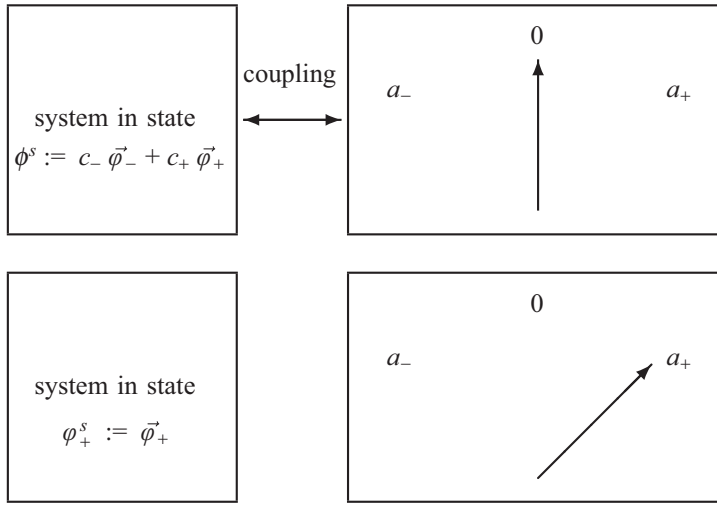


Figure 34.1 Coupling of quantum system and measuring device.

couple the quantum system to a measuring device which will force the quantum system into a classical mixture of states φ^s_- and φ^s_+, i.e., the system is forced into a pure state, either φ^s_- or φ^s_+ after each measurement act. It would then possess a value. One way to achieve such a transition is to require the measuring device and the measuring process to have the following two properties:

P34.7.2(1) The measuring device should be a mixed quantum system (*see* Definition 32.1(4)) capable of coupling to an orthodox quantum system under measurement to form a compound system.

P34.7.2(2) The coupling interaction should be such that the compound system inherits the superselection rule of the measuring device which would lead to the transition from a coherent superposition to a classical mixture to achieve a measurement.

P34.7.2(3) The first diagram in Fig. 34.1 above shows that at the start of the measurement the measuring device on the right has its pointer in the neutral 0 position.

The quantum system on the left is in a coherent superposition of φ^s_- and φ^s_+ given by Eq. (34.10). The second diagram in Fig. 34.1 shows that at the completion of the measurement the quantum

system is in state φ_+^s and the pointer has rotated to the position on the right, a position showing the value a_+.

34.7.2.2 Mathematical formulation

A formulation of the above measuring process is set out below[26]:

1. We adopt the example in §32.3 as a model for our measuring device:

(1) The model system possesses a superselection rule with three one-dimensional supersectors $\vec{\mathcal{S}}^{(-)}$, $\vec{\mathcal{S}}^{(0)}$ and $\vec{\mathcal{S}}^{(+)}$ spanned by three corresponding unit vectors $\vec{\eta}^{(-)}$, $\vec{\eta}^{(0)}$ and $\vec{\eta}^{(+)}$.[27] In other words, the state space of the system is a direct sum of these subspaces, i.e., the state space is $\vec{\mathcal{H}}^{(m)}$ in Eq. (32.11).

(2) Observables are represented by decomposable selfadjoint operators on $\vec{\mathcal{H}}^{(m)}$, i.e., $\widehat{B}^{(m)}$ in Eq. (32.19). These operators are also diagonalisable since all supersectors are one-dimensional.

(3) We can define a **pointer observable** in terms of a selfadjoint diagonalisable operator $\widehat{B}_{po}^{(m)}$ on $\vec{\mathcal{H}}^{(m)}$ which has $\vec{\eta}^{(-)\oplus}$, $\vec{\eta}^{(0)\oplus}$ and $\vec{\eta}^{(+)\oplus}$ as its eigenvectors corresponding to eigenvalues a_-, a_0, a_+, respectively, i.e.,[28]

$$\widehat{B}_{po}^{(m)} = a_- \, \widehat{I\!I}^{(-)} \oplus a_0 \, \widehat{I\!I}^{(0)} \oplus a_+ \, \widehat{I\!I}^{(+)}. \tag{34.11}$$

The vector $\vec{\eta}^{(0)\oplus}$ corresponds to the pointer of the measuring device in the neutral 0 position while $\vec{\eta}^{(-)\oplus}$ and $\vec{\eta}^{(+)\oplus}$ correspond, respectively, to the pointer pointing to the left at the value a_- and right at the value a_+. These vectors represent pointer states corresponding to the pointer pointing in different positions.

(4) There is no coherent superposition of different pointer states.

2. At the start of measurement the measuring device is coupled to the orthodox system to form a compound system. The state space

[26] Wan pp. 384–391. Beltrametti and Cassinelli (1981), Bub (1988), and van Fraassen (1991).

[27] *See* §32.3 for a discussion of supersectors.

[28] These vectors are defined by Eqs. (32.14) to (32.16).

$\vec{\mathcal{H}}^{(q,m)}$ of the compound system is the following tensor product:

$$\vec{\mathcal{H}}^{(q,m)} := \vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{H}}^{(m)} = \vec{\mathcal{H}}^{(q)} \otimes \big(\vec{\mathcal{S}}^{(-)} \oplus \vec{\mathcal{S}}^{(0)} \oplus \vec{\mathcal{S}}^{(+)}\big). \quad (34.12)$$

This tensor product space has the following direct sum decomposition:

$$\vec{\mathcal{H}}^{(q,m)} = \big(\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(-)}\big) \oplus \big(\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(0)}\big) \oplus \big(\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(+)}\big). \quad (34.13)$$

3. On account of P34.7.2(2) we assume that the interaction between the quantum system and the measuring device is such that the above direct sum decomposition generates a superselection rule for the compound system with

$$\big(\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(-)}\big),\ \big(\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(0)}\big),\ \big(\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(+)}\big) \quad (34.14)$$

as supersectors. It is in this sense that we say the compound system inherits the superselection rule of the measuring device. Physically this means that the interaction between measuring device and the quantum system would destroy the coherence of the two pure states φ^s_- and φ^s_+ of the quantum system. In other words, the following state vector

$$c_-\big(\vec{\varphi}_- \otimes \vec{\eta}^{(-)}\big) \oplus c_+\big(\vec{\varphi}_+ \otimes \vec{\eta}^{(+)}\big) \quad (34.15)$$

describes a classical mixture of states corresponding to state vectors $\vec{\varphi}_- \otimes \vec{\eta}^{(-)}$ and $\vec{\varphi}_+ \otimes \vec{\eta}^{(+)}$.

4. Suppose at the start of measurement at $t = 0$ the compound system is in a state described by the following vector:

$$\vec{\Phi}^{(q,m)}(0) := \vec{\phi} \otimes \vec{\eta}^{(0)\oplus} = \big(c_-\vec{\varphi}_- + c_+\vec{\varphi}_+\big) \otimes \vec{\eta}^{(0)\oplus}. \quad (34.16)$$

This vector which is in the supersector $\vec{\mathcal{H}}^{(q)} \otimes \vec{\mathcal{S}}^{(0)}$ represents a pure state with φ^s_- and φ^s_+ in a coherent superposition. The observable $\widehat{A}^{(q)}$ of the quantum system does not possess a value in this pure state.

5. We need to engineer an interaction to evolve the state vector in Eq. (34.16) into the state vector in Eq. (34.15) at the end of a measuring process at time $t = T$. This can be done with an

evolution operator generated by an operator on $\mathcal{H}^{(q,m)}$ which is not decomposable. The following operator satisfies our need:

$$\widehat{H}^{(q,m)} := \rho \left(\widehat{P}_{-}^{(q)} \otimes \widehat{L}_{-}^{(m)} + \widehat{P}_{+}^{(q)} \otimes \widehat{L}_{+}^{(m)} \right), \quad \rho \in I\!R, \tag{34.17}$$

where $\widehat{L}_{-}^{(m)}$, $\widehat{L}_{+}^{(m)}$ are defined by Eqs. (24.78) and (24.79), and

$$\widehat{P}_{-}^{(q)} = |\vec{\varphi}_{-}\rangle\langle\vec{\varphi}_{-}|, \quad \widehat{P}_{+}^{(q)} = |\vec{\varphi}_{+}\rangle\langle\vec{\varphi}_{+}|. \tag{34.18}$$

All these operators are selfadjoint. We now consider the evolution of the compound system in the Schrödinger picture with $\widehat{H}^{(q,m)}$ as the Hamiltonian. The evolution operators are given by Eq. (29.9), i.e.,

$$\widehat{U}(\widehat{H}^{(q,m)}, t) := \exp\left(- \dot{\imath}\, \widehat{H}^{(q,m)}\, t \right). \tag{34.19}$$

The initial state vector in Eq. (34.16) will evolve according to

$$\begin{aligned} \vec{\Phi}^{(q,m)}(t) &:= \widehat{U}(\widehat{H}^{(q,m)}, t)\, \vec{\Phi}^{(q,m)}(0) \\ &= \widehat{U}(\widehat{H}^{(q,m)}, t) \left\{ \left(c_{-}\vec{\varphi}_{-} + c_{+}\vec{\varphi}_{+} \right) \otimes \vec{\eta}^{\,(0)\oplus} \right\}. \end{aligned} \tag{34.20}$$

We also have, on account of Stone's theorem, the following Schrödinger equation:

$$i\hbar \frac{d\vec{\Phi}^{(q,m)}(t)}{dt} = \widehat{H}^{(q,m)} \vec{\Phi}^{(q,m)}(t). \tag{34.21}$$

For the initial state vector $\vec{\Phi}^{(q,m)}(0)$ the solution of Eq. (34.21) is

$$\begin{aligned} \vec{\Phi}^{(q,m)}(t) = {} & c_{-}\, \vec{\varphi}_{-} \otimes \left(\cos(\rho t/\hbar)\, \vec{\eta}^{\,(0)\oplus} - i\, \sin(\rho t/\hbar)\, \vec{\eta}^{\,(-)\oplus} \right) \\ & + c_{+}\vec{\varphi}_{+} \otimes \left(\cos(\rho t/\hbar)\, \vec{\eta}^{\,(0)\oplus} - i\, \sin(\rho t/\hbar) \eta^{\,(+)\oplus} \right). \end{aligned} \tag{34.22}$$

At time $t = T = \pi\hbar/2\rho$ the solution becomes[29]

$$\vec{\Phi}^{(q,m)}(T) = -i \left(c_{-}\, \vec{\varphi}_{-} \otimes \vec{\eta}^{\,(-)\oplus} + c_{+}\, \vec{\varphi}_{+} \otimes \vec{\eta}^{\,(+)\oplus} \right). \tag{34.23}$$

[29] The phase factor $-i$ has no significance in the interpretation of the state.

6. The interpretation is as follows:

(1) Physically we can visualise the measuring device having a pointer which would be at a neutral position at the start of the measurement. This situation is represented by the vector $\vec{\Phi}^{(q,m)}(0)$ in Eq. (34.16).

(2) The Hamiltonian $\widehat{H}^{(q,m)}$ which generates the evolution of the compound system is not an observable of the compound system.

(3) Under the interaction between the orthodox quantum system and the measuring device due to the Hamiltonian $\widehat{H}^{(q,m)}$ the initial state vector $\vec{\Phi}^{(q,m)}(0)$ evolves into $\vec{\Phi}^{(q,m)}(T)$ at the end of the measuring process at $t = T$. The final vector $\vec{\Phi}^{(q,m)}(T)$ represents a classical mixture of pure states described by the state vectors $\vec{\varphi}_- \otimes \vec{\eta}^{(-)\oplus}$ and $\vec{\varphi}_+ \otimes \vec{\eta}^{(+)\oplus}$ on account of the superselection rule. Physically this corresponds to the pointer pointing the left or to the right at the end of each measurement with probabilities $|c_-|^2$ and $|c_+|^2$, respectively.

(4) Observable $\widehat{A}^{(q)}$ would possess a value after each measurement. The value would be a_- if the pointer swings to the left and a_+ if the pointer swings to the right.

(5) The duration of measurement T can be made very short by choosing the parameter ρ to be suitably large.

34.7.2.3 Concluding remarks

The above model gives us an understanding of quantum measurement beyond the projection postulate. However, actually measurement processes are more complicated than the simple model described above. An example is the case of position measurement. Let us look into how position measurement is achieved physically. One way to measure position is to use a photographic plate[30]: The physical processes involved is well-known[31]:

(1) *The interaction stage* A photographic plate is composed of a large numbers of grains each of which is formed by billions

[30] Wan pp. 310–330. Geiger counter and cloud chamber work in a similar way.
[31] Hey and Walters pp. 19–20. Mees.

of silver bromide molecules. With sufficient energy an ionizing particle, the quantum system whose position is to be measured, coming into the photographic plate may interact with a few silver bromide molecules in a grain of billions of silver bromide molecules. The interaction may lead to these few silver bromide molecules to be ionised, i.e., the silver atoms get separated from the molecules. The ionisation process is describable as a quantum mechanical evolution process.

(2) *The magnification stage* Having been separated out from the molecules these few silver atoms will form a small but stable cluster of neutral silver atoms. Such a cluster is known as a *latent image*, an image not yet visible to the naked eye. The photographic plate can then be developed during which a silver cluster can cause all the silver atoms in the grain to separate out to form a visible grain of silver atoms. This magnification process does not have to follow the interaction stage immediately, i.e., a photographic plate can be developed a long time after it was exposed. In other words, the measuring process is effectively completed when the ionisation process is achieved.

So, to establish a theory for position measurement, we have to have a theory for the ionisation of process.[32]

34.8 Quantum Theory, Relativity and Photons

There are many attempts to render quantum theory compatible with Special Relativity with only various degree of success. As yet we do not have a totally satisfactory unification of two of the most amazing and successful theories in physics. The problem becomes even more problematical when it comes to combine quantum theory with General Relativity. We shall not consider the theory of massless particles like photons here.[33] This book deals with massive particles and non-relativistic theories. Massless particles

[32] Wan pp. 324–327.
[33] Most experiments on entanglement are performed on photons.

are necessarily relativistic in nature. There are fundamentally new issues to be addressed. Even the definition of the position observable of a massless particle is very complicated.[34] Take the example of photons. Any questions on the position, distances travelled and the "size" of photons are difficult to answer. Consider intuitively the situation in Bohr's theory of the hydrogen atom which says that

(1) A photon is emitted by the atom when the orbiting electron makes a transition from a higher energy orbit to a lower energy orbit.
(2) The reverse process also takes place, i.e., the atom can absorb an appropriate amount of energy in the form of a photon. As a result the orbiting electron makes a transition from a lower energy orbit to a higher energy orbit.

The atom is of a linear dimension of 0.1 nanometre. Then the question on the size of the photon emitted presents itself. For the sake of argument, it seems to be intuitively reasonable to assume the size of the photon to be of the order of its wavelength which has a linear dimension of several hundred nanometres. Then one has a situation of a tiny atom emitting something several thousand times bigger than itself. This is like an explosion which produces a shock wave much bigger than the explosive which the source of the explosion. Even harder to imagine is the absorption process, i.e., a tiny atom has to "swallow" something several thousand times bigger than itself! Similar problems emerge when considering how a photon travels in matter. In condense matter the inter-atomic distances are of the order of nanometres. It is not easy to visualise how a photon, which is huge compared with inter-atomic distances, travels in the free space between atoms and how one can measure how fast such a huge photon would travel between two atoms separated by only a fraction of the size of the photon.[35]

[34] Kraus (1977), Schroeck (1996), and Bacry (1988).
[35] In classical physics this is not a problem since the distances involved are huge compared with the wavelength of the photon.

Exercise and Problem

Q34(1) Verify that $\vec{\Phi}^{(q,m)}(t)$ in Eq. (34.22) satisfies the initial condition in Eq. (34.16) and the Schrödinger equation (34.21).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Section IV

Illustrative Applications

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 35

Harmonic and Isotropic Oscillators

35.1 Harmonic Oscillators

Harmonic oscillators are important in physical applications and in demonstrating the working of the postulates of quantum formalism. A classical harmonic oscillator has a state space, i.e., its phase space, coordinated by the oscillator's position x and momentum p. Observables correspond to real-valued functions on the state space. The evolution of the state is determined by Hamilton's equations, which reduce to Eq. (27.12). The time dependence of the state automatically determines the time dependence of observables.

The description of a quantum harmonic oscillator is fundamentally different from that of a classical oscillator:

1. States Postulate 25.1(PS) applies. The state space of a quantum harmonic oscillator is an infinite-dimensional Hilbert space. A pure state ϕ^s corresponds to a unit vector $\vec{\phi}$ of the Hilbert space. A state cannot be determined by two numbers as in the classical case.

2. Observables Postulate 26.1(OV) applies. Quantum observables correspond to selfadjoint operators in the state space. Moreover, we consider a quantised oscillator as an orthodox quantum system.

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

3. Explicit representation To obtain explicit expressions to describe the state and observables we have to choose an explicit Hilbert space to represent the state space. In the coordinate representation the Hilbert space is chosen to be $\vec{L}^2(I\!R)$ defined by the space $L^2(I\!R)$ of square-integrable complex-valued functions of the position variable x. A generally accepted choice for the position and momentum operators are $\widehat{x}$ and $\widehat{p}$, a choice consistent with Postulate 27.2(CQ) on quantisation.[1] The Hamiltonian is initially quantised in accordance with Eq. (27.93). This results in the operator

$$\widehat{H}_{ho} = \frac{1}{2m}\,\widehat{p}^{\,2} + \frac{1}{2}\,m\omega\widehat{x}^{\,2}, \tag{35.1}$$

defined on the domain $\mathcal{D}(\widehat{H}_{ho}) = \mathcal{D}(\widehat{p}^{\,2}) \cap \mathcal{D}(\widehat{x}^{\,2})$. This is the operator in Eq. (19.54). Its restriction to the Schwartz space $\vec{S}_s(I\!R)$ is denoted by[2]

$$\widehat{H}_{ho\vec{S}_s} = \frac{1}{2m}\,\widehat{p}^{\,2}_{\vec{S}_s} + \frac{1}{2}\,m\omega\widehat{x}^{\,2}_{\vec{S}_s}, \tag{35.2}$$

This operator is essentially selfadjoint with its unique selfadjoint extension equal to $\widehat{H}_{ho}$.[3] The quantised operator is therefore $\widehat{H}_{ho}$. This quantised operator is well-known to have a discrete set of nondegenerate eigenvalues eigenvalues E_n corresponding to a complete orthonormal set of eigenvectors $\vec{\varphi}_n$, i.e., we have.[4]

$$\widehat{H}_{ho}\vec{\varphi}_n = E_n\,\vec{\varphi}_n, \qquad E_n = \Big(n + \frac{1}{2}\Big)\,\hbar\omega, \tag{35.3}$$

where $n = 0, 1, 2\ldots$. These eigenvectors generate a complete orthogonal family of eigenprojectors $\widehat{P}_{\vec{\varphi}_n} = |\vec{\varphi}_n\rangle\langle\vec{\varphi}_n|$.

4. Measurement and probability distribution Postulates 28.1(PDDO) and 30.1.1(PPDO) apply. A measurement will have to be performed to yield a value of an observable. This will generally cause a change of the initial state. As an example consider energy measurement. There are two distinct cases:

---

[1] *See* Eq. (27.78).

[2] $\widehat{p}_{\vec{S}_s}$ and $\widehat{x}_{\vec{S}_s}$ are the restrictions of $\widehat{p}$ and $\widehat{x}$ to the Schwartz space $\vec{S}_s(I\!R)$ previously denoted by $\widehat{p}_{\vec{S}_s}(I\!R)$ and $\widehat{x}_{\vec{S}_s}(I\!R)$. *See* Eq. (27.94) and its footnote.

[3] Boretti p. 401 (the expressions for A^* and A have a different sign). We shall discuss this further in §35.2.2. *See* also comments after Eq. (35.15).

[4] *See* §35.2 for a derivation of these values.

(1) If the initial state φ_2^s is given by the eigenvector $\vec{\varphi}_2$ in Eq. (35.3) then an energy measurement will yield the value $E_2 = \frac{5}{2}\hbar\omega$ and the state vector $\vec{\varphi}_2$ will remain unchanged during the measurement, i.e., the state vector before the measurement, will remain the state vector after the measurement. In this case the measured result E_2 can be predicted with certainty from the initial state before the measurement, a situation similar to the case of a classical measurement. The oscillator is said to possess the value E_2 in state φ_2^s.

(2) If the initial state ϕ^s is given by a state vector $\vec{\phi}$ which is not an eigenvector of the Hamiltonian, e.g.,[5]

$$\vec{\phi} = \frac{1}{\sqrt{3}}\,\vec{\varphi}_1 + \sqrt{\frac{2}{3}}\,\vec{\varphi}_2. \tag{35.4}$$

Then Postulate 28.1(PDDO) tells us that

(a) The oscillator in state ϕ^s does not possess an energy value. We cannot predict with certainty what value a measurement will give, it could be E_1 or E_2.

(b) We cannot predict with certainty what the final state will be after a measurement. It depends on the result of the measurement. If a measurement results in the value E_2, then according to Postulate 30.1.1(PPDO) the state after will be φ_2^s corresponding to state vector $\vec{\varphi}_2$.

(c) We can predict the probability distribution of measured results E_n and the expectation value. The probabilities are given by $\langle \vec{\phi} \mid \widehat{P}_{\vec{\varphi}_n} \vec{\phi} \rangle$.[6] Explicitly we have

$$\wp^{H_{ho}}(\phi^s, E_1) = 1/3, \tag{35.5}$$

$$\wp^{H_{ho}}(\phi^s, E_2) = 2/3, \tag{35.6}$$

$$\wp^{H_{ho}}(\phi^s, E_n) = 0 \text{ if } n \neq 1, 2. \tag{35.7}$$

The expectation value

$$\mathcal{E}(H_{ho}, \phi^s) = \sum_{n=0}^{\infty} E_n \wp^{H_{ho}}(\phi^s, E_n) = \frac{13}{6}\hbar\omega. \tag{35.8}$$

We can verify that $\mathcal{E}(H_{ho}, \phi^s) = \langle \vec{\phi} \mid \widehat{H}_{ho}\vec{\phi} \rangle$.

[5] This means that $\widehat{H}_{ho}\vec{\phi} \neq c\,\vec{\phi}$ for any $c \in I\!R$.

[6] The eigenvalues are nondegenerate so that $\widehat{P}^{\hat{H}_{ho}}(E_n) = \widehat{P}_{\vec{\varphi}_n}$.

The position and momentum of the oscillator are continuous observables. The discussion in §28.3 and §28.3.3 remains valid here since the discussion does not depend on any particular physical system.

35.2 Energy Eigenvalues and Eigenvectors

35.2.1 Derivation

The idea is to introduce a pair of annihilation and creation operators $\widehat{a}$ and $\widehat{a}^\dagger$ in the Hilbert space $\vec{L}^2(I\!R)$ in terms of linear combinations the position and momentum operators $\widehat{x}$ and $\widehat{p}$. A selfadjoint number operator $\widehat{N} = \widehat{a}^\dagger \widehat{a}$ can then be constructed to relate to the Hamiltonian $\widehat{H}_{ho}$ in such a way that the eigenvalues and eigenvectors of $\widehat{H}_{ho}$ can be obtained from that of the number operator.

Let us start with the defining expressions for $\widehat{a}$ and $\widehat{a}^\dagger$, i.e.,

$$\widehat{a} := \sqrt{\frac{\lambda}{2}} \left(\widehat{x} + \frac{i}{m\omega} \widehat{p} \right), \quad \widehat{a}^\dagger := \sqrt{\frac{\lambda}{2}} \left(\widehat{x} - \frac{i}{m\omega} \widehat{p} \right), \tag{35.9}$$

where $\lambda = m\omega/\hbar$. Both these operators are defined on the domain $\mathcal{D}(\widehat{p}\,) \cap \mathcal{D}(\widehat{x}\,)$. Moreover these operators are adjoint to each other, and they are therefore closed operators.[7]

The operators $\widehat{a}$ and $\widehat{a}^\dagger$ satisfy properties P27.9(1) to P27.9(4) of a pair of creation and annihilation operators presented in §27.9, i.e., we have the following properties[8]:

P35.2.1(1) These two operators are defined on the same dense domain and they are the adjoint of each other.

P35.2.1(2) They obey the following commutation relation[9]

$$[\,\widehat{a}, \widehat{a}^\dagger\,] = \widehat{I\!I}. \tag{35.10}$$

[7] Takhtajan pp. 103–104. We cannot just use Eq. (17.95) to prove this since the operators are unbounded. However, we can follow the comments in §19.3.4 on the proof of selfadjointness of $\widehat{p}(I\!R)$ to verify this.

[8] *See* Fano pp. 354–360 for an explicit derivation.

[9] The equality holds only in the domain $\mathcal{D}(\widehat{p}^2) \cap \mathcal{D}(\widehat{x}^2)$ since the product $\widehat{a}^\dagger \widehat{a}$ is involved (*see* the domain for $\widehat{N}$ in Eq. (35.14)).

