Chapter 8
Quantum Formulation

In this chapter, we present the mathematical formalism used in quantum mechanics first and then derive expressions for canonical and microcanonical partition functions for quantum mechanical systems. This will help you develop familiarity with the basic ideas of quantum mechanics and the bra–ket notation you may encounter when consulting more advanced textbooks on statistical mechanics.

The important conclusion of this chapter is that (4.24) and (4.72) we obtained by means of classical statistical mechanics remain applicable in quantum statistical mechanics as well. Thus, the distinction between classical and quantum mechanical versions of statistical mechanics stems from the explicit expressions for $\overline{\Omega}(E)$ these mechanics predict and from the manner in which a given system populates the microstates accessible to it.

Many optional sections are included to provide explicit derivations of several key results from quantum mechanics we have already used in earlier chapters, but should probably be omitted upon the first reading. Keeping with our most immediate goals, we will not concern ourselves with experimental findings that forced the radical departure from classical mechanics and the eventual formulation of quantum mechanics in the early twentieth century. Interested readers can find these accounts in Refs. [7, 9] as well as in earlier chapters of many textbooks on quantum mechanics.

8.1 Vector Space

In classical mechanics, a microstate of a system with f mechanical degrees of freedom is specified by $2f$ variables (q^f, p^f). In quantum mechanics, a microstate of a system is specified by a vector in a complex vector space. In this section, we spend some time familiarizing ourselves with the basic properties of a vector space.

Before proceeding with the abstract vector space, it may be helpful for you to review the materials in Appendix A, in which vectors in the ordinary three-

I. Kusaka, *Statistical Mechanics for Engineers*,
DOI 10.1007/978-3-319-13809-1_8

dimensional space is discussed. This provides a useful analogy as we talk about abstract vector space in a more general term.

8.1.1 Definition

Let $\mathbb{K}$ denote either the set $\mathbb{C}$ of all complex numbers or the set $\mathbb{R}$ of all real numbers. We shall take for granted the rules for addition and multiplication of two numbers in $\mathbb{K}$. On this basis, we construct a vector space as follows. Let V be a set of objects and suppose that:

a. There is an operation called *addition* that assigns for each $x \in V$ and for each $y \in V$ another element in V, which we denote by $x+y$.
b. There is an operation called *multiplication by a number* that assigns for each number $\alpha \in \mathbb{K}$ and each $x \in V$ another element in V, which we denote by αx.

When the following properties are satisfied, the set V is said to be a **vector space**:

V1: $x+y=y+x$ for all $x,y \in V$.
V2: $x+(y+z)=(x+y)+z$ for all $x,y,z \in V$.
V3: There is an element $\theta \in V$ such that $x+\theta=x$ for all $x \in V$.
V4: To each $x \in V$, there corresponds an element $\bar{x} \in V$ such that $x+\bar{x}=\theta$.
V5: $\alpha(x+y)=\alpha x+\alpha y$ for all $x,y \in V$ and for all $\alpha \in \mathbb{K}$.
V6: $(\alpha+\beta)x=\alpha x+\beta x$ for all $\alpha,\beta \in \mathbb{K}$ and for all $x \in V$.
V7: $\alpha(\beta x)=(\alpha\beta)x$ for all $\alpha,\beta \in \mathbb{K}$ and for all $x \in V$.
V8: $1x=x$ for all $x \in V$, in which 1 is a number.

An element of V is called a **vector** and θ is called the **zero vector**. If $\mathbb{K}=\mathbb{R}$, V is a **real vector space**. If $\mathbb{K}=\mathbb{C}$ instead, V is a **complex vector space**.

We note that the standard notation for $\bar{x}$ appearing in V4 is $-x$. Thus, $x+\bar{y}=x+(-y)$, which is more commonly written as $x-y$. We note that $-x$ is a vector that gives θ when added to another vector x. In contrast, $(-1)x$ is a vector obtained by multiplying x by a number called -1. Thus, they are conceptually distinct objects. It is in order to stress this distinction that we use $\bar{x}$ for $-x$, even though they *turn out to be* the same vector as you can convince yourself in Exercise 8.1d.

Example 8.1. A real vector space: The set V of all pairs of real numbers, that is, $V=\{(x,y)|x,y \in \mathbb{R}\}$, is a real vector space with the addition defined by

$$(x_1,y_1)+(x_2,y_2)=(x_1+x_2,y_1+y_2) \tag{8.1}$$

and the multiplication by a number defined by

$$\alpha(x,y)=(\alpha x,\alpha y) \text{ for each } \alpha \in \mathbb{R}\,. \tag{8.2}$$

As a slightly less trivial example, let us consider a set of functions.