P35.2.1(3) There exists a vector $\vec{\varphi}_0$ which $\widehat{a}$ would annihilate, i.e.,

$$\widehat{a}\,\vec{\varphi}_0 = \vec{0} \quad \text{or} \quad \Big(\widehat{x} + \frac{i}{m\omega}\,\widehat{p}\Big)\vec{\varphi}_0 = \vec{0}. \tag{35.11}$$

This vector corresponds to the normalised solution $\varphi_0(x)$ of the following equation:

$$\Big(x + \frac{\hbar}{m\omega}\,\frac{d}{dx}\Big)\varphi_0(x) = 0. \tag{35.12}$$

Expliclty we have

$$\varphi_0(x) = \left(\frac{\lambda}{\pi}\right)^{\frac{1}{4}} e^{-\frac{1}{2}\lambda x^2}, \quad \lambda = \frac{m\omega}{\hbar}. \tag{35.13}$$

P35.2.1(4) The operators $\widehat{a}$ and $\widehat{a}^\dagger$ are irreducible since $\widehat{x}$ and $\widehat{p}$ are irreducible in $\vec{L}^2(I\!R)$.[10]

It follows that $\widehat{a}$ and $\widehat{a}^\dagger$ are a pair of annihilation and creation operators in $\vec{L}^2(I\!R)$. The number operator $\widehat{N} := \widehat{a}^\dagger\widehat{a}$ is selfadjoint. Explicitly we have[11]

$$\widehat{N} := \widehat{a}^\dagger\widehat{a} = \frac{m\omega}{2\hbar}\left(\widehat{x}^{\,2} + \frac{1}{(m\omega)^2}\,\widehat{p}^{\,2} - \frac{\hbar}{m\omega}\right). \tag{35.14}$$

This operator is defined on the domain $\mathcal{D}(\widehat{p}^2) \cap \mathcal{D}(\widehat{x}^2)$. The Hamiltonian $\widehat{H}_{ho}$ in Eq. (35.1) is related to $\widehat{N}$ by[12]

$$\widehat{H}_{ho} = \hbar\omega\left(\widehat{N} + \frac{1}{2}\right). \tag{35.15}$$

We can see clearly that this Hamiltonian is selfadjoint and it possesses a discrete and nondegenerate spectrum

$$\big\{E_n = (n + 1/2)\hbar\omega : \; n = 0, 1, 2, 3 \ldots\big\}, \tag{35.16}$$

with a corresponding complete orthonormal set eigenvectors

$$\vec{\varphi}_n = \frac{1}{\sqrt{n!}}\,\big(\widehat{a}^\dagger\big)^n\,\vec{\varphi}_0. \tag{35.17}$$

---

[10] *See* E20.7(1). Jauch p. 214.

[11] We have expressed $\hbar/m\omega\,\widehat{I\!I}$ as $\hbar/m\omega$ for brevity.

[12] Takhtajan p. 104.

The corresponding functions $\varphi_n(x)$ are obtained from the function $\varphi_0(x)$ in Eq. (35.13) by successive differentiations which produce functions of the form

$$\varphi_n(x) = \varphi_0(x) \times \text{(polynomial of order } n \text{ in } x). \tag{35.18}$$

Detailed calculations show that these eigenfunctions, when normalised, are the same as the Hermite functions $\varphi_{Hn}(x)$ in Eq. (16.18).[13] Hence we shall denote $\vec{\varphi}_n$ in Eq. (35.17) by $\vec{\varphi}_{Hn}$ and rewrite the equation as[14]

$$\vec{\varphi}_{Hn} = \frac{1}{\sqrt{n!}} \left(\hat{a}^\dagger\right)^n \vec{\varphi}_{H0}. \tag{35.19}$$

Equations (16.22) to (16.23) apply, i.e.,

$$\hat{a}\vec{\varphi}_{H0} = \vec{0}, \tag{35.20}$$

$$\hat{a}\vec{\varphi}_{Hn} = \sqrt{n}\,\vec{\varphi}_{H(n-1)}, \quad n \geq 1 \tag{35.21}$$

$$\hat{a}^\dagger\vec{\varphi}_{Hn} = \sqrt{n+1}\,\vec{\varphi}_{H(n+1)} \quad n \geq 0. \tag{35.22}$$

We can appreciate the concept of creation and annihilation clearly here. When acting on the state vector $\vec{\varphi}_{Hn}$ having energy E_n the operator $\hat{a}$ leads to a new state vector $\vec{\varphi}_{H(n-1)}$ with a lower energy value E_{n-1}. Hence an amount of energy $E_n - E_{n-1} = \hbar\omega$ is annihilated. In particular, acting on the ground state vector $\vec{\varphi}_{H_0}$ the operator destroys the ground state itself together with the ground state energy E_0. The operator $\hat{a}^\dagger$ does the exact opposite. When acting on the state vector $\vec{\varphi}_{Hn}$ having energy E_n the operator $\hat{a}^\dagger$ leads to a new state vector $\vec{\varphi}_{H(n+1)}$ with a higher energy value E_{n+1}, e.g., acting on the ground state vector $\vec{\varphi}_{H0}$ the operator will lead to the first excited state vector $\vec{\varphi}_{H1}$. An amount of energy $E_{n+1} - E_n = \hbar\omega$ is created.

35.2.2 Mathematical Discussion

It is instructive to follow the procedure set out in P27.2(4) more closely in quantising the harmonic oscillator Hamiltonian. This

[13] Greiner p. 118 and Fano pp. 355–360. Jauch pp. 211–214. Moretti pp. 401–402. *See* Greiner p. 119 and Zettili pp. 232–234 for explicit examples.

[14] As pointed out in E16.2.2(4) these vectors form an orthonormal basis for $\vec{L}^2(I\!R)$.

means that we should introduce the annihilation and creation operators in the Schwartz space $\vec{S}_s(I\!R)$ to start with, i.e., we define two new operators $\widehat{a}^{\star}_{\vec{S}_s}$ and $\widehat{a}_{\vec{S}_s}$ on $\vec{S}_s(I\!R)$ by[15]

$$\widehat{a}_{\vec{S}_s} := \sqrt{\frac{\lambda}{2}}\left(\widehat{x}_{\vec{S}_s} + \frac{i}{m\omega}\,\widehat{p}_{\vec{S}_s}\right). \tag{35.23}$$

$$\widehat{a}^{\star}_{\vec{S}_s} := \sqrt{\frac{\lambda}{2}}\left(\widehat{x}_{\vec{S}_s} - \frac{i}{m\omega}\,\widehat{p}_{\vec{S}_s}\right), \tag{35.24}$$

These two operators satisfy the commutation relation[16]

$$[\,\widehat{a}_{\vec{S}_s}, \widehat{a}^{\star}_{\vec{S}_s}\,] = \widehat{I\!I}_{\vec{S}_s}. \tag{35.25}$$

We also have

$$\widehat{N}_{\vec{S}_s} := \widehat{a}^{\star}_{\vec{S}_s}\,\widehat{a}_{\vec{S}_s}, \tag{35.26}$$

$$\widehat{H}_{ho\vec{S}_s} = \hbar\omega\big(\,\widehat{N}_{\vec{S}_s} + \frac{1}{2}\big). \tag{35.27}$$

Then:

(1) We want to find appropriate extensions of $\widehat{a}_{\vec{S}_s}$ and $\widehat{a}^{\star}_{\vec{S}_s}$ to serve as annihilation and creation operators. What we need is to find a complete orthonormal set of vectors within $\vec{S}_s(I\!R)$ such that these extensions would satisfy Definitions 17.10(1) and 17.10(2). As expected from the discussion in the preceding subsection the desired complete orthonormal set is the set of vectors $\vec{\varphi}_{Hn}$ corresponding to the Hermite functions in Eq. (16.18). Since Hermite functions $\varphi_{Hn}(x)$ belong to $S_s(I\!R)$ their correspoding vectors $\vec{\varphi}_{Hn}$ would belong $\vec{S}_s(I\!R)$.

(2) The operators $\widehat{a}_{\vec{S}_s}$ and $\widehat{a}^{\star}_{\vec{S}_s}$ act like $\widehat{a}$ and $\widehat{a}^{\dagger}$ on $\vec{S}_s(I\!R)$.[17] It follows from Eqs. (35.20) to (35.22) that[18]

$$\widehat{a}_{\vec{S}_s}\,\vec{\varphi}_{H0} = \vec{0}, \tag{35.28}$$

$$\widehat{a}_{\vec{S}_s}\,\vec{\varphi}_{Hn} = \sqrt{n}\,\vec{\varphi}_{H(n-1)}, \quad n = 1, 2, 3\ldots, \tag{35.29}$$

$$\widehat{a}^{\star}_{\vec{S}_s}\,\vec{\varphi}_{Hn} = \sqrt{n+1}\,\vec{\varphi}_{H(n+1)}, \quad n = 0, 1, 2, 3\ldots. \tag{35.30}$$

---

[15] These are restrictions of $\widehat{a}$ and $\widehat{a}^{\dagger}$ in Eq. (35.9). Because of Eqs. (17.97) and (17.101) the operator $\widehat{a}^{\star}_{\vec{S}_s}$ is not the adjoint of $\widehat{a}_{\vec{S}_s}$, hence the notation (*see* Moretti p. 402). Here $\widehat{x}_{\vec{S}_s}$ and $\widehat{p}_{\vec{S}_s}$ are the same as $\widehat{x}_{\vec{S}_s}(I\!R)$ in E17.6(2) and $\widehat{p}_{\vec{S}_s}(I\!R)$ in Eq. (17.48) for the restrictions of $\widehat{x}$ and of $\widehat{p}$ to $\vec{S}_s(I\!R)$.

[16] Here $\widehat{I\!I}_{\vec{S}_s}$ is the restriction of $\widehat{I\!I}$ to $\vec{S}_s$. The equality holds without further restrictions

[17] The operators $\widehat{a}$ and $\widehat{a}^{\dagger}$ are defined by Eq. (35.9).

[18] *See* Eqs. (16.21) to (16.23).

(3) We can now extend the operation of $\widehat{a}^{\star}_{\vec{S}_s}$ and $\widehat{a}_{\vec{S}_s}$ to the domain[19]

$$\vec{\mathcal{D}} = \{\vec{\phi} \in \vec{L}^2(I\!R) : \sum_{n=0}^{\infty} \left|\langle \vec{\varphi}_{Hn} \mid \vec{\phi} \rangle\right|^2 n < \infty\}. \tag{35.31}$$

The above domain is introduced so that the extended operators, denoted by $\widehat{a}^{\star}_{\vec{\mathcal{D}}}$, and $\widehat{a}_{\vec{\mathcal{D}}}$, acting on[20]

$$\vec{\phi} = \sum_{n=0}^{\infty} c_n \vec{\varphi}_{Hn} \in \vec{\mathcal{D}}, \quad c_n = \langle \vec{\varphi}_{Hn} \mid \vec{\phi} \rangle \tag{35.32}$$

would lead to a new vector with a finite norm, e.g., the square of the norm of the vector

$$\widehat{a}_{\vec{\mathcal{D}}}\vec{\phi} = \sum_{n=0}^{\infty} c_n \widehat{a}_{\vec{\mathcal{D}}}\vec{\varphi}_{Hn} = \sum_{n=1}^{\infty} c_n \sqrt{n}\, \vec{\varphi}_{H(n-1)} \tag{35.33}$$

given by

$$||\widehat{a}_{\vec{\mathcal{D}}}\vec{\phi}\,||^2 = \langle \widehat{a}_{\vec{\mathcal{D}}}\vec{\phi} \mid \widehat{a}_{\vec{\mathcal{D}}}\vec{\phi} \rangle = \sum_{n=1}^{\infty} |c_n|^2 n \tag{35.34}$$

is finite. Following the discussion on Eqs. (17.122) to (17.124) we can conclude that $\widehat{a}_{\vec{\mathcal{D}}}$ and $\widehat{a}^{\star}_{\vec{\mathcal{D}}}$ are the adjoints of each other, e.g., $\widehat{a}^{\star}_{\vec{\mathcal{D}}} = \widehat{a}^{\dagger}_{\vec{\mathcal{D}}}$. It follows that $\widehat{a}_{\vec{\mathcal{D}}}$ and $\widehat{a}^{\dagger}_{\vec{\mathcal{D}}}$ constitute a pair of annihilation and creation operators.

(4) Their associated number operator $\widehat{N}_{\vec{\mathcal{D}}} := \widehat{a}^{\dagger}_{\vec{\mathcal{D}}}\widehat{a}_{\vec{\mathcal{D}}}$ is selfadjoint. The operator admits $\vec{\varphi}_{Hn}$ as eigenvectors corresponding to eigenvalue n and it acts on the domain[21]

$$\vec{\mathcal{D}}(\widehat{N}_{\vec{\mathcal{D}}}) := \{\vec{\phi} \in \vec{L}^2(I\!R) : \sum_{n=0}^{\infty} \left|\langle \vec{\varphi}_{Hn} \mid \vec{\phi} \rangle\right|^2 n^2 < \infty\}. \tag{35.35}$$

This operator is a selfadjoint extension of $\widehat{N}_{\vec{S}_s}$. It follows that

$$\hbar\omega\big(\,\widehat{N}_{\vec{\mathcal{D}}} + 1/2\,\big). \tag{35.36}$$

is a selfadjoint extension of $\widehat{H}_{ho\vec{S}_s}$ in Eq. (35.2). This selfadjoint extension is unique since $\widehat{H}_{ho\vec{S}_s}$ is essentially selfadjoint. The essential selfadjointness of $\widehat{H}_{ho\vec{S}_s}$ can be proved as follows:

---

[19] Jauch p. 45.

[20] Bearing in mind that $\vec{\varphi}_{Hn}$ form an orthonormal basis for $\vec{L}^2(I\!R)$.

[21] This domain is established by the same reasoning which establishes the domain $\vec{\mathcal{D}}$ in Eq. (35.31).

(a) First we note that $\widehat{H}_{ho\vec{S}_s}$ can act on all the vectors $\vec{\varphi}_{Hn}$ since they are in $\vec{S}_s(I\!R)$. Moreover these vectors are eigenvectors of $\widehat{H}_{ho\vec{S}_s}$.

(b) Being eigenvectors they are automatically analytic vectors of $\widehat{H}_{ho\vec{S}_s}$. Since these vectors form a complete set it follows from Theorem 19.5(1) that $\widehat{H}_{ho\vec{S}_s}$ is essentially selfadjoint.

(5) Following P27.2(4) the quantised Hamiltonian for the oscillator is taken to be the unique selfadjoint extension of $\widehat{H}_{ho\vec{S}_s}$ given in Eq. (35.36). This selfadjoint extension can be identified with $\widehat{H}_{ho}$ in Eq. (35.15) since they are both selfadjoint and they have the same eigenvalues and eigenvectors.[22]

In the remainder of this chapter we shall adopt an intuitive discussion using the results and notation of §35.2.1 to avoid excessive mathematical discussion.

35.3 Time Evolution

35.3.1 In the Schrödinger Picture

A classical oscillator would execute periodic motion around the origin. If the particle is placed at rest at a position $x_c(0)$ to the right of the origin at time $t = 0$ it would move to position $x_c(t)$ with momentum $p_c(t)$ at time $t > 0$ given in Eq. (27.13). The evolution of a quantised harmonic oscillator is more complicated. There are two clear cases:

1. Since Hamiltonian $\widehat{H}_{ho}$ possesses a discrete set of eigenvalues $E_n 0$ with corresponding eigenvectors $\vec{\varphi}_{Hn}$ the evolution of an eigenstate takes a simple form in the Schrödinger picture, i.e., if at time $t = 0$ the state vector $\vec{\phi}_{Hn}(0)$ is equal to the eigenvector $\vec{\varphi}_{Hn}$ then, in accordance with Eq. (29.4), the evolved state vector is given by

$$\vec{\phi}_{Hn}(t) = e^{-\frac{i}{\hbar}E_n t}\vec{\varphi}_{Hn}. \tag{35.37}$$

[22] They would have the same spectral decomposition. This shows that the domain in Eq. (35.35) is the same as $\mathcal{D}(\widehat{H}_{ho}) = \mathcal{D}(\widehat{p}^2) \cap \mathcal{D}(\widehat{x}^2)$.

These states are known as *stationary states* since these states generate position probability distribution functions which are time-independent. They also lead to time-independent expectation values for observables which are not explicitly time-dependent

2. If the initial state vector is not an eigenvector of $\widehat{H}_{ho}$, the solution is more complicated. One way to find the evolved state vector is to use Eq. (29.5). Alternatively we can solve the Schrödinger equation directly. As an example consider an initial state at $t = 0$ given in the coordinate representation by the following wave function centered at $x = x_c(0)$:

$$\phi(x, 0) = \left(\frac{\lambda}{\pi}\right)^{\frac{1}{4}} \exp\left(-\frac{1}{2}\lambda\left(x - x_c(0)\right)^2\right), \quad \lambda = \frac{m\omega}{\hbar}. \quad (35.38)$$

The evolved the wave function $\phi(x, t)$ is

$$\phi(x, t) = C(t) \exp\left(\frac{i}{\hbar} p_c(t)x - \frac{1}{2}\lambda\left(x - x_c(t)\right)^2\right), \quad (35.39)$$

where

$$x_c(t) = x_c(0)\cos\omega t, \quad (35.40)$$

$$p_c(t) = -m\omega x_c(0)\sin\omega t, \quad (35.41)$$

$$C(t) = \left(\frac{\lambda}{\pi}\right)^{\frac{1}{4}} \exp\left(-\frac{i}{\hbar}\left(\frac{1}{2}x_c(t)p_c(t) + \frac{1}{2}\hbar\omega t\right)\right). \quad (35.42)$$

In other words, $\phi(x, t)$ is a solution of the Schrödinger equation satisfying the initial condition in Eq. (35.38). The corresponding position probability density function is

$$|\phi(x, t)|^2 = \left(\frac{\lambda}{\pi}\right)^{\frac{1}{2}} \exp\left(-\lambda\left(x - x_c(t)\right)^2\right). \quad (35.43)$$

This position probability density function is time-dependent and at time t it reaches the maximum at $x = x_c(t)$ which can be regarded as the center of the wave packet. We can visualise the motion of the wave function $\phi(x, t)$ as an oscillating wave packet shown in the plot of $|\phi(x, t)|^2$ versus x below.

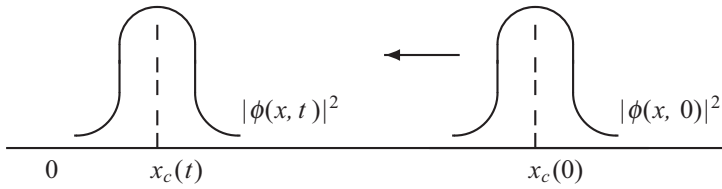


Figure 35.1 Oscillating wave packet.

Figure 35.1 has the following interpretation:

(1) The initial wave packet $\phi(x, 0)$ is centered at $x = x_c(0)$ which is at a maximum distance on the right of the origin.

(2) As t increases the center of the wave packet at $x = x_c(t)$ moves left.

(3) As t increases further the center of the wave packet would move through the origin $x = 0$ to the left until it reaches a maximum distance away from the origin at $x = -x_c(0)$.

(4) As t continues to increase the center of the wave packet would move back to the right until it reaches $x = x_c(0)$.

(5) As t increases still further the center of the wave packet would oscillate cosinusoidally about the origin like a classical harmonic oscillator. This is a rare example of a wave packet oscillating in time without dispersing.

The physical interpretation of the motion of the wave function based on the time dependence of expectation values is as follows:

(1) The position expectation value is given by[23]

$$\mathcal{E}\left(\widehat{x}, \vec{\phi}(t)\right) = \int_{-\infty}^{\infty} \phi(x, t)^* x\, \phi(x, t)\, dx = x_c(t). \tag{35.44}$$

(2) The momentum expectation value is given by

$$\mathcal{E}\left(\widehat{p}, \vec{\phi}(t)\right) = \int_{-\infty}^{\infty} \phi^*(x, t) \left(-i\hbar \frac{d}{dx}\right) \phi(x, t)\, dx = p_c(t). \tag{35.45}$$

[23] Here $\vec{\phi}(t)$ denotes the vector defined by the wave function $\phi(x, t)$ in Eq. (35.39).

(3) The energy expectation value is given by

$$\mathcal{E}\left(\widehat{H}_{ho}, \vec{\phi}(t)\right) = \int_{-\infty}^{\infty} \phi^*(x,t)\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right)\phi(x,t)\,dx$$
$$= \frac{1}{2m}p_c^2(t) + \frac{1}{2}\,m\omega^2\,x_c^2(t) + \frac{1}{2}\,\hbar\omega. \tag{35.46}$$

This energy expectation value which differs from the classical value by an additional term $\frac{1}{2}\,\hbar\omega$ is conserved while the position and momentum are not conserved. This feature can be seen more clearly in the Heisenberg picture.

35.3.2 In the Heisenberg Picture

The Heisenberg picture describes the evolution of the system in terms of the evolution of observables, particularly in terms of the time dependence of position and momentum. This is not a strange way to describe evolution as this is how evolution is described in classical mechanics. Let us examine how the position and momentum operators depend on time:

(1) The Hamiltonian at time t is given in accordance with Eq. (29.31) by[24]

$$\widehat{H}_{hoHei}(t) = \frac{1}{2m}\,\widehat{p}_{Hei}^{\,2}(t) + \frac{1}{2}\,m\omega^2\widehat{x}_{Hei}^{\,2}(t), \tag{35.47}$$

(2) The Heisenberg equations for position and momentum in Eq. (29.32) become[25]

$$\frac{d}{dt}\,\widehat{x}_{Hei}(t) = \frac{1}{m}\,\widehat{p}_{Hei}(t) \quad \text{or} \quad \widehat{p}_{Hei}(t) = m\frac{d}{dt}\,\widehat{x}_{Hei}(t), \tag{35.48}$$
$$\frac{d}{dt}\,\widehat{p}_{Hei}(t) = -m\omega^2\,\widehat{x}_{Hei}(t). \tag{35.49}$$

[24] The subscript *Hei* (already used in Eq. (29.35)) signifies quantities in the Heisenberg picture. This notation aims to avoid confusion with subscript H used elsewhere, e.g., in the vectors in Eq. (35.19).

[25] The commutators $[\,\widehat{x}_{Hei}(t),\, \widehat{H}_{hoHei}(t)\,]$ and $[\,\widehat{p}_{Hei}(t),\, \widehat{H}_{hoHei}(t)\,]$ can be evaluated using Eq. (29.25).

(3) Differentiating Eq. (35.48) leads to

$$\frac{d^2}{dt^2}\,\widehat{x}_{\text{Hei}}(t) = -\omega^2\,\widehat{x}_{\text{Hei}}(t). \tag{35.50}$$

(4) The general solution to the above equation gives us the time dependence of $\widehat{x}_{\text{Hei}}(t)$ and $\widehat{p}_{\text{Hei}}(t)$, i.e., we have

$$\widehat{x}_{\text{Hei}}(t) = \widehat{A}\,\cos\omega t + \widehat{B}\,\sin\omega, \tag{35.51}$$
$$\widehat{p}_{\text{Hei}}(t) = -m\omega\widehat{A}\,\sin\omega t + m\omega\widehat{B}\,\cos\omega t, \tag{35.52}$$

where $\widehat{A}$, $\widehat{B}$ are time-independent.[26] They are determined by initial conditions which are

$$\widehat{x}_{\text{Hei}}(t=0) = \widehat{x} \;\text{ and }\; \widehat{p}_{\text{Hei}}(t=0) = \widehat{p}. \tag{35.53}$$

(a) Initial position at $t = 0$:

$$\widehat{x}_{\text{Hei}}(0) = \widehat{A} \quad\Rightarrow\quad \widehat{A} = \widehat{x}. \tag{35.54}$$

(b) Initial momentum at $t = 0$:

$$\widehat{p}_{\text{Hei}}(0) = \widehat{p} = m\omega\widehat{B} \quad\Rightarrow\quad \widehat{B} = \widehat{p}/m\omega. \tag{35.55}$$

(c) Finally we get

$$\widehat{x}_{\text{Hei}}(t) = (\cos\omega t)\,\widehat{x} + \left(\frac{\sin\omega t}{m\omega}\right)\,\widehat{p}, \tag{35.56}$$
$$\widehat{p}_{\text{Hei}}(t) = (-m\omega\sin\omega t)\,\widehat{x} + (\cos\omega t)\,\widehat{p}. \tag{35.57}$$

(5) We can evaluate the position and momentum expectation values at time t with respect to the state vector $\vec{\phi}_{\text{Hei}}(t) = \vec{\phi}_{\text{Hei}}(0)$ defined by the wave function $\phi(x, 0)$ in Eq. (35.38) to get

$$\mathcal{E}\big(\widehat{x}_{\text{Hei}}(t), \vec{\phi}_{\text{Hei}}(0)\big) = \langle\vec{\phi}_{\text{Hei}}(0) \mid \widehat{x}_{\text{Hei}}(t)\vec{\phi}_{\text{Hei}}(0)\rangle = x_c(t), \tag{35.58}$$
$$\mathcal{E}\big(\widehat{p}_{\text{Hei}}(t), \vec{\phi}_{\text{Hei}}(0)\big) = \langle\vec{\phi}_{\text{Hei}}(0) \mid \widehat{p}_{\text{Hei}}(t)\vec{\phi}_{\text{Hei}}(0)\rangle = p_c(t). \tag{35.59}$$

These are the same as those calculated in the Schrödinger picture in Eqs. (35.44) and (35.45).

[26] These are time-independent operators, not numerical constants.

(6) One can explicitly check that the Hamiltonian is time-independent, despite being a function of the time-dependent position and momentum operators, i.e.,

$$\widehat{H}_{hoHei}(t) = \widehat{H}_{ho}(0) = \frac{1}{2m}\,\widehat{p}^{\,2} + \frac{1}{2}\,m\omega^2\,\widehat{x}^{\,2}. \tag{35.60}$$

It follows that the energy expectation value is conserved.

Equations (35.56) and (35.57) do not appear to have an intuitive meaning. The difficulty is to visualise the wave packet remaining stationary while operators are "moving". To avoid getting into this kind of difficulties one should not try to look at Eqs. (35.56) and (35.57) in isolation. By themselves neither the wave function nor the operators are physical quantities. It is expectation values which are the directly measurable physical quantities. The meaning of Eqs. (35.56) and (35.57) should be understood in terms of their expectation values in Eqs. (35.58) and (35.59). It is not meaningful to keep asking whether it should be the wave function or the operators which are "moving".

35.3.3 In the Interaction Picture

A harmonic oscillator with charge q is placed in a uniform and static external electric field of strength E pointing in the $+x$ direction. With the electrical potential energy taken to be zero at the origin the classical Hamiltonian of the system is

$$H = \frac{1}{2m}\,p^2 + \frac{1}{2}\,m\omega^2 x^2 - qEx. \tag{35.61}$$

This is an interacting system in that the harmonic oscillator is interacting with an external electric field. The Hamiltonian can be written as a sum of a free part H_0 and an interacting part H', where

$$H_0 = \frac{1}{2m}\,p^2 + \frac{1}{2}\,m\omega^2\,x^2, \quad H' = -qEx. \tag{35.62}$$

The quantised Hamiltonian $\widehat{H}$ at time $t = 0$ can be similarly written as a sum of a free part $\widehat{H}_0$ and an interacting part $\widehat{H}'$ where

$$\widehat{H}_0 = \frac{1}{2m}\,\widehat{p}^{\,2} + \frac{1}{2}\,m\omega^2\,\widehat{x}^{\,2}, \quad \widehat{H}' = -qE\,\widehat{x}. \tag{35.63}$$

In the Interaction picture we have to consider the evolution of operators and state separately[27]:

1. Operators evolve like that of the corresponding operators for the free system in the Heisenberg picture, i.e., like that of an unperturbed oscillator in the Heisenberg picture:

(1) For the position and momentum operators we have, from Eqs. (35.56) and (35.57),

$$\widehat{x}_I(t) = (\cos\omega t)\,\widehat{x} + \left(\frac{\sin\omega t}{m\omega}\right)\,\widehat{p}, \tag{35.64}$$

$$\widehat{p}_I(t) = (-m\omega\sin\omega t)\,\widehat{x} + (\cos\omega t)\,\widehat{p}. \tag{35.65}$$

(2) The interaction Hamiltonian is then given by

$$\widehat{H}'_I(t) = -qE\left((\cos\omega t)\,\widehat{x} + \left(\frac{\sin\omega t}{m\omega}\right)\,\widehat{p}\right). \tag{35.66}$$

2. The wave function $\phi_I(x, t)$ corresponding to state vector $\vec{\phi}_I(t)$ satisfies the following equation:

$$i\hbar\frac{\partial}{\partial t}\phi_I(x, t) = -qE\left((\cos\omega t)x + \frac{\sin\omega t}{m\omega}\left(-i\hbar\frac{d}{dx}\right)\right)\phi_I(x, t). \tag{35.67}$$

The solution for a given initial wave function $\phi_I(x, 0)$ is easily verified to be

$$\phi_I(x, t) = \exp\left\{\frac{qE}{i\hbar}\left(-\frac{\sin\omega t}{\omega}x + \left[\frac{\cos\omega t}{m\omega^2} + A\right]\left(-i\hbar\frac{d}{dx}\right)\right)\right\}\phi_I(x, 0).$$

where A is a time-independent constant which arises from the initial condition. Letting $t = 0$ the above equation becomes

$$\phi_I(x, 0) = \exp\left\{\frac{qE}{i\hbar}\left[\frac{1}{m\omega^2} + A\right]\left(-i\hbar\frac{d}{dx}\right)\right\}\phi_I(x, 0) \tag{35.68}$$

$$\Rightarrow \quad A = -\frac{1}{m\omega^2}. \tag{35.69}$$

[27] We have employed the subscript I to signify quantities in the Interaction picture. As before we are working in the coordinate representation.

This results in the following solution:

$$\phi_I(x,t) = \exp\left\{\frac{qE}{i\hbar}\left(-\frac{\sin\omega t}{\omega}x + \left[\frac{\cos\omega t}{m\omega^2} - \frac{1}{m\omega^2}\right]\left(-i\hbar\frac{d}{dx}\right)\right)\right\}\phi_I(x,0). \tag{35.70}$$

Looking at the three pictures we can say that

(1) The Schrödinger picture is easier to visualise in terms of a propagating wave function and convenient for many practical calculations.

(2) Heisenberg picture with its Heisenberg equation similar to the Hamilton's equation is useful in the study of transitions between classical and quantum mechanics, in many-body systems and quantum field theory, and in general theoretical analysis.

(3) Interaction picture is useful for interacting systems.

35.4 Isotropic Oscillators

A simple harmonic oscillator is a particle in one-dimensional motion under a harmonic force. We have similar systems in two and three dimensions.

Consider a classical particle of mass m in two-dimensional motion on the x-y plane under a central force. A central force is one which is derivable from a potential which is a function of the radial distance r from the origin. The Hamiltonian is of the form

$$H = \frac{1}{2m}\left(p_x^2 + p_y^2\right) + V(r). \tag{35.71}$$

The particle is a two-dimensional isotropic oscillator if the potential energy is proportional to r^2, i.e., the Hamiltonian is

$$\begin{aligned} H_{2io} &:= \frac{1}{2m}\left(p_x^2 + p_y^2\right) + \frac{1}{2}m\omega^2 r^2 \\ &= \frac{1}{2m}\left(p_x^2 + p_y^2\right) + \frac{1}{2}m\omega^2\left(x^2 + y^2\right). \end{aligned} \tag{35.72}$$

This Hamiltonian can be rewritten as

$$H_{2iso} = H_{ho,x} + H_{ho,y}, \tag{35.73}$$

where $H_{ho,x}$ is the Hamiltonian of a harmonic oscillator along the x-axis and $H_{ho,y}$ is one along the y-axis.

This isotropic oscillator can be quantised as follows:

1. Take the state space to be the Hilbert space $\vec{L}^2(I\!R^2, dxdy)$.[28]

2. The position x and momentum p_x can be quantised, respectively, as the operator $\widehat{x}(I\!R^2)$ defined in Eq. (17.17) and the operator $\widehat{p}_x(I\!R^2)$ defined by Eq. (17.52). We can similarly quantise y and momentum p_y as operators $\widehat{y}(I\!R^2)$ and $\widehat{p}_y(I\!R^2)$.

3. The classical $H_{ho,x}$ and $H_{ho,y}$ are quantised as

$$\widehat{H}_{sho,x} = \frac{1}{2m}\left(\widehat{p}_x(I\!R^2)\right)^2 + \frac{1}{2}\, m\omega\big(\widehat{x}(I\!R^2)\big)^2. \tag{35.74}$$

$$\widehat{H}_{sho,y} = \frac{1}{2m}\left(\widehat{p}_y(I\!R^2)\right)^2 + \frac{1}{2}\, m\omega\big(\widehat{y}(I\!R^2)\big)^2. \tag{35.75}$$

4. The Hamiltonian of the isotropic oscillator is

$$\widehat{H}_{2io} = \widehat{H}_{sho,x} + \widehat{H}_{sho,y}. \tag{35.76}$$

Since $\vec{L}^2(I\!R^2, dxdy) = \vec{L}^2(I\!R, dx) \otimes \vec{L}^2(I\!R, dy)$, as shown in Eq. (24.60), we can establish the operators in $\vec{L}^2(I\!R^2, dxdy)$ in terms of tensor products of operators defined separately on $\vec{L}^2(I\!R, dx)$ and $\vec{L}^2(I\!R, dy)$:

(1) The position and momentum operators $\widehat{x}(I\!R^2)$ and $\widehat{p}_x(I\!R^2)$ are related to $\widehat{x}$ and $\widehat{p}$ in $\vec{L}^2(I\!R, dx)$ by Eq. (24.65), i.e.,

$$\widehat{x}(I\!R^2) = \widehat{x}(I\!R) \otimes \widehat{I\!I}_y, \quad \widehat{p}_x(I\!R^2) = \widehat{p}(I\!R) \otimes \widehat{I\!I}_y \tag{35.77}$$

We shall rewrite $\widehat{p}$ in $\vec{L}^2(I\!R, dx)$ as $\widehat{p}_x$ in what follows so as to distinguish corresponding operator in $\vec{L}^2(I\!R, dy)$ which will be denoted by $\widehat{p}_y$.

(2) Following Eq. (35.9) we can introduce a pair of annihilation and creation operators in $\vec{L}^2(I\!R, dx)$

$$\widehat{a}_x := \frac{1}{\sqrt{2}}\,\lambda\left(\widehat{x} + \frac{i}{m\omega}\,\widehat{p}_x\right), \tag{35.78}$$

$$\widehat{a}_x^\dagger := \frac{1}{\sqrt{2}}\,\lambda\left(\widehat{x} - \frac{i}{m\omega}\,\widehat{p}_x\right). \tag{35.79}$$

[28] *See* §24.2.2 for notation.

The associated number operator is $\widehat{N}_x = \widehat{a}_x^\dagger \widehat{a}_x$. The corresponding operator $\widehat{N}_x^\otimes$ in $\vec{L}^2(I\!R^2, dxdy)$ is given by Eq. (24.69), i.e.,

$$\widehat{N}_x^\otimes := \widehat{N}_x \otimes \widehat{I\!I}_y. \tag{35.80}$$

The operator $\widehat{H}_{sho,x}$ in Eq. (35.76) is related to $\widehat{N}_x$ by

$$\widehat{H}_{sho,x} = \hbar\omega\Big(\widehat{N}_x + \frac{1}{2}\Big) \otimes \widehat{I\!I}_y. \tag{35.81}$$

(3) We can repeat the procedure to define a pair of annihilation and creation operators $\widehat{a}_y$, $\widehat{a}_y^\dagger$ and the associated number operator $\widehat{N}_y = \widehat{a}_y^\dagger \widehat{a}_y$ in $\vec{L}^2(I\!R, dy)$. The corresponding number operator in $\vec{L}^2(I\!R^2, dxdy)$ is

$$\widehat{N}_y^\otimes := \widehat{I\!I}_x \otimes \widehat{N}_y. \tag{35.82}$$

(4) The Hamiltonian operator in Eq. (35.76) is the operator defined in terms of tensor products in Eq. (24.71). This is effectively the sum of two harmonic oscillator Hamiltonians, one along the x-axis and the other one along the y-axis.

We can solve the eigenvalue problem of this Hamiltonian with the help of the annihilation and creation operators $\widehat{a}_x$, $\widehat{a}_x^\dagger$, $\widehat{a}_y$, $\widehat{a}_y^\dagger$ introduced above. Each pair of these operators, e.g., $\widehat{a}_x$ and $\widehat{a}_x^\dagger$, possesses properties P27.9(1), P27.9(2) and P27.9(3). Since the set of operators $\widehat{x}$, $\widehat{p}_x$, $\widehat{y}$, $\widehat{p}_y$ form an irreducible set in $\vec{L}^2(I\!R^2, dxdy)$ the two pairs of operators $\widehat{a}_x$, $\widehat{a}_x^\dagger$, $\widehat{a}_y$, $\widehat{a}_y^\dagger$ also form an irreducible set in $\vec{L}^2(I\!R^2, dxdy)$. Each pair alone does not form an irreducible set in $\vec{L}^2(I\!R^2, dxdy)$.

The number operators $\widehat{N}_x$ and $\widehat{N}_y$ would have the same set of eigenvalues as the number operator $\widehat{N}$ in Eq. (27.123), i.e., 0, 1, 2, ... Let us list some of the properties of these number operators:

(1) Let $\vec{\varphi}_{xn}$ and $\vec{\varphi}_{ym}$ be the eigenvectors of $\widehat{N}_x$ and $\widehat{N}_y$ in $\vec{L}^2(I\!R, dx)$ and $\vec{L}^2(I\!R, dy)$, respectively. Then

$$\vec{\varphi}_{n,m}^{\,\otimes} := \vec{\varphi}_{xn} \otimes \vec{\varphi}_{ym}, \quad n, m = 0, 1, 2, \ldots, \tag{35.83}$$

are the common eigenvectors of $\widehat{N}_x^\otimes$ and $\widehat{N}_y^\otimes$, i.e., we have

$$\widehat{N}_x^\otimes \vec{\varphi}_{n,m}^{\,\otimes} = n\,\vec{\varphi}_{n,m}^{\,\otimes}, \quad \widehat{N}_y^\otimes \vec{\varphi}_{n,m}^{\,\otimes} = m\,\vec{\varphi}_{n,m}^{\,\otimes}. \tag{35.84}$$

(2) The eigenvalues of the number operators $\widehat{N}_x^{\otimes}$ and $\widehat{N}_y^{\otimes}$ are degenerate, a result predictable by the reducible nature of each pair of the annihilation and creation operators. Since the two pairs of annihilation and creation operators together form an irreducible set the eigenvectors $\vec{\varphi}_{n,m}^{\otimes}$ constitute a complete orthonormal set for $\vec{L}^2(I\!R^2, dxdy)$ so that an arbitrary vector in $\vec{L}^2(I\!R^2, dxdy)$ is expressible as a linear combination of $\vec{\varphi}_{n,m}^{\otimes}$, i.e.,

$$\vec{\phi}^{\otimes} = \sum_{n,m} c_{nm}\, \vec{\varphi}_{n,m}^{\otimes}. \tag{35.85}$$

(3) The eigenvalues of the Hamiltonian are

$$\widehat{H}_{2iso}\, \vec{\varphi}_{n,m}^{\otimes} = E_{n,m}\, \vec{\varphi}_{n,m}^{\otimes}, \quad E_{n,m} = \big(n + m + 1\big)\hbar\omega. \tag{35.86}$$

The three-dimensional isotropic oscillator can be similarly treated.

In practical applications it is useful to express the eigenvectors $\vec{\varphi}_{xn}$ and $\vec{\varphi}_{ym}$ in terms of their corresponding eigenfunctions $\varphi_n(x)$ and $\varphi_m(y)$. The tensor product of $\vec{\varphi}_{xn}$ and $\vec{\varphi}_{ym}$ becomes the product $\varphi_n(x)\varphi_m(y)$ which are functions of variables x and y.

Exercises and Problems

Q35(1) Verify the results in Eqs. (35.5) to (35.8).

Q35(2) Verify Eq. (35.10).

Q35(3) Verify that $\varphi_0(x)$ in Eq. (35.13) satisfies Eq. (35.12).

Q35(4) Obtain the first and the second excited state eigenvectors from the expression in Eq. (35.19) in terms of $\varphi_{H0}(x)$.

Q35(5) Verify Eq. (35.14) and (35.15).

Q35(6) Verify that $\phi(x, t)$ in Eq. (35.39) is normalised and that it also satisfies the Schrödinger equation for time evolution of the harmonic oscillator.

Q35(7) Verify Eqs. (35.44), (35.45).

Q35(8) Verify Eq. (35.46).

Q35(9) Using the expressions in Eqs. (35.40) and (35.41) for $x_c(t)$ and $p_c(t)$ show that energy expectation value

in Eq. (35.46) is conserved, i.e., the value is time-independent. Explain how the expectation value can approximate the energy of a classical harmonic oscillator.

Q35(10) Verify Eqs. (35.58) and (35.59).

Q35(11) Verify Eq. (35.60) by explicit calculation of $\widehat{H}_{hoHei}(t)$ using $\widehat{x}_{Hei}(t)$ and $\widehat{p}_{Hei}(t)$ in Eqs. (35.56) and (35.57).

Q35(12) The Hamiltonian of a forced harmonic oscillator in the Schrödinger picture is given in the usual notation by[29]

$$\widehat{H}_{Sch} = \frac{1}{2m}\,\widehat{p}^{\,2}_{Sch} + \frac{1}{2}\,m\omega^2\,\widehat{x}^{\,2}_{Sch} - g\left(\widehat{x}_{Sch} + \frac{1}{m\omega}\,\widehat{p}_{Sch}\right), \quad (35.87)$$

where g is a real number. Let $\widehat{a}_{Sch}$ and $\widehat{a}^{\dagger}_{Sch}$ be a pair of operators related to $\widehat{x}_{Sch}$ and $\widehat{p}_{Sch}$ by Eq. (35.9).

(a) Show that

$$\widehat{H}_{Sch} = \hbar\omega\left(\widehat{a}^{\dagger}_{Sch}\,\widehat{a}_{Sch} + \frac{1}{2}\right) + \gamma\,\widehat{a}_{Sch} + \gamma^*\,\widehat{a}^{\dagger}_{Sch}, \quad (35.88)$$

where

$$\gamma = -\sqrt{\frac{\hbar}{2m\omega}}\,(1-i)g.$$

(b) In the Heisenberg picture the annihilation and creation operators become $\widehat{a}_{Hei}(t)$ and $\widehat{a}^{\dagger}_{Hei}(t)$ and the Hamiltonian becomes

$$\widehat{H}_{Hei}(t) = \hbar\omega\left(\widehat{a}^{\dagger}_{Hei}(t)\widehat{a}_{Hei}(t) + \frac{1}{2}\right) + \gamma\,\widehat{a}_{Hei}(t) + \gamma^*\,\widehat{a}^{\dagger}_{Hei}(t).$$

Show that $\widehat{a}_{Hei}(t)$ satisfies the Heisenberg equation of motion and that the Heisenberg equation can be wrtten as

$$\frac{d}{dt}\widehat{a}_{Hei}(t) = -i\omega\widehat{a}_{Hei}(t) - \frac{i}{\hbar}\gamma^*.$$

Show further that this equation can be rewritten in the form

$$\frac{d}{dt}\left(\widehat{a}_{Hei}(t)e^{i\omega t}\right) = -\frac{i}{\hbar}\gamma^*\,e^{i\omega t}.$$

Integrate this equation to obtain an explicit expression for the time dependence of $\widehat{a}_{Hei}(t)$.

Q35(13) What is the degeneracy of the eigenvalues $E_{1,1}$, $E_{2,2}$ of the two-dimensional isotropic oscillator? Are all eigenvalues degenerate?

[29] Merzbacher pp. 335–336.

Chapter 36

Angular Momenta

36.1 Orbital Angular Momentum

36.1.1 Orbital Angular Momentum Operators

The quantised angular momentum operators are obtained in §27.4. These operators do not commute. They satisfy the commutation relations in Eqs. (27.111) to (27.113). It follows that

(1) $\widehat{L}_{cx}$, $\widehat{L}_{cy}$ and $\widehat{L}_{cz}$ are essentially incompatible. They do not possess a complete orthonormal set of common eigenvectors. However, there is a state, the state of zero orbital angular momentum values, in which they do possess a definite value, i.e., the value 0.[1]

(2) $\widehat{L}_{cz}$, $\widehat{L}_c^2$ are compatible and they possess common eigenvectors and are simultaneously measurable. This is also true for $\widehat{L}_{cx}$, $\widehat{L}_c^2$ and for $\widehat{L}_{cy}$, $\widehat{L}_c^2$. However, the eigenvectors of $\widehat{L}_{cz}$, $\widehat{L}_c^2$ are different from that of $\widehat{L}_{cx}$, $\widehat{L}_c^2$ or that of $\widehat{L}_{cy}$, $\widehat{L}_c^2$.

Orbital angular momentum is relevant to rotational motion for which spherical coordinates are more convenient to use. The

[1] *See* Definition 28.4.1(1). The state corresponds to $\ell = 0$ in Eq. (36.30).

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

Cartesian coordinates x, y, z are related to spherical coordinates r, θ, φ by Eq. (16.41), or conversely by

$$r^2 = x^2 + y^2 + z^2, \quad \cos\theta = z/r, \quad \tan\varphi = y/x. \tag{36.1}$$

The derivatives with respect to the Cartesian and spherical coordinates are related by[2]

$$\frac{\partial}{\partial x} = \sin\theta\cos\varphi\,\frac{\partial}{\partial r} + \frac{1}{r}\cos\theta\cos\varphi\,\frac{\partial}{\partial\theta} - \frac{1}{r}\,\frac{\sin\varphi}{\sin\theta}\,\frac{\partial}{\partial\varphi}, \tag{36.2}$$

$$\frac{\partial}{\partial y} = \sin\theta\sin\varphi\,\frac{\partial}{\partial r} + \frac{1}{r}\cos\theta\sin\varphi\,\frac{\partial}{\partial\theta} + \frac{1}{r}\,\frac{\cos\varphi}{\sin\theta}\,\frac{\partial}{\partial\varphi}, \tag{36.3}$$

$$\frac{\partial}{\partial z} = \cos\theta\,\frac{\partial}{\partial r} - \frac{1}{r}\sin\theta\,\frac{\partial}{\partial\theta}. \tag{36.4}$$

Using these we can arrive at the expressions for all the operators relevant to orbital angular momentum in spherical coordinates:

$$\widehat{L}_{cx} = i\hbar\left(\sin\varphi\,\frac{\partial}{\partial\theta} + \cot\theta\cos\varphi\,\frac{\partial}{\partial\varphi}\right), \tag{36.5}$$

$$\widehat{L}_{cy} = -i\hbar\left(\cos\varphi\,\frac{\partial}{\partial\theta} - \cot\theta\sin\varphi\,\frac{\partial}{\partial\varphi}\right), \tag{36.6}$$

$$\widehat{L}_{cz} = -i\hbar\,\frac{\partial}{\partial\varphi}, \tag{36.7}$$

$$\widehat{L}_c^2 = -\hbar^2\left(\frac{1}{\sin\theta}\,\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\,\frac{\partial^2}{\partial\varphi^2}\right). \tag{36.8}$$

These differential expressions do not involve the radial variable r. Hence, they can be used to define four selfadjoint operators

$$\widehat{L}_x(\mathcal{S}_u),\ \widehat{L}_y(\mathcal{S}_u),\ \widehat{L}_z(\mathcal{S}_u),\ \widehat{L}^2(\mathcal{S}_u)$$

to act in the space $\vec{L}^2(\mathcal{S}_u)$.[3] The pair $\widehat{L}_z(\mathcal{S}_u)$, $\widehat{L}^2(\mathcal{S}_u)$ are given explicitly by Eqs. (17.42) and (19.51). As shown in Eqs. (24.72) and (24.73), operators $\widehat{L}_z(\mathcal{S}_u)$ and $\widehat{L}^2(\mathcal{S}_u)$ in $\vec{L}^2(\mathcal{S}_u)$ can be extended to $\vec{L}^2(I\!R^3)$. These extensions can be identified with the quantised operators $\widehat{L}_{cz}$ and $\widehat{L}_c^2$. Similar results apply to $\widehat{L}_x(\mathcal{S}_u)$, $\widehat{L}_y(\mathcal{S}_u)$.[4] It

[2] Zettili pp. 633–635.
[3] $\widehat{L}^2(\mathcal{S}_u)$ is an operator acting on the vector space $\vec{L}^2(\mathcal{S}_u)$ given in §16.1.2.9.
[4] Amrein, Jauch and Sinha pp. 458–459. Blank, Exner and Havlíček p. 395.

follows that $\widehat{L}_x(\mathcal{S}_u)$, $\widehat{L}_y(\mathcal{S}_u)$, $\widehat{L}_z(\mathcal{S}_u)$ and $\widehat{L}^2(\mathcal{S}_u)$ satisfy the angular momentum commutation relations in Eqs. (27.111) to (27.114). These commutation relations enable us to obtain quite a lot of information on the eigenvalues and eigenvectors of these operators, as shown in the next subsection. This information is also applicable to the quantised angular momentum operators.[5]

36.1.2 Eigenvalues and Eigenfunctions

As pointed out in Eqs. (17.128) and (19.51), $\widehat{L}_z(\mathcal{S}_u)$ and $\widehat{L}^2(\mathcal{S}_u)$ admit $\vec{Y}_{\ell,m_\ell}$ as their common eigenvectors. As before let the corresponding eigenvalues be denoted by $m_\ell\hbar$ and $\ell(\ell+1)\hbar^2$, i.e.,[6]

$$\widehat{L}_z(\mathcal{S}_u)\,\vec{Y}_{\ell,m_\ell} = m_\ell\hbar\,\vec{Y}_{\ell,m_\ell}, \tag{36.9}$$

$$\widehat{L}^2(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} = \ell(\ell+1)\hbar^2\,\vec{Y}_{\ell,m_\ell}, \tag{36.10}$$

where m and ℓ are dimensionless real numbers on account of the selfadjointness of the operators.[7] Introduce two new operators

$$\widehat{L}_+(\mathcal{S}_u) = \widehat{L}_x(\mathcal{S}_u) + i\,\widehat{L}_y(\mathcal{S}_u) = \hbar e^{i\varphi}\left(\frac{\partial}{\partial\theta} + i\cot\theta\,\frac{\partial}{\partial\varphi}\right), \tag{36.11}$$

$$\widehat{L}_-(\mathcal{S}_u) = \widehat{L}_x(\mathcal{S}_u) - i\,\widehat{L}_y(\mathcal{S}_u) = -\hbar e^{-i\varphi}\left(\frac{\partial}{\partial\theta} - i\cot\theta\,\frac{\partial}{\partial\varphi}\right). \tag{36.12}$$

These new operators satisfy the following properties[8]:

$$\widehat{L}_+(\mathcal{S}_u)\widehat{L}_-(\mathcal{S}_u) = \widehat{L}^2(\mathcal{S}_u) - \widehat{L}_z^2(\mathcal{S}_u) + \hbar\widehat{L}_z(\mathcal{S}_u), \tag{36.13}$$

$$\widehat{L}_-(\mathcal{S}_u)\widehat{L}_+(\mathcal{S}_u) = \widehat{L}^2(\mathcal{S}_u) - \widehat{L}_z^2(\mathcal{S}_u) - \hbar\widehat{L}_z(\mathcal{S}_u), \tag{36.14}$$

$$[\widehat{L}^2(\mathcal{S}_u),\,\widehat{L}_\pm(\mathcal{S}_u)\,] = \widehat{0}, \tag{36.15}$$

$$[\,\widehat{L}_z(\mathcal{S}_u),\,\widehat{L}_\pm(\mathcal{S}_u)\,] = \pm\hbar\,\widehat{L}_\pm(\mathcal{S}_u), \tag{36.16}$$

$$\widehat{L}_z(\mathcal{S}_u)\widehat{L}_\pm(\mathcal{S}_u) = \widehat{L}_\pm(\mathcal{S}_u)\widehat{L}_z(\mathcal{S}_u) \pm \hbar\widehat{L}_\pm(\mathcal{S}_u). \tag{36.17}$$

[5] Merzbacher pp. 238–255. Zettili pp. 272–273. Gasiorowicz pp. 121–124.