Example 8.2. A real vector space: The set V of all real-valued functions defined on the interval $[a,b] \subset \mathbb{R}$ is a vector space with the addition defined by

$$(f+g)(x) := f(x)+g(x) \text{ for each } f,g \in V \tag{8.3}$$

and the multiplication by a number defined by

$$(\alpha f)(x) := \alpha f(x) \text{ for each } f \in V \text{ and each } \alpha \in \mathbb{R}\,. \tag{8.4}$$

It is important to recognize the meaning of (8.3), in which $f+g$ is the name of a function in V. Being a function, there has to be a rule of assigning a real number to each $x \in \mathbb{R}$. This rule is given by (8.3) in terms of the rule for adding two real numbers and the rules that assign real numbers $f(x)$ and $g(x)$ to each $x \in \mathbb{R}$. Likewise, αf on the left-hand side of (8.4) is the name of a function in V. The rule of assigning a number to each $x \in \mathbb{R}$ is given in terms of the rule for multiplying two numbers, α and $f(x)$, and the rule of assignment for the function f.

Finally, if we replace $\alpha \in \mathbb{R}$ in (8.4) by $\alpha \in \mathbb{C}$ while insisting that V is still a set of all real-valued functions only, then V is no longer a vector space since $\alpha f \notin V$ if $\alpha \notin \mathbb{R}$.

An example of a complex vector space can be generated easily with a slight modification to the previous one.

Example 8.3. A complex vector space: The set V of all complex-valued functions defined on the interval $[a,b] \subset \mathbb{R}$ is a vector space with the addition defined by

$$(f+g)(x) := f(x)+g(x) \text{ for each } f,g \in V \tag{8.5}$$

and the multiplication by a number defined by

$$(\alpha f)(x) := \alpha f(x) \text{ for each } f \in V \text{ and each } \alpha \in \mathbb{C}\,. \tag{8.6}$$

Even if we limit ourselves to $\alpha \in \mathbb{R}$, V still is a vector space. According to our definition, however, V now becomes a real vector space.

Using the properties of a vector space enumerated above, we can prove various useful facts about it.

Example 8.4. Zero vector is unique: Take property V3, for example. This property demands that there must be a zero vector in a vector space. But,

it does not say how many zero vectors a particular vector space might contain. Nevertheless, we can show that, in any vector space V, there is only one zero vector. To see this, let both θ_1 and θ_2 be zero vectors of V. From V3, we have

$$x + \theta_1 = x \text{ for any } x \in V \tag{8.7}$$

and

$$x + \theta_2 = x \text{ for any } x \in V\ . \tag{8.8}$$

Since (8.7) holds for *any* $x \in V$, it holds for $\theta_2 \in V$ as well:

$$\theta_2 + \theta_1 = \theta_2\ . \tag{8.9}$$

Using the property V1 of a vector space, we may rewrite this as

$$\theta_1 + \theta_2 = \theta_2\ . \tag{8.10}$$

From (8.8) with $x = \theta_1 \in V$,

$$\theta_1 + \theta_2 = \theta_1\ , \tag{8.11}$$

Comparing the last two equations, we find that $\theta_1 = \theta_2$.

Exercise 8.1. Let V be a vector space:

a. Suppose that θ' satisfies $x + \theta' = x$ for a *particular* $x \in V$. Show that θ' is a zero vector of V.
b. Show that $0x = \theta$.
c. Show that, if $x + y = \theta$, then $y = \bar{x}$.
d. Show that $(-1)x = \bar{x}$.
e. Show that $\alpha\bar{x} = (-\alpha)x$ for any $\alpha \in \mathbb{K}$.
f. Show that $\alpha\theta = \theta$ for any $\alpha \in \mathbb{K}$.
g. Show that $\alpha\bar{x} = \overline{\alpha x}$ for any $\alpha \in \mathbb{K}$.
h. Show that

$$\alpha(x - y) = \alpha x - \alpha y \text{ for any } \alpha \in \mathbb{K} \tag{8.12}$$

and that

$$(\alpha - \beta)x = \alpha x - \beta x \text{ for any } \alpha, \beta \in \mathbb{K}. \tag{8.13}$$