[6] $\vec{Y}_{\ell,m_\ell}$ are normalised vectors in $\vec{L}^2(\mathcal{S}_u)$ defined by the spherical harmonics $Y_{\ell,m_\ell}(\theta,\varphi)$ which do not involve radial variable r.

[7] No assumption on the values of ℓ and m_ℓ are made here.

[8] Eqs. (36.13) to (36.17) are derived using Eqs. (27.111) to (27.113).

It follows that

$$\begin{aligned}\widehat{L}^2(\mathcal{S}_u)\widehat{L}_\pm(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} &= \widehat{L}_\pm(\mathcal{S}_u)\widehat{L}^2(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}\\ &= \ell(\ell+1)\hbar^2\,\widehat{L}_\pm(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}, \qquad (36.18)\\ \widehat{L}_z(\mathcal{S}_u)\widehat{L}_\pm(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} &= \big(\widehat{L}_\pm(\mathcal{S}_u)\widehat{L}_z(\mathcal{S}_u) \pm \hbar\widehat{L}_\pm(\mathcal{S}_u)\big)\vec{Y}_{\ell,m_\ell}\\ &= (m_\ell \pm 1)\,\hbar\,\widehat{L}_\pm(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}. \qquad (36.19)\end{aligned}$$

These results mean that $\widehat{L}_\pm(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}$ are again eigenvectors of $\widehat{L}^2(\mathcal{S}_u)$ and $\widehat{L}_z(\mathcal{S}_u)$, i.e., $\widehat{L}_\pm(\mathcal{S}_u)$ acting on $\vec{Y}_{\ell,m_\ell}$ would raise or lower the eigenvalues of $\widehat{L}_z(\mathcal{S}_u)$ by $\hbar$ while leaving the eigenvalue of $\widehat{L}(\mathcal{S}_u)^2$ unchanged. These eigenvectors are not normalised. They are related to the normalised eigenvectors $\vec{Y}_{\ell,m_\ell\pm1}$ by[9]

$$\widehat{L}_+(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} = \hbar\sqrt{(\ell-m_\ell)(\ell+m_\ell+1)}\;\vec{Y}_{\ell,m_\ell+1}, \qquad (36.20)$$

$$\widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} = \hbar\sqrt{(\ell+m_\ell)(\ell-m_\ell+1)}\;\vec{Y}_{\ell,m_\ell-1}. \qquad (36.21)$$

$$\Rightarrow\quad \widehat{L}_+(\mathcal{S}_u)\vec{Y}_{\ell,\ell} = \vec{0}, \qquad \widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,-\ell} = \vec{0}, \qquad (36.22)$$

$$\vec{Y}_{\ell,-\ell+1} = \frac{1}{\hbar\sqrt{2\ell}}\,\widehat{L}_+(\mathcal{S}_u)\vec{Y}_{\ell,-\ell}. \qquad (36.23)$$

Possible values of ℓ and m_ℓ and their relationship can be found:

(1) The eigenvalues of $\widehat{L}^2(\mathcal{S}_u)$, being the sum of the squares of three selfadjoint operators are non-negative. i.e., $\ell(\ell+1) \geq 0$. Moreover we can choose ℓ to be non-negative. If we choose ℓ to be negative, then $\ell+1$ would have to be negative. Then $\ell' = -(\ell+1)$ would be non-negative with $\ell'(\ell'+1) = \ell(\ell+1)$.

(2) Acting on the eigenvectors the operators $\widehat{L}_\pm(\mathcal{S}_u)$ satisfy[10]

$$\begin{aligned}\langle\,\vec{Y}_{\ell,m_\ell}\mid \widehat{L}_+(\mathcal{S}_u)\widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}\,\rangle &= \langle\,\widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}\mid \widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}\,\rangle\\ &= ||\,\widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}\,||^2 \geq 0. \qquad (36.24)\end{aligned}$$

By Eq. (36.13) the left-hand side of Eq. (36.24) becomes

$$\begin{aligned}&\langle\vec{Y}_{\ell,m_\ell}\mid \big(\widehat{L}^2(\mathcal{S}_u) - \widehat{L}_z^2(\mathcal{S}_u) + \hbar\widehat{L}_z(\mathcal{S}_u)\big)\vec{Y}_{\ell,m_\ell}\,\rangle\\ &= \hbar^2\big(\ell(\ell+1) - m_\ell^2 + m_\ell\big) \geq 0\\ &\Rightarrow \qquad \ell^2+\ell \geq m_\ell^2 - m_\ell. \qquad (36.25)\end{aligned}$$

[9] *See* Gasiorowicz pp. 122–124 for normalisation.

[10] Use the fact that $\big(\widehat{L}_\pm(\mathcal{S}_u)\big)^\dagger \supset \widehat{L}_\mp(\mathcal{S}_u)$ on account of Eq. (17.101) in P17.8(6). The adjoint of $\widehat{L}_+(\mathcal{S}_u)$ acts like $\widehat{L}_-(\mathcal{S}_u)$, i.e., $\widehat{L}^\dagger_+(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell} = \widehat{L}_-(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}$.

Similarly we have

$$\langle \vec{Y}_{\ell,m_\ell} \mid \widehat{L}_-(\mathcal{S}_u)\widehat{L}_+(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}\rangle = ||\widehat{L}_+(\mathcal{S}_u)\vec{Y}_{\ell,m_\ell}||^2 \geq 0, \quad (36.26)$$

and using Eq. (36.14) we get

$$\ell^2 + \ell \geq m_\ell^2 + m_\ell. \quad (36.27)$$

Since $\ell \geq 0$ we can conclude from Eqs. (36.25) and (36.27) that $\ell^2 \geq m_\ell^2$. The assumption $\ell^2 < m_\ell^2$ would contradict Eqs. (36.25) and (36.27).[11] In other words, ℓ and m_ℓ are related by

$$m_\ell^2 \leq \ell^2 \quad \text{or equivalently} \quad -\ell \leq m_\ell \leq \ell. \quad (36.28)$$

(3) The above results show that m_ℓ is bounded from above and from below. The minimum value of m_ℓ must then be $-\ell$ corresponding to eigenvectors $\vec{Y}_{\ell,-\ell}$ so that we cannot generate new eigenvector corresponding to a lower eigenvalue of m_ℓ, i.e., lower than $-\ell$ by using Eq. (36.21). Similarly the maximum m_ℓ is equal ℓ.

(4) On account of Eq. (36.20) the eigenvalue m_ℓ can increase by 1 in $2\ell + 1$ steps from its minimum value $-\ell$ to its maximum value ℓ, i.e.,

$$m_\ell = -\ell, -\ell + 1, -\ell + 2, \ldots, \ell - 1, \ell. \quad (36.29)$$

We can conclude that $2\ell + 1$ must an integer, i.e., ℓ can only be integers or half integers. There are no other values of m_ℓ.[12]

(5) Our final conclusion here is that the commutation relations between the angular momentum operators imply both integer and half integer values for ℓ and m_ℓ, i.e.,

$$\ell = \{0, +\text{half integers}, +\text{integers}\}, \quad (36.30)$$

$$m_\ell = \{0, \pm\text{half integers}, \pm\text{integers}\}, \quad (36.31)$$

$$-\ell \leq m_\ell \leq \ell. \quad (36.32)$$

[11] Adding Eqs. (36.25) and (36.27) we get $\ell^2 + \ell \geq m^2$, i.e., we cannot have $\ell^2 < m^2$ since $\ell \geq 0$.

[12] This is to avoid creating a value of m_ℓ higher than ℓ by Eq. (36.20). An assumption that m_ℓ has other values, e.g., when $\ell = 2$ the value of m_ℓ is 1.3, would contradict Eq. (36.28) since Eq. (36.20) can then be used to produce $\vec{Y}_{\ell=2,m_\ell=2.3}$ with a value of $m_\ell = 2.3$ which is higher than $\ell = 2$.

Further properties of the orbital angular momentum operators embodied in the explicit expressions for the operators show that half integers values should be excluded. We call the resulting ℓ and m_ℓ the **orbital angular momentum quantum number** and the **orbital magnetic quantum number**, respectively.[13]

Let us now examine why half integer values must be excluded. In spherical coordinates the differential expression the eigenvalue equation of $\widehat{L}_z(\mathcal{S})$ can be written as

$$-i\hbar \frac{\partial}{\partial \varphi} Y_{\ell,m_\ell}(\theta, \varphi) = m_\ell \hbar\, Y_{\ell,m_\ell}(\theta, \varphi). \tag{36.33}$$

Since φ and $\varphi + 2\pi$ refer to the same spatial position we would intuitively expect the functions in the domain of the operator of $\widehat{L}_z(\mathcal{S}_u)$ to satisfy the periodic boundary condition. Mathematically such condition renders $\widehat{L}_z(\mathcal{S}_u)$ selfadjoint.[14] The eigenfunctions and eigenvalues of $\widehat{L}_z(\mathcal{S}_u)$ are similar to those for $\widehat{p}(\mathcal{C}_a)$ given by Eqs. (19.36) and (19.37). It is instructive to find these quantities from Eq. (36.33).

Let us start with lowest magnetic quantum number for a given angular momentum quantum number. We can verify easily that the solution $Y_{\ell,-\ell}(\theta, \varphi)$ of Eq. (36.33) for $m_\ell = -\ell$ is

$$Y_{\ell,-\ell}(\theta, \varphi) = c f(\theta)\, e^{-i\ell\varphi}, \tag{36.34}$$

where c is a normalisation constant. Applying the periodic boundary condition $Y_{\ell,-\ell}(\theta, 0) = Y_{\ell,-\ell}(\theta, 2\pi)$ we get

$$Y_{\ell,-\ell}(\theta, \varphi) = Y_{\ell,-\ell}(\theta, \varphi + 2\pi) \quad \Rightarrow \quad 1 = e^{-i2\pi\ell} \tag{36.35}$$

$$\Rightarrow\ 1 = \cos 2\pi\ell - i \sin 2\pi\ell \ \Rightarrow\ \cos 2\pi\ell = 1, \quad \sin 2\pi\ell = 0$$

$$\Rightarrow\ \ell = 0, 1, 2, 3, \cdots. \tag{36.36}$$

It follows that ℓ can be zero or positive integers, not half integers. Since m_ℓ ranges from $-\ell$ to ℓ the magnetic quantum numbers can only be zero and integers, both positive and negative, i.e., we have

$$m_\ell = 0, \pm 1, \pm 2, \cdots, \quad -\ell \le m_\ell \le \ell. \tag{36.37}$$

[13] As will be seen in Eq. (37.42) later, m_ℓ is related to the magnetic property of the particle.

[14] *See* Eq. (17.42).

We can find the θ-dependence of $Y_{\ell,-\ell}(\theta, \varphi)$ from Eq. (36.22), i.e.,[15]

$$\left(\frac{\partial}{\partial\theta} - i\cot\theta\,\frac{\partial}{\partial\varphi}\right) \quad Y_{\ell(-\ell)}(\theta, \varphi) = 0 \tag{36.38}$$

$$\Rightarrow \quad f(\theta) = (\sin\theta)^{\ell}. \tag{36.39}$$

$$\Rightarrow \quad Y_{\ell,-\ell}(\theta, \varphi) = c\,(\sin\theta)^{\ell}\,e^{-i\ell\varphi}. \tag{36.40}$$

The constant c is determined by normalisation with respect to the angle variables, i.e.,

$$\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta\, d\theta\ |Y_{\ell,-\ell}(\theta, \varphi)|^2 = 1. \tag{36.41}$$

We can obtain other eigenfunctions from $Y_{\ell,-\ell}$ by Eq. (36.20). The general expression is[16]

$$Y_{\ell,m_\ell}(\theta, \varphi) = C_{\ell m_\ell}\,e^{im_\ell\varphi}\,\frac{1}{(\sin\theta)^{m_\ell}}\,\frac{d^{\ell-m_\ell}}{d(\cos\theta)^{\ell-m_\ell}}(\sin\theta)^{2\ell}, \tag{36.42}$$

where

$$C_{\ell m_\ell} = \frac{(-1)^{\ell}}{2^{\ell}\,\ell!}\sqrt{\left(\frac{2\ell+1}{4\pi}\right)\frac{(\ell+m_\ell)!}{(\ell-m_\ell)!}}. \tag{36.43}$$

These are the spherical harmonics mentioned in §16.2.2. The corresponding vectors $\vec{Y}_{\ell,m_\ell}$ form a complete orthonormal set in the Hilbert space $\vec{L}^2(\mathcal{S}_u)$.[17] There are no other eigenvalues and eigenvectors for $\widehat{L}^2(\mathcal{S}_u)$ and $\widehat{L}_z(\mathcal{S}_u)$. The functions $Y_{\ell,m_\ell}(\theta, \varphi)$ can be extended to $L^2(I\!R^3)$ by including a factor $R(r) \in L^2(I\!R^+, r^2dr)$.[18]

36.2 Annihilation and Creation Operators

36.2.1 Introduction

Orbital angular momentum operators satisfy the commutation relations in Eqs. (27.111) to (27.113). It is possible to construct

[15] Any multiplicative constant can be absorbed into the normalisation constant c in Eq. (36.34).
[16] Zettili pp. 291–295.
[17] Amrein, Jauch and Sinha pp. 459–460. Zettili p. 292.
[18] *See* Eqs. (37.54) and (24.63).

three operators satisfying these commutation relations in terms of annihilation and creation operators in an appropriate Hilbert space. Let us suppose that in a given Hilbert space $\vec{\mathcal{H}}$ it is possible to introduce two commuting pairs of annihilation and creation operators $\widehat{a}_1$, $\widehat{a}_1^\dagger$ and $\widehat{a}_2$, $\widehat{a}_2^\dagger$ such that each pair possessing properties P27.9(1) to P27.9(3) and the two pairs together forming an irreducible set in $\vec{\mathcal{H}}$.[19] Since $\widehat{a}_1$, $\widehat{a}_1^\dagger$ commute with $\widehat{a}_2$, $\widehat{a}_2^\dagger$ the following number operators

$$\widehat{N}_1 := \widehat{a}_1^\dagger \widehat{a}_1, \quad \widehat{N}_2 := \widehat{a}_2^\dagger \widehat{a}_2 \tag{36.44}$$

also commute with each other. They are selfadjoint and they share a complete orthonormal set of common eigenvectors which can be denoted by $\vec{\varphi}_{n_1,n_2}$ where $n_1, n_2 = 0, 1, 2, \cdots$, i.e., we have[20]

$$\widehat{N}_1 \vec{\varphi}_{n_1,n_2} = n_1 \vec{\varphi}_{n_1,n_2}, \quad \widehat{N}_2 \vec{\varphi}_{n_1,n_2} = n_2 \vec{\varphi}_{n_1,n_2}. \tag{36.45}$$

These eigenvectors can be generated from $\vec{\varphi}_{0,0}$, i.e.,

$$\vec{\varphi}_{n_1,n_2} = \frac{(\widehat{a}_1^\dagger)^{n_1}}{\sqrt{n_1!}} \frac{(\widehat{a}_2^\dagger)^{n_2}}{\sqrt{n_2!}} \vec{\varphi}_{0,0}, \tag{36.46}$$

The sum $\widehat{N} := \widehat{N}_1 + \widehat{N}_2$ admits $\vec{\varphi}_{n_1,n_2}$ as its eigenvectors, i.e.,

$$\widehat{N} \vec{\varphi}_{n_1,n_2} = n \vec{\varphi}_{n_1,n_2}, \quad n = (n_1 + n_2). \tag{36.47}$$

We can define three new operators $\widehat{J}_x$, $\widehat{J}_y$ and $\widehat{J}_z$ in terms of $\widehat{a}_1$, $\widehat{a}_1^\dagger$, $\widehat{a}_2$ and $\widehat{a}_2^\dagger$ by[21]

$$\widehat{J}_x := \frac{\hbar}{2}\left(\widehat{a}_1^\dagger \widehat{a}_2 + \widehat{a}_2^\dagger \widehat{a}_1\right), \tag{36.48}$$

$$\widehat{J}_y := \frac{\hbar}{2i}\left(\widehat{a}_1^\dagger \widehat{a}_2 - \widehat{a}_2^\dagger \widehat{a}_1\right), \tag{36.49}$$

$$\widehat{J}_z := \frac{\hbar}{2}\left(\widehat{a}_1^\dagger \widehat{a}_1 - \widehat{a}_2^\dagger \widehat{a}_2\right). \tag{36.50}$$

[19] *See* also properties P35.2(1) to P35.2(4). In the state space $\vec{L}^2(I\!R^2, dxdy)$ of an isotropic oscillator studied in §35.4, we have two commuting pairs of annihilation and creation operators, i.e., $\widehat{a}_x$ and $\widehat{a}_x^\dagger$ in Eqs. (35.78) and (35.79) and a similar pair $\widehat{a}_y$ and $\widehat{a}_y^\dagger$ defined in terms of $\widehat{y}$ and $\widehat{p}_y$.

[20] An example is the discussion in §35.4 on isotropic oscillators.

[21] Baym pp. 380–386. The subscripts x, y, z are just labels unrelated to the Cartesian coordinates. We could have labelled the operators as $\widehat{J}_1, \widehat{J}_2, \widehat{J}_3$.

These operators have the following properties:

P36.2.1(1) $\widehat{J}_x$, $\widehat{J}_y$ and $\widehat{J}_z$ satisfy the following commutation relations, i.e.,[22]

$$[\widehat{J}_x, \widehat{J}_y] = i\hbar \widehat{J}_z, \tag{36.51}$$
$$[\widehat{J}_z, \widehat{J}_x] = i\hbar \widehat{J}_y, \tag{36.52}$$
$$[\widehat{J}_y, \widehat{J}_z] = i\hbar \widehat{J}_x. \tag{36.53}$$

It follows that the sum of their squares $\widehat{J}^{\,2} = \widehat{J}_x^{\,2} + \widehat{J}_y^{\,2} + \widehat{J}_z^{\,2}$ commutes with $\widehat{J}_x$, $\widehat{J}_y$ and $\widehat{J}_z$.

P36.2.1(2) The operator $\widehat{J}_z$ is equal to

$$\widehat{J}_z = \frac{\hbar}{2}(\widehat{N}_1 - \widehat{N}_2). \tag{36.54}$$

P36.2.1(3) The operator $\widehat{J}^{\,2}$ is equal to[23]

$$\widehat{J}^{\,2} = \frac{\widehat{N}}{2}\left(\frac{\widehat{N}}{2} + 1\right)\hbar^2. \tag{36.55}$$

P36.2.1(4) Having expressed $\widehat{J}_z$ and $\widehat{J}^{\,2}$ in terms of $\widehat{N}_1$, $\widehat{N}_2$ and $\widehat{N}$ we can see that $\vec{\varphi}_{n_1,n_2}$ in Eq. (36.46) are the eigenvectors of $\widehat{J}_z$ and $\widehat{J}^{\,2}$ corresponding to eigenvalues

$$\frac{1}{2}(n_1 - n_2)\hbar \quad \text{and} \quad \frac{n}{2}\left(\frac{n}{2} + 1\right)\hbar^2, \tag{36.56}$$

where $n = n_1 + n_2$, i.e.,

$$\widehat{J}_z \vec{\varphi}_{n_1,n_2} = \frac{1}{2}(n_1 - n_2)\,\hbar\,\vec{\varphi}_{n_1,n_2}. \tag{36.57}$$
$$\widehat{J}^{\,2} \vec{\varphi}_{n_1,n_2} = \frac{n}{2}\left(\frac{n}{2} + 1\right)\hbar^2\,\vec{\varphi}_{n_1,n_2}. \tag{36.58}$$

The operators $\widehat{J}_z$ and $\widehat{J}^{\,2}$ are both selfadjoint, having a complete orthonormal set of eigenvectors $\vec{\varphi}_{n_1,n_2}$ with real eigenvalues.

[22] These are typical angular momentum commutation relations.
[23] It is straightforward, albeit tedious, to verify the result.

P36.2.1(5) We can introduce two new operators

$$\widehat{J}_+ := \widehat{J}_x + i\widehat{J}_y \quad \text{and} \quad \widehat{J}_- := \widehat{J}_x - i\widehat{J}_y. \tag{36.59}$$

These two operators are not selfadjoint. They are related to the creation and annihilation operators by

$$\widehat{J}_+ = \hbar\,\widehat{a}_1^\dagger \widehat{a}_2, \quad \widehat{J}_- = \hbar\,\widehat{a}_2^\dagger \widehat{a}_1. \tag{36.60}$$

Acting on $\vec{\varphi}_{n_1,n_2}$ these two operators would change the eigenvalue of $\widehat{J}_z$. When acting on $\vec{\varphi}_{0,0}$, the operators $\widehat{J}_+$ and $\widehat{J}_-$ will result in the zero vector. Details are given in P36.2.2(3) in the next subsection.

36.2.2 Notation

Let the eigenvalues and normalised eigenvectors of $\widehat{J}_z$ and $\widehat{J}^{\,2}$ be denoted, respectively, by[24]

$$m_j\hbar, \quad j(j+1)\hbar^2, \quad |j, m_j\rangle. \tag{36.61}$$

Then we have where

$$m_j = \frac{1}{2}(n_1 - n_2), \quad j = \frac{n}{2} = \frac{1}{2}(n_1 + n_2). \tag{36.62}$$

We can rewrite Eqs. (36.57) and (36.58) as

$$\widehat{J}_z|j, m_j\rangle = m\hbar\,|j, m_j\rangle, \tag{36.63}$$

$$\widehat{J}^{\,2}|j, m_j\rangle = j(j+1)\hbar^2\,|j, m_j\rangle, \tag{36.64}$$

The following properties are obvious:

P36.2.2(1) Possible values of m and j are

$$m_j = 0, \ \pm\frac{1}{2}, \ \pm 1, \ \pm\frac{3}{2}, \ 2, \ldots \tag{36.65}$$

$$= \{0, \ \pm\,\text{half integers}, \ \pm\,\text{integers}\}, \tag{36.66}$$

$$j = 0, \ \frac{1}{2}, \ 1, \ \frac{3}{2}, \ 2, \ \frac{5}{2}, \ldots\ldots \tag{36.67}$$

$$= \{0, \ +\,\text{half integers}, \ +\,\text{integers}\}. \tag{36.68}$$

[24] The subscript j in m_j indicates that m_j is related to j. Their relationship is shown in Eq. (36.69).

P36.2.2(2) Relationship between m and j is

$$m_j^2 \leq j^2 \quad \text{or equivalently} \quad -j \leq m_j \leq j. \tag{36.69}$$

P36.2.2(3) The operators $\widehat{J}_+ = \widehat{J}_x + i\widehat{J}_y$ and $\widehat{J}_- = \widehat{J}_x - i\widehat{J}_y$ act on $|j, m\rangle$ as "raising" and "lowering" operators by increasing and decreasing the value of m_ℓ while keeping the value of j unchanged, i.e.,

$$\widehat{J}_z\Big(\widehat{J}_+|j, m_j\rangle\Big) = (m_j + 1)\hbar\Big(\widehat{J}_+|j, m_j\rangle\Big), \tag{36.70}$$

$$\widehat{J}_z\Big(\widehat{J}_-|j, m\rangle\Big) = (m_j - 1)\hbar\Big(\widehat{J}_-|j, m_j\rangle\Big), \tag{36.71}$$

$$\widehat{J}^2\Big(\widehat{J}_\pm|j, m_j\rangle\Big) = j(j+1)\hbar^2\Big(\widehat{J}_\pm|j, m_j\rangle\Big). \tag{36.72}$$

More explicitly we have[25]

$$\widehat{J}_+|j, m_j\rangle = \hbar\sqrt{(j - m_j)(j + m_j + 1)}\,|j, m_j + 1\rangle, \tag{36.73}$$

$$\widehat{J}_-|j, m_j\rangle = \hbar\sqrt{(j + m_j)(j - m_j + 1)}\,|j, m_j - 1\rangle. \tag{36.74}$$

A similar derivation can be applied to the pair $\widehat{J}_x$ and $\widehat{J}^2$ and the pair $\widehat{J}_y$ and $\widehat{J}^2$ to obtain the same eigenvalues as $\widehat{J}_z$ and $\widehat{J}^2$.[26]

36.2.3 Summary of Results

Let $\widehat{J}_x, \widehat{J}_y$ and $\widehat{J}_z$ be any three operators defined in terms of two pairs of commuting annihilation and creation operators by Eqs. (36.51), (36.52) and (36.53). Then we have the following results:

R36.2.3(1)

$$[\widehat{J}_x, \widehat{J}_y] = i\hbar\widehat{J}_z, \quad [\widehat{J}_z, \widehat{J}_x] = i\hbar\widehat{J}_y, \quad [\widehat{J}_y, \widehat{J}_z] = i\hbar\widehat{J}_x. \tag{36.75}$$

R36.2.3(2) The sum of their squares $\widehat{J}^2 = \widehat{J}_x^2 + \widehat{J}_y^2 + \widehat{J}_z^2$ would commute with $\widehat{J}_x$, $\widehat{J}_y$ and $\widehat{J}_z$.

---

[25] One can verify the results using Eq. (36.60) for $\widehat{J}_\pm$ and $\vec{\varphi}_{n_1,n_2}$ for $|j, m_j\rangle$.

[26] It follows that the operators $\widehat{J}_x$ and $\widehat{J}_y$ are also selfadjoint.

R36.2.3(3) The operators $\widehat{J}_{\pm} = \widehat{J}_x \pm i\widehat{J}_y$ commute with $\widehat{J}^{\,2}$ but not with $\widehat{J}_z$. We have

$$[\widehat{J}^{\,2}, \widehat{J}_{\pm}] = \widehat{0}, \quad [\widehat{J}_{\pm}, \widehat{J}_z] = \pm\hbar\widehat{J}_{\pm} \tag{36.76}$$

R36.2.3(4) For $\widehat{J}_z$ and $\widehat{J}^{\,2}$, we have

$$\widehat{J}_z|j, m_j\rangle = m_j\hbar\,|j, m_j\rangle, \tag{36.77}$$
$$\widehat{J}^{\,2}|j, m_j\rangle = j(j+1)\hbar^2\,|j, m_j\rangle, \tag{36.78}$$

where

$$m_j = \{0, \pm\text{half integers}, \pm\text{integers}\}, \tag{36.79}$$
$$j = \{0, +\text{half integers}, +\text{integers}\}, \tag{36.80}$$
$$-j \le m_j \le j. \tag{36.81}$$

Similar results apply to $\widehat{J}_x$ and $\widehat{J}^{\,2}$ and to $\widehat{J}_y$ and $\widehat{J}^{\,2}$.

R36.2.3(5) The eigenvectors of the z-component angular momentum can be raised or lowered using the operators $\widehat{J}_{\pm} = \widehat{J}_x \pm i\widehat{J}_y$ without changing the total angular momentum, i.e.,

$$\widehat{J}_+|j, m_j\rangle = \hbar\sqrt{(j-m_j)(j+m_j+1)}\;|j, m_j+1\rangle, \tag{36.82}$$
$$\widehat{J}_-|j, m_j\rangle = \hbar\sqrt{(j+m_j)(j-m_j+1)}\;|j, m_j-1\rangle. \tag{36.83}$$

In particular, we have

$$\widehat{J}_+|j, j\rangle = \vec{0}, \quad \widehat{J}_-|j, -j\rangle = \vec{0}. \tag{36.84}$$

Equations (36.13) to (36.32) are derived using the commutation relations of orbital angular momentum operators in Eqs. (27.111) to (27.114) without reference to any annihilation and creation operators. In the same way, the results in R36.2.3(2) to R36.2.3(5) in this subsection can be derived using the commutation relations in R36.2.3(1). This suggests that:

(1) The restrictions of $\widehat{J}_z$ and $\widehat{J}^{\,2}$ to the subspace of $\vec{\mathcal{H}}$ spanned by the eigenvectors $|j, m_j\rangle$ of integer values of j correspond to the orbital angular momentum operators $\widehat{L}_z(\mathcal{S}_u)$ and $\widehat{L}^2(\mathcal{S}_u)$.

(2) The restriction of $\widehat{J}_z$ and $\widehat{J}^2$ to the two-dimensional subspace spanned by the eigenvectors $|j = \frac{1}{2}, m_j\rangle$, where $m_j = \pm 1/2$, would corresponding to spin angular momentum operators $\widehat{S}_z$ and $\widehat{S}^2$ introduced in §14.1.1.[27]

36.3 Spin Angular Momentum

36.3.1 Introduction

An electron possesses an intrinsic angular momentum similar to a spinning particle and this is unrelated to its orbital motion or any other spatial motion.[28] We call this the electron's spin angular momentum or spin for short. The spin angular momentum has no classical counterpart. Postulate 27.2(CQ) on quantisation does not apply to spin. We have to make some ad hoc assumptions in order to establish appropriate vector space and operators for the description of spin which would conform to Postulates 25.1(PS) and Postulate 26.1(OV). A model theory of spin is set out in §14.1.1. Here we shall start with a description of the properties of spin.

Defining properties of spin operators

P36.3(1) *Spin angular momentum is a quantity with three components S_x, S_y and S_z along the x, y and z directions. These components can be written formally in vector notation as $\vec{S} = (S_x, S_y, S_z)$.*

P36.3(2) *Each of the spin components can take only two values, i.e., $\pm\hbar/2$. Hence, the state space for the spin is assumed to be a two-dimensional Hilbert space $\vec{\mathbb{V}}^2$.*

---

[27] *See* Zettili pp. 277–279 for the restriction to the subspace spanned by the eigenvectors $|j = 1, m_j\rangle$, where $m_j = -1, 0, 1$.

[28] Electron spin was postulated by Goudsmith and Uhlenbeck in 1925 to explain complex atomic spectra. The basis of their hypothesis can be traced back to the Stern–Gerlach experiment performed in 1922. The theoretical origin of electron spin came from Dirac's relativistic quantum theory. Uhlenbeck (1900–1988) and Goudsmith (1902–1978) were Dutch-American theoretical physicists. Stern (1888–1969) and Gerlach (1889–1979) were German physicists.

P36.3(3) *The spin components are represented by three selfadjoint operators $\widehat{S}_x$, $\widehat{S}_y$ and $\widehat{S}_z$ on the state space $\vec{\vec{\mathbb{V}}}^2$ symbolically written in vector notation as*

$$\vec{\widehat{S}} = \left(\widehat{S}_x, \widehat{S}_y, \widehat{S}_z\right). \tag{36.85}$$

These operators possess only two eigenvalues, i.e., $\pm\hbar/2$.

P36.3(4) *Spin operators obey the same commutation relations as that of orbital angular momentum operators, i.e.,*

$$[\widehat{S}_x, \widehat{S}_y] = i\hbar\widehat{S}_z, \quad [\widehat{S}_z, \widehat{S}_x] = i\hbar\widehat{S}_y, \quad [\widehat{S}_y, \widehat{S}_z] = i\hbar\widehat{S}_x. \tag{36.86}$$

P36.3(5) *Spin motion is unrelated to spatial motion. This means that spin operators $\widehat{S}_x$, $\widehat{S}_y$, $\widehat{S}_z$ commute with spatial operators such as $\vec{\widehat{x}}$, $\vec{\widehat{p}}$ and $\vec{\widehat{L}}$.*

The square of the total spin angular momentum operator commutes with each of the component operators, i.e., we have

$$\widehat{S}^2 = \widehat{S}_x^2 + \widehat{S}_y^2 + \widehat{S}_y^2 \tag{36.87}$$

$$[\,\widehat{S}_x, \widehat{S}^2\,] = [\,\widehat{S}_y, \widehat{S}^2\,] = [\,\widehat{S}_z, \widehat{S}^2\,] = 0. \tag{36.88}$$

Spin has been discussed in §10.2.2 and §14.1.1 to illustrate how a probabilistic theory can be established in a complex vector space. In the remainder of this chapter, we shall present an intuitive formulation of the theory electron spin.

36.3.2 Two-Component Wave Functions

At any particular time a spinless particle has associated with it a wave function $\phi(\vec{x})$ so that the position probability density function of the particle is given by $|\phi(\vec{x})|^2$.[29] A spin-$\frac{1}{2}$ particle would have a probability density of being found to be at $\vec{x}$ with spin-up and a separate probability density of being found to be at $\vec{x}$ with spin down. In other words, we require two probability distribution functions, one for spin-up and one for spin-down. A

[29] The time variable in all the wave functions has been suppressed for brevity.

single wave function is unable to meet such a requirement. The need for two probability density functions suggests we should employ a pair of wave functions $\phi_+(\vec{x})$ and $\phi_-(\vec{x})$ with $|\phi_+(\vec{x})|^2$ as the position probability density function of the particle with spin-up, and $|\phi_-(\vec{x})|^2$ as the position probability density of the particle with spin down. If the particle is known to have its spin-up everywhere, then $\phi_-(\vec{x}) = 0$ and if the particle is known to have spin down everywhere, then $\phi_+(\vec{x}) = 0$. These two wave functions are separately normalised, i.e., we have

$$\int_{-\infty}^{\infty} |\phi_+(\vec{x})|^2 \, dxdydz = 1 = \int_{-\infty}^{\infty} |\phi_-(\vec{x})|^2 \, dxdydz = 1. \quad (36.89)$$

We can combine these two wave functions into a single *two-component wave function* in the following manner:

(1) If the particle is known to have its spin-up everywhere, then it is described by a two-component wave function of the form

$$\begin{pmatrix} \phi_+(\vec{x}) \\ 0 \end{pmatrix}. \quad (36.90)$$

(2) If the particle is known to have its spin down everywhere, the corresponding two-component wave function is

$$\begin{pmatrix} 0 \\ \phi_-(\vec{x}) \end{pmatrix}. \quad (36.91)$$

(3) If the particle does not have its spin-up or spin-down everywhere, it is a two-component wave function and is a combination of the two previous ones, i.e.,

$$\begin{pmatrix} c_+\phi_+(\vec{x}) \\ c_-\phi_-(\vec{x}) \end{pmatrix}, \quad c_+, c_- \in \mathbb{C} \text{ and } |c_+|^2 + |c_-|^2 = 1. \quad (36.92)$$

This intuitive approach can be formalised in terms of spin functions.

36.3.3 Spin Functions and Spin Vectors

Introduce a discrete variable s_c, to be referred to as the **spin coordinate**, which takes only two values, i.e., $s_c = \pm 1/2$.[30] Given a

[30] The subscript c in s_c stands for "coordinate."

two-component wave function we can define a new one-component function of $\vec{x}$ and s_c, written as $\Phi(\vec{x}, s_c)$, by

$$\Phi(\vec{x}, 1/2) := \phi_+(\vec{x}), \quad \Phi(\vec{x}, -1/2) := \phi_-(\vec{x}). \tag{36.93}$$

The interpretation is that the function $\Phi(\vec{x}, s_c)$ for the value $s_c = 1/2$, i.e., $\Phi(\vec{x}, 1/2)$, represents a spin-up state and the function $\Phi(\vec{x}, s_c)$ for the value $s_c = -1/2$, i.e., $\Phi(\vec{x}, -1/2)$, represents a spin-down state.

What we have done here is to embody the two components of a two-component wave function into the spin coordinate. This is more than just a matter of notation. The introduction of the spin coordinate leads to the useful concept of spin functions on which spin operators can be defined.

Just as we have functions of spatial coordinates $\vec{x}$, we can introduce functions of the spin coordinate s_c. These functions would have only two values, one for $s_c = 1/2$ and one for $s_c = -1/2$. Such functions are called **spin functions**. Two simple examples, denoted by $\alpha(s_c)$ and $\beta(s_c)$, are defined as follows:

$$\alpha(s_c) := \begin{cases} 1, & s_c = 1/2 \\ 0, & s_c = -1/2 \end{cases}, \tag{36.94}$$

$$\beta(s_c) := \begin{cases} 0, & s_c = 1/2 \\ 1, & s_c = -1/2 \end{cases}. \tag{36.95}$$

A general spin function $\eta(s_c)$ is a linear combination of $\alpha(s_c)$ and $\beta(s_c)$, i.e.,

$$\eta(s_c) := c_+ \alpha(s_c) + c_- \beta(s_c), \qquad c_+, \ c_- \in \mathbb{C}. \tag{36.96}$$

We have

$$\eta(s_c) = \begin{cases} c_+, & s_c = 1/2 \\ c_-, & s_c = -1/2 \end{cases}. \tag{36.97}$$

An immediate application of spin functions is the separation of the wave function $\Phi(\vec{x}, s_c)$ defined by Eq. (36.93) into a spatial part and a spin part. The spin-up and spin-down states in Eqs. (36.90) and (36.91) can be identified, respectively, with

$$\Phi_+(\vec{x}, s_c) = \phi_+(\vec{x})\alpha(s_c), \tag{36.98}$$

$$\Phi_-(\vec{x}, s_c) = \phi_-(\vec{x})\beta(s_c). \tag{36.99}$$

A general spin state can be described by the function

$$\Phi(\vec{x}, s_c) = c_+\Phi_+(\vec{x}, s_c) + c_-\Phi_-(\vec{x}, s_c) \tag{36.100}$$

$$= c_+\phi_+(\vec{x}\,)\alpha(s_c) + c_-\phi_-(\vec{x}\,)\beta(s_c). \tag{36.101}$$

Under the usual addition and multiplication by complex numbers spin functions define a two-dimensional complex vector space with $\alpha(s_c)$ and $\beta(s_c)$ forming a basis for the space. We can also define a scalar product in accordance with Eq. (12.17). Given two spin functions $\eta(s_c)$ and $\eta'(s_c)$

$$\eta(s_c) = c_+\,\alpha(s_c) + c_-\,\beta(s_c), \tag{36.102}$$

$$\eta'(s_c) = c'_+\,\alpha(s_c) + c'_-\,\beta(s_c) \tag{36.103}$$

we define their scalar product as

$$\langle \eta \mid \eta' \rangle := \eta(1/2)^*\eta'(1/2) + \eta(-1/2)^*\eta'(-1/2) \tag{36.104}$$

$$= c_+^* c'_+ + c_-^* c'_- \tag{36.105}$$

$$\Rightarrow \quad \langle \alpha \mid \alpha \rangle = 1, \quad \langle \beta \mid \beta \rangle = 1, \quad \langle \alpha \mid \beta \rangle = 0. \tag{36.106}$$

Let the resulting two-dimensional Hilbert space be denoted by $\vec{\mathbb{W}}^{(2)}$ and vectors corresponding to $\alpha(s_c)$, $\beta(s_c)$, $\eta(s_c)$ be denoted by $\vec{\alpha}$, $\vec{\beta}$, $\vec{\eta}$ without the argument s_c.[31] We call these **spin vectors**. The two spin vectors $\vec{\alpha}$ and $\vec{\beta}$ form an orthonormal basis for $\vec{\mathbb{W}}^{(2)}$. Any spin vector is a linear combination of $\vec{\alpha}$ and $\vec{\beta}$, i.e.,

$$\vec{\eta} = c_+\,\vec{\alpha} + c_-\,\vec{\beta} \;\text{ with }\; c_+ = \langle \vec{\alpha} \mid \vec{\eta} \rangle, \;\; c_- = \langle \vec{\beta} \mid \vec{\eta} \rangle. \tag{36.107}$$

When we want to treat the spatial and spin functions in Eq. (36.101) as vectors, the products $\phi_+(\vec{x}\,)\alpha(s_c)$ and $c_-\phi_-(\vec{x}\,)\beta(s_c)$ should be treated as tensor products. This is already discussed in §33.2, i.e., we would adopt the tensor product space $\vec{\mathcal{H}}^{(s)}(I\!R^3) := \vec{L}^2(I\!R^3) \otimes \vec{\mathbb{W}}^{(2)}$ in Eq. (33.10) as the state space of a spin-$\frac{1}{2}$ particle. A vector in $\vec{\mathcal{H}}^{(s)}(I\!R^3)$ is of the form[32]

$$\vec{\Phi} = c_+\vec{\phi}_+ \otimes \vec{\alpha} + c_-\vec{\phi}_- \otimes \vec{\beta}. \tag{36.108}$$

[31] This is in keeping with our notation when treating functions such as $\phi(x)$ as vectors.
[32] Note that $\vec{\phi}_+$ and $\vec{\phi}_-$ are normalised separately in accordance with Eq. (36.89).

The spin-up and spin-down states correspond to tensor product vectors

$$\vec{\Phi}_+ = \vec{\phi}_+ \otimes \vec{\alpha} \quad \text{and} \quad \vec{\Phi}_- = \vec{\phi}_- \otimes \vec{\beta}. \tag{36.109}$$

36.3.4 Spin Operator $\widehat{S}_z$

Spin is unrelated to spatial motion. Spin operators should not operate on functions of spatial coordinates which describe spatial motion. They should operate on spin functions. Mathematically spin operators should act on $\vec{\mathbb{V}}^{(2)}$. In order to act on $\vec{\Phi}$, the spin operators should take the form given by Eq. (33.11), i.e., when acting on $\vec{\mathcal{H}}^{(s)}(I\!R^3)$ the spin operators along the x, y and z components become $\widehat{I\!I}(I\!R^3) \otimes \widehat{S}_x$, $\widehat{I\!I}(I\!R^3) \otimes \widehat{S}_y$ and $\widehat{I\!I}(I\!R^3) \otimes \widehat{S}_z$. We have

$$\left(\widehat{I\!I}(I\!R^3) \otimes \widehat{S}_z\right)\vec{\Phi} = c_+\vec{\phi}_+ \otimes \left(\widehat{S}_z\vec{\alpha}\,\right) + c_-\vec{\phi}_- \otimes \left(\widehat{S}_z\vec{\beta}\,\right). \tag{36.110}$$

The effects of $\widehat{S}_z$ on spin vectors can be established by the following physical arguments:

(1) For the particle to be in a spin-up state everywhere, there must be no probability of its being found to have its spin down anywhere. In accordance with Eqs. (36.90) and (36.109), this means that the state vector must be of the form of $\vec{\Phi}_+$

(2) This state vector must also be an eigenvector of $\widehat{S}_z$ corresponding to the eigenvalue $\frac{1}{2}\hbar$ in accordance with Postulate 28.1(PDDO),[33] i.e., we must have

$$\left(\widehat{I\!I}(I\!R^3) \otimes \widehat{S}_z\right)\vec{\Phi}_+ = \frac{1}{2}\hbar\,\vec{\Phi}_+ \tag{36.111}$$

$$\Rightarrow \qquad \vec{\phi}_+ \otimes \left(\widehat{S}_z\vec{\alpha}\,\right) = \vec{\phi}_+ \otimes \frac{1}{2}\hbar\,\vec{\alpha} \tag{36.112}$$

$$\Rightarrow \qquad \widehat{S}_z\vec{\alpha} = \frac{1}{2}\hbar\,\vec{\alpha}. \tag{36.113}$$

[33] *See* C28.1(PDDO)(2) in particular.

(3) Similarly by considering a spin-down state vector of the form $\vec{\Phi}_-$ we get

$$\left(\widehat{I\!I}(I\!R^3)\otimes \widehat{S}_z\right)\vec{\Phi}_- = -\frac{1}{2}\,\hbar\,\vec{\Phi}_- \tag{36.114}$$

$$\Rightarrow \qquad \widehat{S}_z\vec{\beta} = -\frac{1}{2}\,\hbar\,\vec{\beta}. \tag{36.115}$$

(4) For a general spin vector $\vec{\eta}$, we have

$$\widehat{S}_z\,\vec{\eta} = c_+\widehat{S}_z\,\vec{\alpha} + c_-\widehat{S}_z\vec{\beta} = \frac{1}{2}\,\hbar\,(c_+\vec{\alpha} - c_-\vec{\beta}\,). \tag{36.116}$$

To emphasise the link of these spin functions with the z-component spin and in keeping with the notation used in §14.1.1, we shall re-label them with a subscript z, i.e., we rewrite

$$\boxed{\vec{\alpha} \text{ as } \vec{\alpha}_z, \text{ and } \vec{\beta} \text{ as } \vec{\beta}_z.} \tag{36.117}$$

In this notation, we have

$$\widehat{S}_z\,\vec{\alpha}_z = \frac{1}{2}\hbar\,\vec{\alpha}_z, \quad \widehat{S}_z\vec{\beta}_z = -\frac{1}{2}\hbar\,\vec{\beta}_z, \tag{36.118}$$

$$\widehat{S}_z\,\vec{\eta}\ = c_+\widehat{S}_z\,\vec{\alpha}_z + c_-\widehat{S}_z\vec{\beta}_z = \frac{1}{2}\,\hbar\,(c_+\vec{\alpha}_z - c_-\vec{\beta}_z\,). \tag{36.119}$$

These results imply that

> *the spin vector $\vec{\alpha}_z$ is the eigenvector of $\widehat{S}_z$ corresponding to the eigenvalue $\frac{1}{2}\,\hbar$ and the spin vector $\vec{\beta}_z$ is the eigenvector of $\widehat{S}_z$ corresponding to the eigenvalue $-\frac{1}{2}\,\hbar$.*

Physically the above statement means that $\vec{\alpha}_z$ represents the spin-up state and $\vec{\beta}_z$ represents the spin-down state.

36.3.5 Spin Operators $\widehat{S}_x$ and $\widehat{S}_y$

As pointed out earlier, Eqs. (36.82) to (36.84) are derived using only the commutation relations of the operators involved. It follows that these results can be applied to spin operators provided we restrict ourselves to the two-dimensional subspace corresponding to

$j = 1/2$.[34] This becomes obvious if we change the notation:

1. Replace j by s. This new quantum number s takes only one value, i.e., $s = 1/2$.

2. Replace m_j by m_s. This new quantum number takes only two values, i.e., $m_s = \pm 1/2$.

3. Replace $|\, j = 1/2,\, m_j = 1/2\,\rangle$ by $\vec{\alpha}_z$ and $|\, j = 1/2,\, m_j = -1/2\,\rangle$ by $\vec{\beta}_z$.

4. Replace $\widehat{J}_x, \widehat{J}_y, \widehat{J}_z$ and $\widehat{J}^{\,2}$ by $\widehat{S}_x, \widehat{S}_y, \widehat{S}_z$, and $\widehat{S}^2$, respectively. Then, instead of $\widehat{J}_\pm = \widehat{J}_x \pm i\widehat{J}_y$, we now have

$$\widehat{S}_+ = \widehat{S}_x + i\widehat{S}_y, \quad \widehat{S}_- = \widehat{S}_x - i\widehat{S}_y. \tag{36.120}$$

5. Equations (36.77) and (36.78) become

$$\widehat{S}_z\,\vec{\alpha}_z = \frac{1}{2}\,\hbar\,\vec{\alpha}_z, \quad \widehat{S}_z\,\vec{\beta}_z = -\frac{1}{2}\,\hbar\,\vec{\beta}_z, \tag{36.121}$$

$$\widehat{S}^2\,\vec{\alpha}_z = s(s+1)\hbar^2\,\vec{\alpha}_z = \frac{3}{2}\,\hbar^2\,\vec{\alpha}_z, \tag{36.122}$$

$$\widehat{S}^2\vec{\beta}_z = s(s+1)\hbar^2\,\vec{\beta}_z = \frac{3}{2}\,\hbar^2\,\vec{\beta}_z. \tag{36.123}$$

6. Equations (36.82), (36.83) and (36.84) become[35]

$$(\widehat{S}_x + i\,\widehat{S}_y)\,\vec{\beta}_z = \hbar\,\vec{\alpha}_z, \quad (\widehat{S}_x - i\,\widehat{S}_y)\,\vec{\alpha}_z = \hbar\,\vec{\beta}_z, \tag{36.124}$$

$$(\widehat{S}_x + i\,\widehat{S}_y)\,\vec{\alpha}_z = \vec{0}, \quad (\widehat{S}_x - i\,\widehat{S}_y)\,\vec{\beta}_z = \vec{0}. \tag{36.125}$$

7. It follows that the action of $\widehat{S}_x$ and $\widehat{S}_y$ on $\vec{\alpha}_z$ and $\vec{\beta}_z$ are

$$\widehat{S}_x\vec{\alpha}_z = \frac{1}{2}\,\hbar\,\vec{\beta}_z, \quad \widehat{S}_x\vec{\beta}_z = \frac{1}{2}\,\hbar\,\vec{\alpha}_z, \tag{36.126}$$

$$\widehat{S}_y\vec{\alpha}_z = \frac{1}{2}\,i\hbar\,\vec{\beta}_z, \quad \widehat{S}_y\vec{\beta}_z = -\frac{1}{2}\,i\hbar\,\vec{\alpha}_z. \tag{36.127}$$

8. It is instructive to derive explicitly the action of $\widehat{S}_x$ and $\widehat{S}_y$ on $\vec{\alpha}_z$ and $\vec{\beta}_z$ using the commutation relations of the spin operators and

[34] Alternatively we can consider Eqs. (36.20) to (36.22) restricted to $\ell = 1/2$. *See* Zettili pp. 277–278 for restriction to $j = 1$.

[35] *See* Q36(7).