Note that we interpret $\alpha x - \alpha y$ as $\alpha x + \overline{\alpha y}$ and $\alpha x - \beta x$ as $\alpha x + \overline{\beta x}$. ///

8.1.2 Linear Independence

For vectors $x_1, \ldots, x_m \in V$ and the numbers $c_1, \ldots, c_m \in \mathbb{K}$, the sum

$$c_1 x_1 + \cdots + c_m x_m \tag{8.14}$$

is called a **linear combination** of $x_1, \ldots, x_m$. The set U of *all* possible linear combinations of $x_1, \ldots, x_m$ is a vector space called the **span** of the set $\{x_1, \ldots, x_m\}$, which is denoted by

$$\text{span}(\{x_1, \ldots, x_m\}) \,. \tag{8.15}$$

We say that the vectors $x_1, \ldots, x_m$ **span** this vector space. In the ordinary three-dimensional space, for example, two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, neither of which is a scalar multiple of the other, span a two-dimensional vector space, whose elements are vectors on the plane containing $\boldsymbol{a}$ and $\boldsymbol{b}$.

Vectors $x_1, \ldots, x_m$ are said to be **linearly independent** if

$$c_1 x_1 + \cdots + c_m x_m = \theta \tag{8.16}$$

requires

$$c_i = 0 \text{ for all } i = 1, \ldots, m \,. \tag{8.17}$$

Equation (8.17) is clearly sufficient for (8.16) and is referred to as the **trivial solution** of (8.16). The vectors are **linearly dependent** if there is a nontrivial solution of (8.16), for which not every c_i is zero.

Example 8.5. Linear independence:

a. Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in the ordinary three-dimensional space are linearly dependent if one of them is a scalar multiple of the other:

$$\boldsymbol{a} = c\boldsymbol{b} \text{ for some } c \in \mathbb{R}. \tag{8.18}$$

On the other hand, if neither is a scalar multiple of the other, then they are linearly independent.

b. The zero vector, taken by itself, is linearly dependent since

$$c\theta = \theta \text{ for any } c \in \mathbb{K}. \tag{8.19}$$

c. Vectors $\theta, x_1, \ldots, x_m$ are linearly dependent since

$$c_0\theta + c_1 x_1 + \cdots + c_m x_m = \theta \tag{8.20}$$

is satisfied for any value of $c_0 \in \mathbb{K}$ if $c_1 = \cdots = c_m = 0$.

8.1.3 Basis

A set of vectors

$$\{\phi_1, \ldots, \phi_r\}\,, \tag{8.21}$$

where $\phi_i \in V$ for $i = 1, \ldots, r$, is said to be a **basis** of V if *any* $x \in V$ can be expressed as a linear combination of $\phi_1, \ldots, \phi_r$ in one and only one way. A member of the basis is referred to as a **basis vector**.

A vector space V may have multiple bases. However, the number of basis vectors is the same for all of them. It is therefore an *intrinsic property* of V and is called the **dimension** of V. In other words, if the dimension of V is r, a set of m vectors cannot be a basis if $m > r$ or $m < r$. This is established by the following two theorems, the proofs of which are provided in the next optional subsection.

Theorem 8.1. *A set of vectors* $\mathscr{B} := \{\phi_1, \ldots, \phi_r\}$ *is a basis of* V *if and only if* $\text{span}(\mathscr{B}) = V$ *and* $\phi_1, \ldots, \phi_r$ *are linearly independent.*

Theorem 8.2. *Suppose that* $\mathscr{B} := \{\phi_1, \ldots, \phi_r\}$ *is a basis of* V.

a. Vectors $\psi_1, \ldots, \psi_m$ *are not linearly independent if* $m > r$.
b. Vectors $\psi_1, \ldots, \psi_m$ *does not span* V *if* $m < r$.

8.1.4 †Proofs of Theorems

For completeness, we prove of the theorems just introduced.