$\widehat{S}_\pm$ instead of using Eqs. (36.82) to (36.84). We can start with the following commutation relations:

$$[\,\widehat{S}_z,\,\widehat{S}_\pm\,] = \pm\hbar\,\widehat{S}_\pm, \qquad [\widehat{S}^2,\,\widehat{S}_\pm\,] = \widehat{0}. \tag{36.128}$$

$$\widehat{S}_z\widehat{S}_\pm = \widehat{S}_\pm\widehat{S}_z \pm \hbar\widehat{S}_\pm, \qquad \widehat{S}^2\widehat{S}_\pm = \widehat{S}_\pm\widehat{S}^2. \tag{36.129}$$

(1) Letting $\widehat{S}_z\widehat{S}_+$ act on $\vec{\alpha}_z$ and $\widehat{S}_z\widehat{S}_-$ act on $\vec{\beta}_z$ we get

$$\widehat{S}_z\widehat{S}_+\vec{\alpha}_z = \big(\widehat{S}_+\widehat{S}_z + \hbar\widehat{S}_+\,\big)\vec{\alpha}_z = \frac{3}{2}\hbar\,\widehat{S}_+\vec{\alpha}_z, \tag{36.130}$$

$$\widehat{S}_z\widehat{S}_-\vec{\beta}_z = \big(\widehat{S}_-\widehat{S}_z - \hbar\,\widehat{S}_+\big)\vec{\beta}_z = -\frac{3}{2}\hbar\,\widehat{S}_+\vec{\beta}_z. \tag{36.131}$$

These results appear to imply that $\widehat{S}_+\vec{\alpha}_z$ is an eigenvector of $\widehat{S}_z$ corresponding to eigenvalue $3\hbar/2$ and $\widehat{S}_-\vec{\beta}_z$ is an eigenvector of $\widehat{S}_z$ corresponding to eigenvalue $-3\hbar/2$. Such a conclusion contradicts P36.3(2). To avoid contradiction, we must not allow $\widehat{S}_+$ to create a new eigenvector from $\vec{\alpha}_z$ and we must also not allow $\widehat{S}_-$ to create a new eigenvector from $\vec{\beta}_z$, i.e., P36.3(2) imposes the following condition:

$$(\widehat{S}_x + i\,\widehat{S}_y)\,\vec{\alpha}_z = \vec{0}, \quad (\widehat{S}_x - i\,\widehat{S}_y)\,\vec{\beta}_z = \vec{0}. \tag{36.132}$$

(2) Next letting $\widehat{S}_z\widehat{S}_+$ act on $\vec{\beta}_z$ and $\widehat{S}_z\widehat{S}_-$ act on $\vec{\alpha}_z$ we get

$$\widehat{S}_z\widehat{S}_+\vec{\beta}_z = \frac{1}{2}\hbar\,\widehat{S}_+\vec{\beta}_z \quad\Rightarrow\quad \widehat{S}_+\vec{\beta}_z = \hbar\vec{\alpha}_z, \tag{36.133}$$

$$\widehat{S}_z\widehat{S}_-\vec{\alpha}_z = -\frac{1}{2}\hbar\,\widehat{S}_-\vec{\alpha}_z \quad\Rightarrow\quad \widehat{S}_-\vec{\alpha}_z = \hbar\vec{\beta}_z. \tag{36.134}$$

Explicitly we have

$$(\widehat{S}_x + i\,\widehat{S}_y)\,\vec{\beta}_z = \hbar\,\vec{\alpha}_z, \quad (\widehat{S}_x - i\,\widehat{S}_y)\,\vec{\alpha}_z = \hbar\,\vec{\beta}_z. \tag{36.135}$$

The factor $\hbar$ on the right-hand side can be verified by

$$\begin{aligned}\langle\widehat{S}_+\vec{\beta}_z \mid \widehat{S}_+\vec{\beta}_z\rangle &= \langle\vec{\beta}_z \mid \widehat{S}_-\widehat{S}_+\vec{\beta}_z\rangle \\ &= \langle\vec{\beta}_z \mid \big(\widehat{S}^2 - \widehat{S}_z^2 - \hbar\widehat{S}_z\big)\vec{\beta}_z\rangle = \hbar^2.\end{aligned} \tag{36.136}$$

Then Eqs. (36.126) and (36.127) follow as before.

9. The normalised eigenvectors of $\widehat{S}_x$ and $\widehat{S}_y$ are seen to be

$$\vec{\alpha}_x := \frac{1}{\sqrt{2}}(\vec{\alpha}_z + \vec{\beta}_z), \qquad \vec{\beta}_x := \frac{1}{\sqrt{2}}(\vec{\alpha}_z - \vec{\beta}_z), \tag{36.137}$$

$$\vec{\alpha}_y := \frac{1}{\sqrt{2}}(\vec{\alpha}_z + i\vec{\beta}_z), \qquad \vec{\beta}_y := \frac{1}{\sqrt{2}}(\vec{\alpha}_z - i\vec{\beta}_z). \tag{36.138}$$

We can summarise our discussion on spin as follows:

(1) The *state space of the spin motion* is taken to be $\vec{\mathbb{W}}^{(2)}$. The spin operators $\widehat{S}_x$, $\widehat{S}_y$, and $\widehat{S}_z$ defined on $\vec{\mathbb{W}}^{(2)}$ satisfy P36.3(1) to P36.3(5) of the defining properties of spin.

(2) When both spin motion and spatial motion are included, the *state space of a spin-$\frac{1}{2}$ particle* is taken to be the tensor product of $\vec{L}^2(I\!R^3)$ and $\vec{\mathbb{W}}^{(2)}$, i.e., $\mathcal{H}^{(s)}(I\!R^3) = \vec{L}^2(I\!R^3) \otimes \vec{\mathbb{W}}^{(2)}$.

(3) A general state vector is given by Eq. (36.108). The spin-up and spin-down state vectors are given by Eq. (36.109). These are consistent with the results in Eqs. (36.113) and (36.115).

(4) The scalar product for vectors like $\vec{\Phi}$ in Eq. (36.108) is calculated using the linearity properties of scalar product and tensor product and Eq. (24.44).

(5) To act on $\mathcal{H}^{(s)}(I\!R^3)$ operators for the spatial motion take the form $\widehat{A} \otimes \widehat{I\!I}(\vec{\mathbb{W}}^{(2)})$, e.g., the kinetic energy operator $\widehat{K}$ in Eq. (27.91) becomes $\widehat{K} \otimes \widehat{I\!I}(\vec{\mathbb{W}}^{(2)})$.[36]

36.3.6 Matrix Representation

Being two-dimensional the state space $\vec{\mathbb{W}}^2$ of spin motion can be represented by the two-dimensional space $\vec{\mathbb{C}}^2$. We can obtain explicit representations of spin vectors and spin operators on $\vec{\mathbb{C}}^2$ by two-component column vectors and 2×2 matrices, respectively.

36.3.6.1 Two-component column vectors

In the orthonormal basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$, the eigenvectors $\vec{\alpha}_z$, $\vec{\beta}_z$, $\vec{\alpha}_x$, $\vec{\beta}_x$, $\vec{\alpha}_y$ and $\vec{\beta}_y$ of the spin operators have the following representation[37]:

$$\boldsymbol{C}_{\vec{\alpha}_z} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{C}_{\vec{\beta}_z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \tag{36.139}$$

$$\boldsymbol{C}_{\vec{\alpha}_x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{C}_{\vec{\beta}_x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}; \tag{36.140}$$

[36] The identity operator on $\vec{\mathbb{W}}^{(s)}$ is $\widehat{I\!I}(\vec{\mathbb{W}}^{(2)})$.

[37] In the notation of §7.5. These results can be easily established using Eq. (13.107) for matrix representation of operators.

$$\boldsymbol{C}_{\vec{\alpha}_y} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \qquad \boldsymbol{C}_{\vec{\beta}_y} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{36.141}$$

A general spin vector $\vec{\eta} = c_+\vec{\alpha}_z + c_-\vec{\beta}_z$ has the following representation:

$$\boldsymbol{C}_{\vec{\eta}} = \begin{pmatrix} c_+ \\ c_- \end{pmatrix}. \tag{36.142}$$

36.3.6.2 Spin operators as 2×2 matrices

To act on two-component column vectors, the matrix representation of spin operators must be 2×2 matrices. Let us denote the matrix representation of $\widehat{S}_z$, $\widehat{S}_x$ and $\widehat{S}_y$ in the basis $\{\vec{\alpha}_z,\ \vec{\beta}_z\}$ for $\vec{\mathbb{C}}^2$ by $\boldsymbol{M}_{\hat{S}_z}$, $\boldsymbol{M}_{\hat{S}_x}$ and $\boldsymbol{M}_{\hat{S}_y}$, respectively.[38] Using Eq. (13.107), we get the following representations:

$$\boldsymbol{M}_{\hat{S}_z} = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{36.143}$$

$$\boldsymbol{M}_{\hat{S}_x} = \frac{1}{2}\hbar \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{36.144}$$

$$\boldsymbol{M}_{\hat{S}_y} = \frac{1}{2}\hbar \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{36.145}$$

It is easily verified that

$$\boldsymbol{M}_{\hat{S}_z}\boldsymbol{C}_{\vec{\alpha}_z} = \frac{1}{2}\hbar\boldsymbol{C}_{\vec{\alpha}_z}, \qquad \boldsymbol{M}_{\hat{S}_z}\boldsymbol{C}_{\vec{\beta}_z} = -\frac{1}{2}\hbar\boldsymbol{C}_{\vec{\beta}_z}; \tag{36.146}$$

$$\boldsymbol{M}_{\hat{S}_x}\boldsymbol{C}_{\vec{\alpha}_x} = \frac{1}{2}\hbar\boldsymbol{C}_{\vec{\alpha}_x}, \qquad \boldsymbol{M}_{\hat{S}_x}\boldsymbol{C}_{\vec{\beta}_x} = -\frac{1}{2}\hbar\boldsymbol{C}_{\vec{\beta}_x}; \tag{36.147}$$

$$\boldsymbol{M}_{\hat{S}_y}\boldsymbol{C}_{\vec{\alpha}_y} = \frac{1}{2}\hbar\boldsymbol{C}_{\vec{\alpha}_y}, \qquad \boldsymbol{M}_{\hat{S}_y}\boldsymbol{C}_{\vec{\beta}_y} = -\frac{1}{2}\hbar\boldsymbol{C}_{\vec{\beta}_y}. \tag{36.148}$$

The above matrices are related to Pauli matrices in Eq. (7.9) by

$$\boldsymbol{M}_{\hat{S}_x} = \frac{1}{2}\hbar\sigma_x, \quad \boldsymbol{M}_{\hat{S}_y} = \frac{1}{2}\hbar\sigma_y, \quad \boldsymbol{M}_{\hat{S}_z} = \frac{1}{2}\hbar\sigma_z. \tag{36.149}$$

The angular momentum commutation relations are clearly satisfied by these matrices. All these agree with the theory given in §14.1.1. It

[38] Following the notation in §8.2.3 and §13.5. These are called **spin matrices**.

is a standard practice to use the same notation for the spin operators and their matrix representations, e.g., to write $\boldsymbol{M}_{\hat{S}_z}$ in Eq. (36.143) as $\widehat{S}_z$.

Exercises and Problems

Q36(1) The spherical harmonics $Y_{1,1}(\theta, \varphi)$, $Y_{1,0}(\theta, \varphi)$, $Y_{1,-1}(\theta, \varphi)$ are given by Eqs. (16.65) to (16.67). In Cartesian coordinates, these functions are given by[39]

$$Y_{1,-1}(x, y, z) = \sqrt{\frac{3}{8\pi}}\,\frac{x - iy}{r}, \tag{36.150}$$

$$Y_{1,0}(x, y, z) = \sqrt{\frac{3}{4\pi}}\,\frac{z}{r}, \tag{36.151}$$

$$Y_{1,1}(x, y, z) = -\sqrt{\frac{3}{8\pi}}\,\frac{x + iy}{r}. \tag{36.152}$$

(a) Using the expression in Cartesian coordinates for the quantised orbital angular momentum operator $\widehat{L}_{cz}$ in Eq. (27.86), verify by explicit calculations that $Y_{1,1}(x, y, z)$, $Y_{1,0}(x, y, z)$ and $Y_{1,-1}(x, y, z)$ are the eigenfunctions of $\widehat{L}_{cz}$ corresponding to eigenvalues $\hbar$, 0 and $-\hbar$.

(b) Consider the following coordinate transformations:

$$x \to z', \quad y \to x', \quad z \to y'. \tag{36.153}$$

(i) Show that the component of the orbital angular momentum operator along the z'-direction is the same as that along the x-direction, i.e., show that $\widehat{L}_{cz'} = \widehat{L}_{cz}$.

(ii) Show that the simultaneous eigenfunctions of $\widehat{L}_c^2$ and $\widehat{L}_{cx}$ corresponding to eigenvalues of $\widehat{L}_c^2$ equal

[39] Zettili pp. 293–934. For convenience we have used the same symbols, e.g., $Y_{1,0}$ for the functions in Cartesian coordinates.

to $2\hbar^2$ are given Cartesian coordinates by

$$X_{1,-1}(x,y,z) = \sqrt{\frac{3}{8\pi}}\,\frac{y-iz}{r}, \tag{36.154}$$

$$X_{1,0}(x,y,z) = \sqrt{\frac{3}{4\pi}}\,\frac{x}{r}, \tag{36.155}$$

$$X_{1,1}(x,y,z) = -\sqrt{\frac{3}{8\pi}}\,\frac{y+iz}{r}. \tag{36.156}$$

Q36(2) Suppose $\widehat{L}_c^2$ and $\widehat{L}_{cz}$ are measured giving the eigenvalues $2\hbar^2$ and $-\hbar$, respectively. A measurement of $\widehat{L}_{cx}$ is then made. What are the possible results of the measurement of $\widehat{L}_{cx}$? Find the probability of each of these possible results.

Q36(3) The Hamiltonian of a classical particle of mass m constrained to move freely on the surface of a sphere of radius a is

$$H = \frac{1}{2I}\,L^2, \tag{36.157}$$

where $I = ma^2$ is the moment of inertia of the particle and L^2 is the total orbital angular momentum square of the particle, both with respect to the origin. Quantise the system and find the energy eigenvalues and the corresponding eigenfunctions of the quantised system. What are the degeneracy of the energy eigenvalues?[40]

Q36(4) Verify the commutation relation in Eq. (36.51).

Q36(5) Verify Eqs. (36.54), (36.55) and (36.60).

Q36(6) Verify properties P36.2.2(1), P36.2.2(2) and P36.2.2(3). including Eqs. (36.73) and (36.74).

Q36(7) How is the vector $\vec{\Phi}$ in Eq. (36.108) normalised?

Q36(8) Prove Eqs. (36.124) and (36.125) directly using the defining properties of spin operators given in P36.3(1) to P36.3(4) without using Eqs. (36.82), (36.83) and (36.84).

[40] Zettili pp. 296–297. Such a system is known as a *rigid rotator* which can be used to model a diatomic molecule. For the state space of a quantum rigid rotator, see the comment on a footnote in §27.8.

Q36(9) Using the matrix representation of $\widehat{S}_{(1)x}$, $\widehat{S}_{(1)y}$, $\widehat{S}_{(1)z}$ for a spin-1 particle in Eqs. (14.48), (14.49) and (14.50) show that the matrices for

$$\widehat{S}_{(1)+} = \widehat{S}_{(1)x} + i\widehat{S}_{(1)y}, \quad \widehat{S}_{(1)-} = \widehat{S}_{(1)x} - i\widehat{S}_{(1)y}, \quad \widehat{S}^2_{(1)}, \tag{36.158}$$

are

$$\boldsymbol{M}_{\hat{S}_{(1)+}} = \sqrt{2}\hbar \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \tag{36.159}$$

$$\boldsymbol{M}_{\hat{S}_{(1)-}} = \sqrt{2}\hbar \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \tag{36.160}$$

$$\boldsymbol{M}_{\hat{S}^2_{(1)}} = 2\hbar^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{36.161}$$

Q36(10) Using Eqs. (36.126) and (36.127) verify that $\vec{\alpha}_x$, $\vec{\beta}_x$ defined by Eq. (36.137) are eigenvectors of $\widehat{S}_x$ and $\vec{\alpha}_y$ and $\vec{\beta}_y$ defined by Eq. (36.138) are eigenvectors of $\widehat{S}_y$.

Q36(11) The z-component spin is measured giving a value $-\frac{1}{2}\hbar$. What are the possible outcomes of a measurement of the x-component spin? Find the probabilities of these possible measured outcomes.

Q36(12) Verify that Eqs. (36.146), (36.148) and (36.147) are satisfied by the matrices in Eqs. (36.143), (36.145) and (36.144).

Q36(13) [41]A unit vector $\vec{n}$ in the 3-dimensional $\vec{I\!E}^3$ aligned at an angle θ to the z-axis on the xz plane is given by $\vec{n} = \sin\theta\,\vec{i} + \cos\theta\,\vec{k}$. The spin operator $\widehat{S}_{\vec{n}}$ in the direction of the unit vector $\vec{n}$ is given by $\widehat{S}_{\vec{n}} = \vec{n}\cdot\widehat{S}$, where

$$\vec{n}\cdot\widehat{S} = n_x\widehat{S}_x + n_y\widehat{S}_y + n_z\widehat{S}_z = \sin\theta\,\widehat{S}_x + \cos\theta\,\widehat{S}_z. \tag{36.162}$$

[41] *See* Zettili pp. 298–316 for more examples. $\vec{I\!E}^3$ corresponds to the physical space we live in.

(a) Show that the matrix representation of $\widehat{S}_{\vec{n}}$ in basis $\{\vec{\alpha}_z, \vec{\beta}_z\}$ is

$$\boldsymbol{M}_{\hat{S}_{\vec{n}}} = \frac{1}{2}\hbar \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}, \tag{36.163}$$

and that $\boldsymbol{M}_{\hat{S}_{\vec{n}}}$ admits

$$\boldsymbol{C}_{\vec{\eta}_{\vec{n}+}} = \begin{pmatrix} \cos(\theta/2) \\ \\ \sin(\theta/2) \end{pmatrix}, \quad \boldsymbol{C}_{\vec{\eta}_{\vec{n}-}} = \begin{pmatrix} -\sin(\theta/2) \\ \\ \cos(\theta/2) \end{pmatrix} \tag{36.164}$$

as eigenvectors corresponding eigenvalues $\pm\hbar/2$.

(b) Find matrix representation of the state vector of a spin aligned in the direction $\vec{n}$.[42]

(c) A spin is aligned in the positive direction $\vec{n}$. Find the probabilities of a measurement of spin along the z-axis resulting in the values $\pm\frac{1}{2}\hbar$.

(d) A beam of spin-$\frac{1}{2}$ particles with its spin aligned in the positive direction $\vec{n}$ is fed into a Stern–Gerlach apparatus oriented to measure the component of the spin along the z-axis. The incoming beam will split into two with the upper beam corresponding to spin-up along the z-axis and the lower beam corresponding to spin-down along the z-axis. Find the ratio of the intensities of the emerging beams.

Q36(14) The Pauli matrix σ_y is given by

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

(a) Show that the Pauli matrix σ_y possesses the following properties[43]:

[42] The notation $\boldsymbol{C}_{\vec{\eta}_{\vec{n}}}$ in Eq. (36.164) shows that the spin aligned in the direction $\vec{n}$ should be denoted by $\vec{\eta}_{\vec{n}+}$.

[43] For the exponential function, we can use the expansion

$$e^{c\sigma_y} = \sum_{n=0}^{\infty} \frac{1}{n!} (c\sigma_y)^n. \tag{36.165}$$

$$\sigma_y^0 = \sigma_y^2 = \sigma_y^4 = \cdots = \sigma_y^{2k} = \boldsymbol{I}_{2\times 2}, \tag{36.166}$$

$$\sigma_y = \sigma_y^3 = \sigma_y^5 = \cdots = \sigma_y^{2k+1}, \tag{36.167}$$

$$e^{c\sigma_y} = \sum_{k=0}^{\infty} \frac{1}{(2k)!}\, c^{2k}\, \boldsymbol{I}_{2\times 2} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!}\, c^{2k+1}\, \sigma_y, \tag{36.168}$$

where $\boldsymbol{I}_{2\times 2}$ is the 2×2 identity matrix. Here $k = 0, 1, 2, 3, \ldots$, and $c \in \mathbb{C}$.

(b) Show that[44]

$$e^{-i\frac{1}{2}\theta\sigma_y} = \begin{pmatrix} \cos\frac{1}{2}\theta & -\sin\frac{1}{2}\theta \\ \sin\frac{1}{2}\theta & \cos\frac{1}{2}\theta \end{pmatrix}. \tag{36.170}$$

(c) Let

$$\boldsymbol{U}(\theta) = e^{-\frac{1}{2}i\theta\sigma_y}. \tag{36.171}$$

Show that $\boldsymbol{U}(\theta)$ is unitary and evaluate the unitary transform of $\boldsymbol{C}_{\vec{\alpha}_z}$ in Eq. (36.139) generated by $\boldsymbol{U}(\theta)$. Give an account on the physical meaning of the transformation.

[44] Use the following expansions of $\cos x$ and $\sin x$:

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!}\,(x)^{2k}, \quad \sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\,(x)^{2k+1}. \tag{36.169}$$

Chapter 37

Particles in Static Magnetic Field

37.1 Static Magnetic Fields

37.1.1 Vector Potentials

As discussed in §27.1.2, a static magnetic field $\vec{B}$ is describable in terms of a magnetic vector potential $\vec{A}(\vec{x})$ in that the field can be derived from $\vec{A}$ by $\vec{B} = \nabla \times \vec{A}$, where the curl operation is given in Cartesian coordinates by Eq. (27.33). A uniform magnetic field pointing in the positive z-axis is expressible as $\vec{B} = B_z \vec{k}$, $B_z > 0$. Its magnitude $B = |\vec{B}|$ is equal to B_z. A suitable magnetic vector potential for such a field is given by its components

$$A_x = -\frac{1}{2} yB, \quad A_y = \frac{1}{2} xB, \quad A_z = 0. \tag{37.1}$$

Using the expression for the curl in Eq. (27.33) we can verify that

$$\vec{B} = \nabla \times \vec{A} = B\vec{k}. \tag{37.2}$$

37.1.2 Uniform Field in Cylindrical Coordinates

A uniform magnetic field along the z-axis has cylindrical symmetry. It is easier to employ cylindrical coordinates (r, θ, z). We shall denote

Quantum Mechanics: A Fundamental Approach
K. Kong Wan

ISBN 978-981-4774-65-9 (Hardcover), 978-0-203-70241-3 (eBook)
www.jennystanford.com

the unit vectors in the r, θ and z directions by $\vec{e}_r$, $\vec{e}_\theta$, and $\vec{e}_z$.[1] These unit vectors are related to $\vec{i}$, $\vec{j}$, $\vec{k}$ by

$$\vec{e}_r = \cos\theta\,\vec{i} + \sin\theta\,\vec{j}, \quad \vec{e}_\theta = -\sin\theta\,\vec{i} + \cos\theta\,\vec{j}, \quad \vec{e}_z = \vec{k}. \tag{37.3}$$

Any vector is a linear combination these unit vectors, e.g.,[2]

$$\vec{B} = B_r\vec{e}_r + B_\theta\vec{e}_\theta + B_z\vec{e}_z, \tag{37.4}$$

$$\vec{A} = A_r\vec{e}_r + A_\theta\vec{e}_\theta + A_z\vec{e}_z. \tag{37.5}$$

We can also conveniently denote the above equations by

$$\vec{B} = (B_r,\, B_\theta,\, B_z), \quad \text{and} \quad \vec{A} = (A_r,\, A_\theta,\, A_z). \tag{37.6}$$

For the particular example of a constant magnetic field along the z-direction, we have

$$\vec{B} = B_z\vec{e}_z = B\vec{e}_z \quad \text{with} \quad B_r = B_\theta = 0. \tag{37.7}$$

The components of $\vec{B} = \nabla \times \vec{A}$ in cylindrical coordinates become[3]

$$B_r = \frac{1}{r}\left(\frac{\partial A_z}{\partial\theta} - \frac{\partial(rA_\theta)}{\partial z}\right), \tag{37.8}$$

$$B_\theta = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}, \tag{37.9}$$

$$B_z = \frac{1}{r}\left(\frac{\partial(rA_\theta)}{\partial r} - \frac{\partial A_r}{\partial\theta}\right). \tag{37.10}$$

We can derive a vector potential for the above magnetic field by Stokes' theorem in vector calculus which says that the line integral of a vector field around a closed curve is equal to the surface integral of the curl of the vector field over any surface bounded by the curve.[4] In our present case, we have a vector field in the form of the vector

[1] Spiegel (3) p. 142 (*see* Fig. 4 on p. 138). These unit vectors are position-dependent.
[2] These are vectors in the three-dimensional physical space.
[3] Spiegel (3) pp. 153–154. These expressions are different from those in Cartesian coordinates.
[4] Spiegel (3) p. 106. Gasiorowicz p. 257. Stokes (1819–1903) was an Irish physicist and mathematician.

potential $\vec{A}$. Consider a circle of radius r on the x-y plane centred at the origin. This circle encloses a disc S_r of radius r the x-y plane centred at the origin. Stokes' theorem for our present geometry means[5]

$$\oint \vec{A} \cdot d\vec{\ell} = \int_{S_r} \nabla \times \vec{A} \cdot d\vec{S}. \tag{37.11}$$

We can evaluate the integrals since (1) A_θ of the vector potential should be independent of both θ and z on account of the cylindrical symmetry of the magnetic field, and (2) the field $\vec{B}$ and surface vector $d\vec{S}$ are parallel along the z-direction. In other words, we have

$$\oint \vec{A} \cdot d\vec{\ell} = \oint A_\theta(r) r d\theta = 2\pi r\, A_\theta(r), \tag{37.12}$$

$$\int_S \nabla \times \vec{A} \cdot d\vec{S} = \int_S \vec{B} \cdot d\vec{S} = \pi r^2\, B. \tag{37.13}$$

Stokes' theorem then implies $A_\theta = \frac{1}{2} Br$. Using Eqs. (37.8) to (37.10) we can verify that a vector potential with zero components along the r and z directions and with its θ component given above, i.e.,

$$A_r = 0, \quad A_\theta(r) = \frac{1}{2} Br, \quad A_z = 0 \tag{37.14}$$

would satisfy Eq. (37.7).

Potentials are generally not unique. Two vector potentials $\vec{A}$ and $\vec{A}'$ related by the gradient of a scalar function $f(\vec{x})$, i.e.,[6]

$$\vec{A}' = \vec{A} + \nabla f(\vec{x}) \tag{37.15}$$

define the same magnetic field, since the curl of a gradient is zero.[7]

---

[5] Here $d\vec{\ell}$ is the line element along the circle and $d\vec{S}$ is the surface element of the disc S_r.

[6] From (37.8) to (37.10) we can see that adding a constant vector $\vec{c} = c_r\vec{e}_r + c_\theta\vec{e}_\theta + c_z\vec{e}_z$ to the vector potential in Eq. (37.14) will lead to the same magnetic field. A similar statement is true for the vector potential in Eq. (37.1) in Cartesian coordinates.

[7] Gasiorowicz p. 247. The change from $\vec{A}$ to $\vec{A}'$ is known as a *gauge transformation*.

37.1.3 Field Confined in a Cylindrical Region

Consider a uniform magnetic field of magnitude B aligned along the z-direction but confined within a cylindrical region of radius $r < b$.[8] In cylindrical coordinates, such a field is specified by

$$\vec{B}(r) = \begin{cases} (0, 0, B) & r < b \\ (0, 0, 0) & r > b \end{cases}. \tag{37.16}$$

The corresponding vector potential $\vec{A}$ in the cylindrical region $r < b$ is given by Eq. (37.14). We can obtain the vector potential outside the circle again by Stokes' theorem:

(1) As before we choose the components A_r and A_z of $\vec{A}$ along the r and the z directions to be zero. The line integral in Eq. (37.11) along any circle of radius $r > b$ centred at the origin on the x-y plane has the value of $2\pi r A_\theta(r)$.

(2) The surface integral of $\nabla \times \vec{A} = \vec{B}$ over the area enclosed by the circle reduces to the surface integral over the disc $S_{r=b}$ of radius b since $\vec{B}$ is zero outside the disc, i.e., the integral has a value $\pi b^2 B$.[9] This is equal to the *magnetic flux* Φ_b enclosed by the circle of radius b in the x-y plane centred at the origin. It follows that for $r > b$ the θ-component A_θ is a function of r given by $\Phi_b/2\pi r$.[10]

(3) Combining the above results we get

$$A_\theta(r) = \begin{cases} Br/2 & r < b \\ \Phi_b/2\pi r & r > b \end{cases}. \tag{37.17}$$

This vector potential is proportional to r inside the cylindrical region and is inversely proportional to r outside. It is continuous

[8] An infinitely long solenoid of radius b lying along the z-direction with its centre coinciding with the z-axis with a current flowing through it would trap a magnetic field within the solenoid.

[9] The magnetic field, being zero for $r > b$, makes no contribution to the surface integral in area outside the circle of radius b.

[10] Gasiorowicz pp. 257–258. *See* Gasiorowicz p. 258 for a gauge transformation of this vector potential. We can directly verify that $\nabla \times \vec{A}(r) = \vec{B}$ for $r < b$ and $\nabla \times \vec{A}(r) = 0$ for $r > b$. At $r = b$ we have to treat Eq. (37.10) with care. We get $\nabla \times \vec{A}(r) = \vec{B}$ as the limit from the left, i.e., $\nabla \times \vec{A}(r)$ as $r \to b$ where $r < b$. The limit from the right, using $\vec{A}(r)$ for $r > b$, will give the value 0.

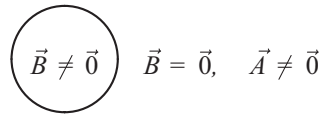


Figure 37.1 Magnetic field confined to a cylindrical region.

across the boundary at $r = b$ with a value $A_\theta(b) = Bb/2$. The above diagram shows the magnetic field and the vector potential on the x-y plane:

In classical electromagnetism, it is the electric and the magnetic fields which are physical and directly measurable; potentials which are generally not even unique are often regarded as mathematical constructs to facilitate calculations. As already discussed in §27.1.2, a classical particle with charged q moving with velocity $\vec{v}$ in a static magnetic field $\vec{B}$ will experience a Lorentz force. This force is zero if $\vec{B} = 0$. In our present case, a charged classical particle moving outside the cylindrical region will not experience any Lorentz force, despite a non-zero and varying magnetic vector potential. It follows that its motion is not affected by the vector potential. In other words, the magnetic field in the region $r < b$ does not affect the motion of the particle outside the region.

The situation is quite different in quantum mechanics. The magnetic field confined inside a cylindrical region can affect the motion of a charged quantum particle moving outside the cylindrical region where the magnetic field is zero. This gives rise to what is known as the *Aharonov–Bohm effect* which we shall discuss in more details in §37.6.[11]

The Aharonov–Bohm effect has its origin in the classical Hamiltonian. In the absence of any mechanical or electrical potential, a classical particle of mass m and charge q in a static magnetic field specified by a magnetic vector potential $\vec{A}$ has a Hamiltonian given by Eq. (27.47). The quantised Hamiltonian is given by Eqs. (27.98), i.e., the following operator in the Hilbert space $L^2(\mathbb{R}^3)$[12]:

$$\widehat{H} := \frac{1}{2m}\left(\widehat{\vec{p}} - q\widehat{\vec{A}}\right)^2. \tag{37.18}$$

[11] Aharonov (1932–) is an Israeli physicist. Bohm (1917–1992) was an American physicist. Historically there had been controversy about such an effect.

[12] For a spinless particle.

The fact that it is the vector potential, not the magnetic field, which appears in the Hamiltonian would suggest that the vector potential would play a significant role in the quantised theory. Details are set out in the following sections.

37.2 Charged Quantum Particles in Uniform Field

We shall start with a spinless particle of mass m and charge q in a static electromagnetic field specified by a vector potential $\vec{A}$. The Hamiltonian of the particle is given in Eq. (37.18). The expression $\left(\hat{\vec{p}} - q\hat{\vec{A}}\right)^2$ is given by Eq. (27.99) in the form of a sum. Each term in the sum can be expanded, e.g., the first term in the above sum can be expanded as follows:

$$\begin{aligned}\left(\hat{p}_x - q\hat{A}_x\right)^2 &= \hat{p}_x^2 - \hat{p}_x\,(q\hat{A}_x) - (q\hat{A}_x)\,\hat{p}_x + q^2\hat{A}_x^2 \\ &= \hat{p}_x^2 + q^2\hat{A}_x^2 - q\left(\hat{p}_x\hat{A}_x + \hat{A}_x\,\hat{p}_x\right). \end{aligned} \tag{37.19}$$

Using the explicit differential expression of $\hat{p}_x$ and replacing the multiplication operator $\hat{A}_x$ by the function A_x we can carry out the following calculation[13]:

$$\begin{aligned}\hat{p}_x\hat{A}_x &:= -i\hbar\frac{\partial}{\partial x}\,A_x = -i\hbar\left(\frac{\partial A_x}{\partial x} + A_x\frac{\partial}{\partial x}\right) \\ &= -i\hbar\frac{\partial A_x}{\partial x} + A_x\hat{p}_x. \end{aligned} \tag{37.20}$$

Writing the multiplication operator $\hat{A}_x$ explicitly as a function A_x we can to express $\left(\hat{p}_x - q\hat{A}_x\right)^2$ as

$$\left(\hat{p}_x - q\,A_x\right)^2 = \hat{p}_x^2 + q^2A_x^2 - 2q\,A_x\,\hat{p}_x + i\hbar q\,\frac{\partial A_x}{\partial x}. \tag{37.21}$$

Similar expressions for $\left(\hat{p}_y - q\,\hat{A}_y\right)^2$ and $\left(\hat{p}_z - q\,\hat{A}_z\right)^2$ can be obtained. Adding these terms we arrive at

$$\left(\hat{\vec{p}} - q\,\vec{A}\right)^2 = \hat{\vec{p}}^{\,2} + q^2\,\vec{A}^{\,2} - 2q\,\vec{A}\cdot\hat{\vec{p}} + i\hbar q\,\nabla\cdot\vec{A}, \tag{37.22}$$

[13] *See* Eqs. (17.80), (20.70) and (27.107).

where

$$\vec{A} \cdot \hat{\vec{p}} := A_x \hat{p}_x + A_y \hat{p}_y + A_z \hat{p}_z, \tag{37.23}$$

$$\nabla \cdot \vec{A} := \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}. \tag{37.24}$$

The Hamiltonian can now be written as

$$\hat{H} = \frac{1}{2m} \hat{\vec{p}}^2 + \frac{1}{2m} q^2 \vec{A}^2 - \frac{q}{m} \vec{A} \cdot \hat{\vec{p}} + \frac{i\hbar q}{2m} \nabla \cdot \vec{A}. \tag{37.25}$$

For the case where the magnetic field is uniform pointing in the z-direction, i.e., $\vec{B} = B\,\vec{k}$ in Cartesian coordinates, the vector potential is given by Eq. (37.1). This vector potential has the following properties:

$$\nabla \cdot \vec{A} = 0, \quad \vec{A}^2 = \frac{1}{4} B^2 \left(x^2 + y^2\right), \quad \vec{A} \cdot \hat{\vec{p}} = \frac{1}{2} B \hat{L}_{cz}. \tag{37.26}$$

where $\hat{L}_{cz} = \hat{y}\,\hat{p}_x - \hat{x}\,\hat{p}_y$ is the z-component orbital angular momentum operator.[14] The Hamiltonian becomes

$$\hat{H} = \frac{1}{2m} \hat{\vec{p}}^2 + \frac{q^2 B^2}{8m} \left(\hat{x}^2 + \hat{y}^2\right) - \frac{q}{2m} B \hat{L}_{cz}. \tag{37.27}$$

In many practical applications, the magnetic field is rather weak so that we can ignore the term containing B^2. This results in a simplified, albeit approximate, Hamiltonian. In the presence of an additional electric potential $\hat{V}$, we have a weak-field Hamiltonian

$$\hat{H}_w = \frac{1}{2m} \hat{\vec{p}}^2 + q\hat{V} - \frac{q}{2m} B \hat{L}_{cz}. \tag{37.28}$$

This Hamiltonian can be interpreted as consisting of a kinetic energy term, an electric energy term and a term containing the magnetic field. This is the magnetic energy term. A study of this term is set out in the following section.

[14] Following the notation in §36.1.2 the operator $\hat{L}_{cz}$ act in $\vec{L}^2(\mathbb{R}^3)$ while $\hat{L}_z(\mathcal{S}_u)$ acts in $\vec{L}^2(\mathcal{S}_u)$.

37.3 Magnetic Moment and Magnetic Energy

37.3.1 For Circular and Orbital Motion

Classically an electric current of magnitude I going anticlockwise round a circular coil of radius r about the origin in the x-y plane behaves like a magnet when it interacts with an external magnetic field. The magnetic property of the current carrying coil manifests itself in the form of a magnetic moment having a value

$$M_z := (\text{area enclosed by the coil}) \times (\text{current}) = \pi r^2 I. \tag{37.29}$$

A magnetic moment is a vector quantity. For the above current carrying coil, it is defined to be directed along the positive z-axis, i.e., the magnetic moment vector is

$$\vec{M} := M_z \vec{k}, \qquad M_z = \pi r^2 I. \tag{37.30}$$

Next consider a particle with charge q constrained to move in a circular orbit of radius r about the origin in the x-y plane with speed v. It would travel a distance equal to v meters per second along the circumference. This is equivalent to going round the circle $n = v/2\pi r$ times. It follows that at any chosen point on the circle there will be an amount of charge nq passing through per second. By definition this is equivalent to a current of magnitude $I = nq$ flowing round the circle, i.e., we have

$$I = nq = \frac{v}{2\pi r}\, q = q\, \frac{mvr}{2\pi mr^2} = q\, \frac{L_z}{2\pi mr^2}, \tag{37.31}$$

where $L_z := mvr$ can be identified with the magnitude of the angular momentum of the particle. The corresponding angular momentum vector is

$$\vec{L} := L_z \vec{k}, \qquad L_z = mvr. \tag{37.32}$$

The current I in Eq. (37.31) gives rise to a *magnetic moment*

$$\vec{M}_z^{(o)} := M_z^{(o)} \vec{k} \quad \text{where} \quad M_z^{(o)} := \pi r^2 I = \frac{q}{2m} L_z. \tag{37.33}$$

The magnetic moment is seen to be related to the angular momentum by

$$\vec{M}^{(o)} = \frac{q}{2m}\,\vec{L}. \tag{37.34}$$

The superscript signifies the origin of the magnetic moment as due to the orbital motion of the charged particle. If the plane of the charged particle's orbit does not lie in the x-y plane, its angular momentum $\vec{L}$ would be directed along the normal the plane.

The presence of a uniform magnetic field $\vec{B}$ will affect the particle's orbital motion in that the magnetic moment arising from the particle's orbital motion will tend to align itself along the direction of the magnetic field. This gives rise to a potential energy. According to classical electromagnetism this *magnetic potential energy* is given by[15]

$$-\vec{B}\cdot\vec{M}^{(o)} = -\frac{q}{2m}\,\vec{B}\cdot\vec{L}. \tag{37.35}$$

For a uniform field in the z-direction, this expression reduces to

$$-B\,M_z^{(o)} = -\frac{q}{2m}\,B\,L_z. \tag{37.36}$$

In accordance with Postulate 27.2(CQ), a quantum particle in circular motion would possess a **magnetic moment operator**[16]

$$\hat{\vec{M}}^{(o)} = \frac{q}{2m}\,\hat{\vec{L}}_c, \tag{37.37}$$

and a **magnetic energy operator**

$$-\vec{B}\cdot\hat{\vec{M}}^{(o)} = -\frac{q}{2m}\,\vec{B}\cdot\hat{\vec{L}}_c. \tag{37.38}$$

For a magnetic field aligned along the z-axis, Eq. (37.37) becomes

$$\hat{M}_z^{(o)} = \frac{q}{2m}\,\hat{L}_{cz}, \tag{37.39}$$

[15] Jackson p. 190. The maximum (minimum) potential energy is achieved when the magnetic moment is anti-parallel (parallel) to the magnetic field. This explains the negative sign in the expression.

[16] Treating $\vec{L}$ as the classical canonical angular momentum it is quantised as the operator $\hat{\vec{L}}_c$ acting in $\vec{L}^2(I\!R^3)$.

and Eq. (37.38) becomes

$$-B\,\widehat{M}_z^{(o)} = -\frac{q}{2m}\,B\,\widehat{L}_{cz}. \tag{37.40}$$

which is the last term in the Hamiltonian $\widehat{H}_w$ in Eq. (37.28).

The above results apply to the electron in circular motion in a hydrogen atom, i.e., for the electron in a hydrogen atom its z-component magnetic moment operator is given by[17]

$$\widehat{M}_z^{(o)} = -\frac{e}{2m}\,\widehat{L}_{cz}. \tag{37.41}$$

The eigenvalues of this operator are

$$-m_\ell \mu_B, \qquad \mu_B = \frac{e\hbar}{2m}, \tag{37.42}$$

where m_ℓ is the orbital magnetic quantum number. The quantity μ_B is known as the **Bohr magneton**. This is the *quantum unit of magnetic moment* in that the values of the quantum magnetic moment are either zero or multiples, both positive and negative, of the Bohr magneton. This statement also applies to spin.

37.3.2 For Spin Motion

It can be established, based on experimental confirmation and theoretical consideration of relativistic quantum mechanics,[18] that an electron also behaves like a magnet when it interacts with an external magnetic field even when it is not in orbital motion. The intuition is that its spinning motion gives rise to a magnetic moment. Under the assumption that the values of the magnetic moment due to spin should also be positive or negative multiple of the Bohr magneton μ_B the spin magnetic momentum operator along the z-direction must be[19]

[17] Wan pp. 486–488. Greiner pp. 161–162. The result comes from Eq. (37.39) with q replaced by the charge of the electron which is $-e$, where e is the (positive) elementary charge.

[18] Dirac pp. 263–267. Roman pp. 131–133.

[19] Goudsmith and Uhlenbeck postulated in 1925 that the magnetic moment arising from electron spin had a magnitude of one Bohr magneton. When compared with Eq. (37.41), we note that there is the absence of the fact $1/2$ in the expression so that the eigenvalues of $\widehat{M}_z^{(s)}$ are $\pm\mu_B$, not $\pm\mu_B/2$.

$$\widehat{M}_z^{(s)} = -\frac{e}{m}\,\widehat{S}_z, \tag{37.43}$$

where $\widehat{S}_z$ is the z-component spin operator and the superscript signifies the origin of the magnetic moment as due to the spin. This operator has eigenvalues $\pm\,\mu_B$.

In a uniform magnetic field $\vec{B} = (0,\, 0,\, B)$, this magnetic moment gives rise to a magnetic potential energy operator

$$\frac{e}{m}\,B\widehat{S}_z. \tag{37.44}$$

This operator possesses eigenvalues

$$E_\pm = \pm\,\frac{1}{2}\,\hbar\omega, \quad \omega = \frac{eB}{m}, \tag{37.45}$$

corresponding to eigenvectors $\vec{\alpha}_z$ and $\vec{\beta}_z$.

For a spin-$\frac{1}{2}$ particle of charge q, Eqs. (37.43) and (37.44) become

$$\widehat{M}_z^{(s)} = \frac{q}{m}\,\widehat{S}_z \quad \text{and} \quad -\frac{q}{m}\,B\widehat{S}_z. \tag{37.46}$$

37.4 Pauli–Schrödinger Equation

For a spin-$\frac{1}{2}$ particle, the Hamiltonian $\widehat{H}_w$ in Eq. (37.28) must be amended to take account of the magnetic energy due to spin. Intuitively we may write down the new Hamiltonian as $\widehat{H}_w - (q/m)\,B\widehat{S}_z$. Mathematically this sum is not well-defined since we are adding operators of two different Hilbert spaces, i.e., the operator $\widehat{H}_w$ acts in $\vec{L}^2(I\!R^3)$ while $\widehat{S}_z$ acts on $\vec{\mathbb{W}}^{(2)}$.

What we should do is to take the state space to be $\vec{\mathcal{H}}^{(s)}(I\!R^3)$ in Eq. (33.10). A state vector $\vec{\Phi}$ is of the form given in Eq. (36.108). Operators act on $\vec{\mathcal{H}}^{(s)}(I\!R^3)$. In accordance with the discussion in §36.3.4 and §36.3.5, spatial operator $\widehat{H}_w$ and the spin operator and $(-q/m)\,B\widehat{S}_z$ should take the form

$$\widehat{H}_w^{(s)} = \widehat{H}_w \otimes \widehat{I\!I}\left(\vec{\mathbb{W}}^{(2)}\right) \quad \text{and} \quad \widehat{I\!I}(I\!R^3) \otimes \left(-\frac{qB}{m}\right)\widehat{S}_z. \tag{37.47}$$

We can now combine these two operators to obtain a Hamiltonian operator $\widehat{H}_w^{(s)}$ acting in $\vec{\mathcal{H}}^{(s)}(I\!R^3)$ to be

$$\widehat{H}_w^{(s)} = \widehat{H}_w \otimes \widehat{\mathbb{I}}\left(\vec{\mathbb{V}}^{(2)}\right) - \frac{qB}{m}\ \widehat{\mathbb{I}}(I\!R^3) \otimes \widehat{S}_z. \tag{37.48}$$

This Hamiltonian is known as the *Pauli–Schrödinger Hamiltonian* for a charged spin-$\frac{1}{2}$ particle in uniform static electric and magnetic fields in the weak field approximation. In the Schrödinger picture, this Hamiltonian will govern the time evolution of the state vector $\vec{\Phi}$ with the following equation of motion:

$$i\hbar \frac{d\vec{\Phi}(t)}{dt} = \widehat{H}_w^{(s)} \vec{\Phi}(t). \tag{37.49}$$

This is known as the **Pauli–Schrödinger equation** for a charged spin-$\frac{1}{2}$ particle in uniform static electric and magnetic fields in the weak field approximation. This may be regarded as a two-component equation corresponding to a two-component wave function introduced in §36.3.2.

In many applications in magnetism and in solid-state physics, we are interested in the effect arising from the electron's spin, not in the electron's spatial motion. So, the spatial part of the wave function is often ignored. In other words, we will work in the state space $\vec{\mathbb{V}}^{(2)}$ of the spin with state vector given by a spin vector $\vec{\eta}(t)$. The Pauli–Schrödinger equation becomes

$$i\hbar \frac{d}{dt}\vec{\eta}(t) = \frac{eB}{m}\, \widehat{S}_z\, \vec{\eta}(t). \tag{37.50}$$

An initial state vector $\vec{\alpha}_z$ will evolve into $\exp(-i\omega t/2)\,\vec{\alpha}_z$, and an initial state vector $\vec{\beta}_z$ will evolve into $\exp(i\omega t/2)\,\vec{\beta}_z$.[20] An initial state vector

$$\vec{\eta}(0) = \frac{1}{\sqrt{2}}\left(\vec{\alpha}_z + \vec{\beta}_z\right) \tag{37.51}$$

will evolve to the state vector

$$\vec{\eta}(t) = \frac{1}{\sqrt{2}}\left(e^{-\frac{1}{2}i\omega t}\vec{\alpha}_z + e^{\frac{1}{2}i\omega t}\vec{\beta}_z\right). \tag{37.52}$$

[20] The constant ω is given by Eq. (37.45), i.e., $\omega = eB/m$.

37.5 The Simple Zeeman Effect

In 1895 Zeeman discovered that in the presence of a uniform static magnetic field there was a splitting of spectral lines in the radiation emitted by atoms and molecules.[21] This is known as Zeeman effect. This is attributed to the splitting of the atomic energy levels. We can now show that this splitting of the atomic energy levels is due to the additional magnetic energy arising from the orbital and spinning motion of the electron. The number of splittings as well as the magnitudes of these splits for the hydrogen atom can be calculated easily in the weak field approximation.

37.5.1 The Hydrogen Atom

In the absence of any magnetic field, the hydrogen atom has the following familiar Hamiltonian:

$$\widehat{H}_h := -\frac{\hbar^2}{2m}\,\nabla^2 - \frac{e^2}{4\pi\epsilon_0\, r}. \tag{37.53}$$

This Hamiltonian has a well-known set of degenerate eigenvalues E_n. The corresponding eigenfunctions are of the form of a product of a function $R_{n\ell}(r)$ of the radial variable r and the spherical harmonics $Y_{\ell,m_\ell}(\theta, \varphi)$, i.e.,[22]

$$\psi_{n\ell m_\ell}(r, \theta, \varphi) = R_{n\ell}(r) Y_{\ell,m_\ell}(\theta, \varphi), \tag{37.54}$$

where n is the *principal quantum number* which fixes the energy eigenvalue, ℓ is the orbital angular momentum quantum number, and m_ℓ is the orbital magnetic quantum number. We have the following eigenvalue equations:

$$\widehat{H}_h \vec{\psi}_{n\ell m_\ell} = E_n \vec{\psi}_{n\ell m_\ell}, \quad \widehat{L}_c^2 \vec{\psi}_{n\ell m_\ell} = \ell(\ell+1)\hbar^2 \vec{\psi}_{n\ell m_\ell}, \tag{37.55}$$

$$\widehat{L}_{cz} \vec{\psi}_{n\ell m_\ell} = m_\ell \hbar\, \vec{\psi}_{n\ell m_\ell}, \quad \widehat{M}_z^{(o)} \vec{\psi}_{n\ell m_\ell} = -\frac{e}{2m}\, m_\ell \hbar\, \vec{\psi}_{n\ell m_\ell}. \tag{37.56}$$

[21] Zeeman (1865–1943) was a Dutch physicist.
[22] Gasiorowicz pp. 132–140.

The quantum numbers are related by

$$n = 1, 2, 3, \ldots, \quad \ell \leq n - 1; \tag{37.57}$$

$$m_\ell = 0, \pm 1, \pm 2, \ldots \text{with} \quad -\ell \leq m_\ell \leq \ell. \tag{37.58}$$

Since m_ℓ ranges from $-\ell$ to ℓ there are $2\ell + 1$ different values of m_ℓ for any given ℓ. In the derivation of the above results, the electron's spin plays no part. As seen in the next section, the spin will play a part when an external magnetic field is applied since spin with its magnetic moment will interact with external magnetic field.[23]

37.5.2 Hydrogen Atom in Magnetic Field

In the presence of a weak and uniform static magnetic field of magnitude B along the z-direction, the electron will acquire additional energy terms due to its orbital and spin magnetic moments. The Pauli–Schrödinger Hamiltonian in Eq. (37.48) applies. The unperturbed Hamiltonian $\widehat{H}_h$ should be replaced by

$$\widehat{H}^{(s)}_{hw} = \left(\widehat{H}_h + \frac{eB}{2m}\,\widehat{L}_{cz}\right) \otimes \widehat{I\!I}(\vec{I\!R}^{(2)}) + \widehat{I\!I}(I\!R^3) \otimes \left(\frac{eB}{m}\,\widehat{S}_z\right). \tag{37.59}$$

The eigenvectors of $\widehat{H}^{(s)}_{hw}$ are easily verified to be

$$\vec{\Psi}_{n\ell m_\ell +} = \vec{\psi}_{n\ell m_\ell} \otimes \vec{\alpha}_z \quad \text{or} \quad \vec{\Psi}_{n\ell m_\ell -} = \vec{\psi}_{n\ell m_\ell} \otimes \vec{\beta}_z. \tag{37.60}$$

For the spin-up state $\vec{\Psi}_{n\ell m_\ell +}$,

$$\begin{aligned} \widehat{H}^{(s)}_{hw}\,\vec{\Psi}_{n\ell m_\ell +} &= \left(\widehat{H}_h\,\vec{\psi}_{n\ell m_\ell} + \frac{eB}{2m}\widehat{L}_z\,\vec{\psi}_{n\ell m_\ell}\right) \otimes \vec{\alpha}_z \\ &\quad + \vec{\psi}_{n\ell m_\ell} \otimes \left(\frac{eB}{m}\,\widehat{S}_z\,\vec{\alpha}_z\right) \\ &= \left(E_n + \frac{eB}{2m}\,m_\ell\hbar + \frac{eB\hbar}{2m}\right)\vec{\psi}_{n\ell m_\ell} \otimes \vec{\alpha}_z \qquad (37.61) \\ \Rightarrow \quad \widehat{H}^{(s)}_{hw}\,\vec{\Psi}_{n\ell m_\ell +} &= E_{nm_\ell +}\,\vec{\Psi}_{n\ell m_\ell +}, \\ E_{nm_\ell +} &= E_n + \frac{e\hbar}{2m}\,B\,(m_\ell + 1). \qquad (37.62) \end{aligned}$$

[23] We have not taken the spin–orbit coupling into account.

Similarly we have, for the spin-down state $\vec{\Phi}_{n\ell m_\ell -}$,

$$\widehat{H}_{hw}\vec{\Psi}_{n\ell m_\ell -} = E_{nm_\ell -}\,\vec{\Psi}_{n\ell m_\ell -}, \tag{37.63}$$

and

$$E_{nm_\ell -} = E_n + \frac{e\hbar}{2m}\,B\,(m_\ell - 1). \tag{37.64}$$

Compared the eigenvalues $E_{nm_\ell +}$ and $E_{nm_\ell -}$ with the unperturbed eigenvalues E_n we can see that each energy level E_n is split into several sub-levels according to the z-component orbital and spin angular momenta. The following examples serve to illustrate the situation:

E37.5.2(1) For the ground state level $n = 1$, we have $\ell = 0$ and hence $m_\ell = 0$. There is no contribution to the energy due to orbital magnetic moment. The ground state level is shifted down for the spin-up state and up by the spin-down state. The energy gap between in this splitting is

$$\Delta^{(s)}E_1 = \frac{e\hbar}{m}\,B. \tag{37.65}$$

the spin gives rise to two sub-levels with energy gap $\Delta^{(s)}E_1$.

E37.5.2(2) For the first excited level $n = 2$, we can have $\ell = 1$ and $m_\ell = -1, 0, 1$. The first excited energy level is split into three sub-levels due to orbital motion for the spin-up state as well as for the spin down state, e.g., for the spin-up state, we have three levels

$$E_{2(-1)+} = E_2, \tag{37.66}$$

$$E_{20+} = E_2 + \frac{e\hbar}{2m}\,B, \tag{37.67}$$

$$E_{21+} = E_2 + \frac{e\hbar}{m}\,B. \tag{37.68}$$

The energy gap is

$$\Delta^{(o)}E_2 = \frac{e\hbar}{2m}\,B. \tag{37.69}$$

Generally for the spin-up (spin-down) state any given ℓ gives rise to $2\ell + 1$ sub-levels with energy gap $\Delta^{(o)}E_n$.

These results agree well with experimental observations.

37.6 Aharonov–Bohm Effect

37.6.1 Circular Motion

A quantum particle of mass m and charge q constrained to move freely in a circle $\mathcal{C}_a$ of radius a centred at the origin in the x-y plane is discussed in §27.8.[24] The state space of the system is the space $\vec{L}^2(\mathcal{C}_a)$ of square-integrable functions on the circle introduced in §16.1.2. These functions satisfy the periodic boundary condition in Eq. (16.39) with scalar product defined by Eq. (16.40). The particle's momentum is represented by the operator $\widehat{p}(\mathcal{C}_a)$ in Eq. (27.116). The eigenvectors $\vec{\varphi}_n(\mathcal{C}_a)$ and eigenvalues $p_n(\mathcal{C}_a)$ of this operator are given by Eqs. (19.36) and (19.37).

For free motion along the circle, the Hamiltonian is

$$\widehat{H}(\mathcal{C}_a) := \frac{1}{2m}\,\widehat{p}^{\,2}(\mathcal{C}_a). \tag{37.70}$$

The operator has the following operator expression:

$$\widehat{H}(\mathcal{C}_a) := -\frac{\hbar^2}{2ma^2}\,\frac{d^2}{d\theta^2}. \tag{37.71}$$

This Hamiltonian shares the same set of eigenvectors with the momentum operator, i.e.,

$$\widehat{H}(\mathcal{C}_a)\vec{\varphi}_n(\mathcal{C}_a) = E_n\,\vec{\varphi}_n(\mathcal{C}_a) \tag{37.72}$$

with eigenvalues

$$E_n = \frac{1}{2m}\,p_n^2 = \frac{1}{2ma^2}\,(\hbar n)^2. \tag{37.73}$$

37.6.2 The Aharonov–Bohm Effect

A constant magnetic field of magnitude B directed along the z-axis is confined to a cylindrical region of radius b with its centre coinciding with the z-axis. A quantum particle of mass m and charge q is

[24] Wan pp. 480–485. Martin pp. 46–47. Physically this is achieved by confinement of the particle to a thin torus (*see* Gasiorowicz p. 259 and Supplement 16-B).

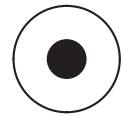


Figure 37.2 Field confined to a cylindrical region of radius $b < a$.

constrained to move along a circular orbit of radius $a > b$ centred at the origin in the x-y plane in the presence of the magnetic field. In Fig. 37.2, the outer circle is the particle's circular orbit of radius a on the x-y plane. The dark circular area is of radius $b < a$. This is the region where the magnetic field is confined so that outside the dark area the magnetic field is zero.

The field and its magnetic vector potential are given in cylindrical coordinates by Eqs. (37.16) and (37.17). The Hamiltonian of the particle is given, in accordance with Eq. (27.98), by

$$\widehat{H} := \frac{1}{2m}\left(\widehat{p}(\mathcal{C}_a) - q\,A_\theta(a)\right)^2, \quad A_\theta(a) = \frac{\Phi_b}{2\pi a} \tag{37.74}$$

This Hamiltonian admits $\vec{\varphi}_n(\mathcal{C}_a)$ in Eqs. (19.37) as eigenvectors corresponding eigenvalues

$$E_n = \frac{1}{2ma^2}\left(\hbar n - \frac{q\Phi_b}{2\pi}\right)^2. \tag{37.75}$$

We can check this by noting that

$$\left(\widehat{p}(\mathcal{C}_a) - q\,A_\theta(a)\right)\vec{\varphi}_n(\mathcal{C}_a) = \frac{1}{a}\left(\hbar\, n - \frac{q\Phi_b}{2\pi}\right)\vec{\varphi}_n(\mathcal{C}_a). \tag{37.76}$$

The magnetic field is zero along the circular orbit of the particle, despite the non-zero value of the vector potential. In classical electromagnetism, a charged particle's motion is affected by the field and not by the potential. In other words, the particle's motion and its energy outside the cylindrical region are independent of the existence of the magnetic flux within the cylindrical region. However, the quantised energy eigenvalues E_n obtained above show that

> *the energy of the quantum particle depends on the magnetic flux Φ_b generated by a magnetic field B which is zero in the region where the particle moves.*

Experiments measuring electron interference through double slits confirm the fact that a magnetic field confined to a region in which electrons are apparently not present can indeed have an effect on the interference pattern.[25] The phenomenon that a charged quantum particle is affected by a magnetic field even though it moves a region in which the magnetic field is zero is a quantum effect known as the **Aharonov–Bohm effect**.

Exercises and Problems

Q37(1) Verify Eq. (37.2).

Q37(2) Using Eqs. (37.8) to (37.10), show that the magnetic field in Eq. (37.7) is derivable from the vector potential in Eq. (37.14) and that the magnetic field in Eq. (37.16) is derivable from the vector potential in Eq. (37.17).

Q37(3) When spatial motion is neglected, the Hamiltonian of an electron of charge $-e$ and mass m in a uniform and static magnetic field of magnitude B pointing along the z-direction is given by

$$\widehat{H}^{(s)} = \frac{e}{m} B \widehat{S}_z. \tag{37.77}$$

Write down the Pauli–Schrödinger equation for the evolution of a spin state in the Schrödinger picture. Show that the following initial spin state

$$\vec{\eta}(0) = \frac{1}{\sqrt{2}} \left\{ \vec{\alpha}_z + \vec{\beta}_z \right\} \tag{37.78}$$

will evolve to a new state $\vec{\eta}(t)$ at time t given

$$\vec{\eta}(t) = \frac{1}{\sqrt{2}} \left\{ e^{-\frac{1}{2} i\omega t} \vec{\alpha}_z + e^{\frac{1}{2} i\omega t} \vec{\beta}_z \right\}, \tag{37.79}$$

where $\omega = eB/m$. What are the spin orientations initially at $t = 0$ and later at $t = \pi/2\omega$?

[25] Gasiorowicz p. 259 and Supplement 16-B.

References

Aczel, A. (2002). *Entanglement: The Greatest Mystery in Physics*, Wiley, New York.

Akhiezer, N. I., and Glazman, I. M. (1963). *Theory of Linear Operators in Hilbert Space Vol. 1*, Frederick Ungar, New York.

Akhiezer, N. I., and Glazman, I. M. (1963). *Theory of Linear Operators in Hilbert Space Vol. 2*, Frederick Ungar, New York.

Amrein, W. O., Jauch, J. M., and Sinha, K. B. (1977). *Scattering Theory in Quantum Mechanics*, W. B. Benjamin, Reading, Mass.

Auletta, G., Fortunato, M., and Parisi, G. (2009). *Quantum Mechanics*, Cambridge University Press.

Bacry, H. (1988). Localizability and Space in Quantum Physics, in (*Lecture Notes in Physics*), Springer-Verlag, Berlin.

Ballentine, L. E. (1990). *Quantum Mechanics*, Prentice-Hall, Eaglewood Cliffs, New Jersey.

Baym, G. (1969). *Lectures on Quantum Mechanics*, W. A. Benjamin, New York.

Beltrametti, E. G., and Cassinelli, G. (1981). *The Logic of Quantum Mechanics*, Addison-Wesley, Reading, Mass.

Berkley, A., Xu, H., Ramos, R., Gubrud, M., Strouch, F., Johnson, P., Anderson, J., Dragt, A., Lobb, C., and Wellstood, F. (2003). *Science*, **300**, June, 1548.

Blank, J., Exner, P., and Havliček, M. (1994). *Hilbert Space Operators in Quantum Physics*, American Institute of Physics Press, New York.

Blum, K. (1981). *Density Matrix Theory and Applications*, Plenum, New York.

Bub, J. (1988). *Found. Phys.*, **18**, 701.

Bub, J. (1997). *Interpreting the Quantum World*, Cambridge University Press, Cambridge.

Busch, P., Lahti, P., and Mittelstaedt, P. (1996). *The Quantum Theory of Measurement*, 2nd edition, Springer-Verlag, Berlin.

Byron, F. W., and Fuller, R. W. (1969). *Mathematics of Classical and Quantum Physics Vol. 1*, Addison-Wesley, Reading, Mass.

Capiński, M., and Kopp, E. (2002). *Measure, Integral and Probability*, Springer-Verlag, London.

Cini, M., and Levy-Leblond, J. M. (1990). *Quantum Theory without Reduction*, Adam Hilger, Bristol.

Cohen, D. W. (1989). *An Introduction to Hilbert Space and Quantum Logic*, Springer-Verlag.

Conway, J. B. (1990). *A Course in Functional Analysis*, 2nd edition, Springer-Verlag, New York.

Crawford, F. S., Jr. (1968). *Waves: Berkeley Physics Course Vol. 3*, MaGraw-Hill, New York.

Croft, A., and Davidson, R. (2004). *Mathematics for Engineers*, Pearson Prentice-Hall, Harlow, England.

de Broglie, L. (1951). *J. Phys. Radium*, **20**, 936.

de Lange, O. L., and Raab, R. E. (1991). *Operator Methods in Quantum Mechanics*, Clarendon Press, Oxford.

D'Espagnat, B. (1999). *Conceptual Foundations of Quantum Mechanics*, 2nd edition, Advanced Books Classics, Perseus Books.

Dicke, R. H., and Wittke, J. P. (1963). *Introduction of Quantum Mechanics*, Addison and Wesley, Reading, Mass.

Dirac, P. A. M. (1958). *The Principles of Quantum Mechanics*, Clarendon Press, Oxford.

Einstein, A., Podolsky, B., and Rosen, N. (1935). *Phys. Rev.*, **47**, 277.

Fano, G. (1992). *Mathematical Methods of Quantum Mechanics*, McGraw-Hill, New York.

Feynman, R. P. (1965). *The Character of Physical Laws*, British Broadcasting Corporation, London. Also published by Penguin Books Ltd, UK (2006) as a paperback.

Feynman, R. P., Leighton, R. B., and Sands, M. (1965). *The Feynman Lectures on Physics Vol. 3*, Addison-Wesley, Reading, Mass.

Finkbeiner II, D. T. (1979). *Introduction to Matrices and Linear Transformations*, Freeman, San Francisco.

Friedman, B. (1956). *Principles and Techniques of Applied Mathematics*, Wiley, New York.

Gallone, F. (2015). *Hilbert Space and Quantum Mechanics*, World Scientific, Singapore.

Gasiorowicz, S. (2003). *Quantum Physics*, Wiley International Edition.

Goldstein, H. (1950). *Classical Mechanics*, Addison-Wesley, Reading, Mass.

Greiner, W. (1989). *Quantum Mechanics Vol. 1: An Introduction*, Springer-Verlag, Berlin.

Hall, B. C. (2013). *Quantum Mechanics for Mathematicians*, Springer Graduate Texts in Mathematics (eBook), New York.

Halliday, D., Resnick, R. and Walker, J. (2008). *Fundamentals of Physics*, 8th edition, John Wiley and Sons, NJ.

Halmos, P. R. (1958). *Finite-Dimensional Vector Spaces*, Van Nostrand, Princeton.

Hey, T., and Walters, P. (1987). *The Quantum Universe*, Cambridge University Press, Cambridge.

Hohn, F. E. (1964). *Elementary Matrix Algebra*, MacMillan, New York.

Isham, C. J. (1995). *Lectures on Quantum Theory*, Imperial College Press, London.

Jackson, J. D. (1999). *Classical Electrodynamics*, 3rd edition, Wiley and Sons, New York.

Jauch, J. M. (1968). *Foundations of Quantum Mechanics*, Addison-Wesley, Reading, Mass.

Johnson, P. R., Strauch, F. W., Gragt, A. J., Ramos, R. C., Lobb, C. J., Anderson, J. R., and Wellstood, F. C. (2003). *Phys. Rev. B*, **67**, 020509-1.

Jordan, T. F. (1969). *Linear Operators for Quantum Mechanics*, Thomas F. Jordan, Duluth, Minnesota, USA.

Kadison, R. V., and Ringrose, J. R. (1983). *Fundamentals of the Theory of Operator Algebras Vol. 1: Elementary Theory*, Academic Press, New York.

Kato, T. (1966). *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin.

Kingman, J. F. C., and Taylor, S. J. (1966). *Introduction to Measure and Probability*, Cambridge University Press, London.

Kraus, K. (1977). Position Observables of the Photon, in *The Uncertainty Principle and the Foundations of Quantum Mechanics*, W. C. Price and S. S. Chissick, eds., Wiley, New York.

Kreyszig, E. (2011). *Advanced Engineering Mathematics*, 10th edition, Wiley, Hoboken, NJ.

Landau, L. D., and Lifshitz, E. M. (2002). *Quantum Mechanics*, 3rd edition, Butterworth-Heinemann.

Lipschutz, S. (1974). *Probability*, Schaum's Outline Series, MaGraw-Hill.

MacKey, G. W. (1963). *The Mathematical Foundations of Quantum Mechanics*, Benjamin, Reading, Mass.

Martin, J. L. (1981). *Basic Quantum Mechanics*, Clarendon Press, Oxford.

Mees, C. R. K. (1963). *The Theory of Photographic Process*, Macmillan, New York.

Merzbacher, E. (1998). *Quantum Mechanics*, 3rd edition, Wiley, New York.

Messiah, A. (l967). *Quantum Mechanics Vol. 1*, North-Holland, Amsterdam.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*, Siam, Philadelphia.

Moretti, V. (2013). *Spectral Theory and Quantum Mechanics*, Springer, Milano.

Naimark, M. A. (1968). *Linear Differential Operators Part 1*, translated by E. R. Dawson, George G. Harrap and Company, London.

Naimark, M. A. (1968). *Linear Differential Operators Part 2*, translated by E. R. Dawson, George G. Harrap and Company, London.

Ohanian, H. C. (1995). *Modern Physics*, 2nd edition, Prentice-Hall, Englewood Cilffe, N.J.

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*, MaGraw-Hill.

Parzen, E. (1960). *Modern Probability Theory and Its Applications*, Wiley, New York.

Pauli, W. (1973). *Pauli Lectures on Physics Vol. 5: Wave Mechanics*, the MIT Press, Cambridge, Mass., edited by C. P. Enz and translated by H. R. Lewis and S. Margulies.

Penrose, O. (1970). *Foundations of Statistical Mechanics*, Pergamon Press, Oxford.

Peres, A. (1995). *Quantum Theory: Concepts and Methods*, Kluwer Academic Publishers, Dordrecht.

Phillips, A. C. (2003). *Introduction to Quantum Mechanics*, Wiley, New York.

Phillips, E. R. (1971). *An Introduction to Analysis and Integration Theory*, Intext Educational Publishers, Scranton, Pennsylvania.

Pitt, H. R. (1963). *Integration, Measure and Probability*, Oliver & Boyd, Edinburgh.

Prugovečki, E. (1981). *Quantum Mechanics in Hilbert Space*, 2nd edition, Academic Press, New York.

Reed, M., and Simon, B. (1972). *Methods of Modern Mathematical Physics Vol. I Functional Analysis*, Academic Press, New York.

Reed, M., and Simon, B. (1975). *Methods of Modern Mathematical Physics Vol. II Fourier Analysis, Seldfadjointess*, Academic Press, New York.

Richtmyer, R. D. (1978). *Principles of Advanced Mathematical Physics Vol. 1*, Springer-Verlag, New York.

Rieze, F., and Nagy, B. (1978). *Functional Analysis*, translated from the 2nd French edition by Boron, L. F., Frederick Ungar Publishing, New York.

Roman, P. (1965). *Advanced Quantum Theory*, Addison-Wesley, Reading, Mass.

Roman, P. (1975). *Some Modern Mathematics for Physicists and Other Outsiders Vol. 1*, Pergamon, New York.

Roman, P. (1975). *Some Modern Mathematics for Physicists and Other Outsiders Vol. 2*, Pergamon, New York.

Schiff, L. I. (1955). *Quantum Mechanics* (International student edition), McGraw-Hill, New York and Kōgakusha, Tokyo.

Schroeck, F. E., Jr. (1996). *Quantum Mechanics on Phase Space*, Kluwer, Dordrecht.

Sewell, G. L. (1986). *Quantum Theory of Collective Phenomena*, Clarendon Press, Oxford.

Smirnov, V. I. (1964). *A Course of Higher Mathematics Volume V*, translated by D. E. Brown, Pergamon Press, Oxford.

Spiegel (1), M. R. (1974). *Advanced Calculus*, a Schaum's Outline Series, McGraw-Hill, New-York.

Spiegel (2), M. R. (1974). *Complex Variables*, a Schaum's Outline Series, McGraw-Hill, New-York.

Spiegel (3), M. R. (1974). *Vector Analysis*, a Schaum's Outline Series, McGraw-Hill, New-York.

Takhtajan, L. A. (2008). *Quantum Mechanics for Mathematicians*, American Mathematical Society, Providence, Rhode Island.

Taylor, J. R. (2005). *Classical Mechanics*, University Science, Books, USA.

van Fraassen, B. C. (1991). *Quantum Mechanics, An Empiricist View*, Clarendon Press, Oxford.

von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, translated from the original German book published in 1932 by R. T. Beyer.

Wan, K. K. (2006). *From Micro to Macro Quantum Systems*, Imperial College Press, London.

Wan, K. K., and Menzies, D. (2009). *J. Theor. Phys.*, **48**, 282–304.

Weidmann, J. (1980). *Linear Operators in Hilbert Spaces*, Springer-Verlag, New York.

Wick, G. C., Wightman, A. S., and Wigner, E. P. (1952). *Phys. Rev.*, **88**, 101.

Williamson, J. H. (1962). *Lebesgue Integration*, Holt, Rinehart and Winston, New York.

Zettili, N. (2001). *Quantum Mechanics*, Wiley, Chichester.

Notation

Classical mechanics

$\{A,\ B\}$, 455
$H,\ K,\ V$, 449–451, 454
H_{ho}, 448
$\mathcal{L}$, 446, 449, 452
$\vec{L}$, L_x, L_y, L_z, 11
$\vec{L}_c$, L_c^2, 454
L_{cx}, L_{cy}, L_{cz}, 456
$\vec{L}_k$, 454
$\vec{p}$, $\vec{v}$, 10, 11
p_{cj}, $\vec{p}_c$, 453, 454, 449, 450
p_{cr}, $p_{c\theta}$, $p_{c\varphi}$, 450
$\vec{p}_l$, p_{lj}, 453, 448
$\vec{x}$, x_j, 8

Functions

$AC(\Lambda)$, $AC(I\!R^+)$, $AC(I\!R)$, 271
$AC(I\!R^2)$, $AC(I\!R^3)$, 272
$C(\Lambda)$, $C(I\!R^+)$, $C(I\!R)$, 270
$C(I\!R^2)$, $C(I\!R^3)$, 271
$C^\infty(\Lambda)$, $C^\infty(I\!R^+)$, $C^\infty(I\!R)$, 271, 272
$C^\infty(I\!R^2)$, $C^\infty(I\!R^3)$, 272
$C_D^\infty(\Lambda)$, $C_c^\infty(I\!R^+)$, $C_c^\infty(I\!R)$, 274
$C_c^\infty(I\!R^2)$, $C_c^\infty(I\!R^3)$, 275
$\chi_{\ \Lambda}(\tau)$, $\chi_{\ (a,b]}(\tau)$, 35, 36
$\chi_{\ (\infty,\tau]}(x)$, 220, 372
$\chi_{\tilde{}\ (-\infty,\tau]}(p)$, 381
$\delta(\tau)$, $\delta(\tau - a)$, 55
$D(x)$, 46
$\mathcal{F}(\tau + 0)$, $\mathcal{F}(\tau - 0)$, 29, 30, 31
$f_p(x)$, 332, 349–342
$f_s(\tau)$, 36
$g_{us}(\tau)$, $g_{as}(\tau)$, $g_s(\tau)$ $g_{nd}(\tau)$, 36, 37
$H_n(y)$, 273
$\mathcal{L}^2(\Lambda)$, 277
$\mathcal{L}^2(I\!R)$, $\mathcal{L}^2(I\!R^+)$, 275, 276
$\varphi(p)$, 340, 342
$\varphi_{Hn}(x)$, 273
$S_s(I\!R)$, $S_s(I\!R^2)$, $S_s(I\!R^3)$, 272, 274
$Y_{\ell,m}(\theta, \varphi)$, 287, 617, 620–621

Function spaces

$\vec{AC}(\Lambda)$, $\vec{AC}(I\!R^+)$, $\vec{AC}(I\!R)$, 271
$\vec{AC}(I\!R^2)$, $\vec{AC}(I\!R^3)$, 271
$\vec{C}(\Lambda)$, $\vec{C}(I\!R^+)$, $\vec{C}(I\!R)$, 204, 270
$\vec{C}(I\!R^2)$, $\vec{C}(I\!R^3)$, 271
$\vec{C}^\infty(\Lambda)$, $\vec{C}^\infty(I\!R^+)$, $\vec{C}^\infty(I\!R)$, 272
$\vec{C}^\infty(I\!R^2)$, $\vec{C}^\infty(I\!R^3)$, 272
$\vec{C}_c^\infty(I\!R^2)$, $\vec{C}_c^\infty(I\!R^3)$, 275
$\vec{C}_D^\infty(\Lambda)$, $\vec{C}_c^\infty(I\!R^+)$, $\vec{C}_c^\infty(I\!R)$, 275
$\vec{\mathcal{L}}^2(\Lambda)$, 277
$\vec{\mathcal{L}}^2(I\!R)$, 275, 276
$\vec{\mathcal{L}}^2(I\!R^+)$, 277
$\vec{L}^2(\Lambda)$, 278, 286
$\vec{L}^2(I\!R)$, $\vec{L}^2(I\!R^+)$, 277, 278
$\vec{L}^2(I\!R^2)$, $\vec{L}^2(I\!R^3)$, 278
$L^2(I\!\underset{\sim}{R})$, $\vec{L}^2(I\!\underset{\sim}{R})$, 344, 345
$\vec{L}^2(I\!R, dx)$, $\vec{L}^2(I\!R, dy)$, 424
$\vec{L}^2(I\!R^2, dxdy)$, $\vec{L}^2(I\!R^3, dxdydz)$, 424, 425
$L^2(I\!R^+, rdr)$, $L^2(I\!R^+, r^2dr)$, 424
$L^2(\mathcal{C}_a)$, $\vec{L}^2(\mathcal{C}_a)$, 278, 286
$L^2(\mathcal{S}_u)$, $\vec{L}^2(\mathcal{S}_u)$, 281, 286
$\vec{S}_s(I\!R)$, 272, 468
$\vec{S}_s(I\!R^2)$, $\vec{S}_s(I\!R^3)$, 274, 459, 460, 464

Matrices

Matrix representation

Measures and integration

Operators (basics)

Operators (differential)

Operators (functions)

Operators (multiplication)

Operators (others)

Probability

Set theory

Symbols (general)

Symbols (specific)

Vectors and vector spaces

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Index