Proof of Theorem 8.1

Let us suppose that $\mathscr{B}$ is a basis and establish the "only if" part of the theorem. We first show that $\text{span}(\mathscr{B}) = V$. If $x \in V$, x may be expressed as a linear combination of $\phi_1, \ldots, \phi_r$, and hence $x \in \text{span}(\mathscr{B})$. On the other hand, if $x \in \text{span}(\mathscr{B})$, then $x \in V$. Thus, the membership of $\text{span}(\mathscr{B})$ is identical to that of V. So, we have $\text{span}(\mathscr{B}) = V$. To establish that $\phi_1, \ldots, \phi_r$ are linearly independent, suppose that

$$c_1\phi_1 + \cdots + c_r\phi_r = \theta\,. \tag{8.22}$$

We note that θ can also be written as

$$\theta = 0\phi_1 + \cdots + 0\phi_r\,. \tag{8.23}$$

But, each vector in V, including θ, has a unique expression when written as a linear combination of the basis vectors in $\mathscr{B}$. Thus, comparing (8.22) and (8.23), we conclude that $c_i = 0$ for all $i = 1, \ldots, r$.

Conversely, suppose that span($\mathscr{B}$) $= V$ and that the members of $\mathscr{B}$ are linearly independent. Then, any member x of V belongs to span($\mathscr{B}$). This guarantees that x may be expressed as a linear combination of $\phi_1, \ldots, \phi_r$. If we suppose

$$x = c_1\phi_1 + \cdots + c_r\phi_r = d_1\phi_1 + \cdots + d_r\phi_r \,, \tag{8.24}$$

then,

$$(c_1 - d_1)\phi_1 + \cdots + (c_r - d_r)\phi_r = \theta \,. \tag{8.25}$$

Since $\phi_1, \ldots, \phi_r$ are linearly independent, only the trivial solution is possible, that is, $c_i = d_i$ for all $i = 1, \ldots, r$. That is, there is only one way to express x as a linear combination of $\phi_1, \ldots, \phi_r$. □

Proof of Theorem 8.2

We prove each part of the theorem by contradiction. That is, we assume that the claim is false and then show that this assumption leads to a contradiction:

a. Let us suppose that $\psi_1, \ldots, \psi_m$ $(m > r)$ are linearly independent. Since $\mathscr{B}$ is a basis,

$$\psi_1 = c_1\phi_1 + \cdots + c_r\phi_r \,, \tag{8.26}$$

in which at least one of $c_1, \ldots, c_r$ is nonzero. To be concrete, we suppose that c_1 is nonzero and solve (8.26) for ϕ_1:

$$\phi_1 = \frac{1}{c_1}\psi_1 - \frac{c_2}{c_1}\phi_2 - \cdots - \frac{c_r}{c_1}\phi_r \,. \tag{8.27}$$

There is no loss of generality here since we can always relabel ϕ_i's if necessary. By means of Theorem 8.1, we now show that

$$\mathscr{B}_1 := \{\psi_1, \phi_2, \ldots, \phi_r\} \tag{8.28}$$

is a basis. Since $\mathscr{B}$ is a basis, for any $x \in V$, we have

$$x = d_1\phi_1 + \cdots + d_r\phi_r = \frac{d_1}{c_1}\psi_1 + \left(d_2 - \frac{d_1c_2}{c_1}\right)\phi_2 + \cdots + \left(d_r - \frac{d_1c_r}{c_1}\right)\phi_r \,, \tag{8.29}$$

where we used (8.27). So, any $x \in V$ can be written as a linear combination of $\psi_1, \phi_2, \ldots, \phi_r$. To see that these vectors are linearly independent, let us consider the equation:

$$e_1\psi_1 + e_2\phi_2 + \cdots + e_r\phi_r = \theta \,. \tag{8.30}$$

If $e_1 \neq 0$, then ψ_1 is a linear combination of $\phi_2, \ldots, \phi_r$. This is contrary to the supposition that $c_1 \neq 0$ in (8.26). Since by definition of basis, the expression (8.26) is unique. So, $e_1 = 0$ and (8.30) reduces to

$$e_2\phi_2 + \cdots + e_r\phi_r = \theta \,. \tag{8.31}$$

But, since $\phi_2, \ldots, \phi_r$ are linearly independent, the only possible solution is that $e_2 = \cdots = e_r = 0$. Thus, only the trivial solution can satisfy (8.30).
Now that $\mathscr{B}_1$ is shown to be a basis, we can write

$$\psi_2 = f_1\psi_1 + f_2\phi_2 + \cdots + f_r\phi_r \,. \tag{8.32}$$

We observe that at least one of $f_2, \ldots, f_r$ is nonzero. Otherwise, we have $\psi_2 = f_1\psi_1$, which is in odds with our supposition that $\psi_1, \ldots, \psi_m$ are linearly independent. For concreteness, we assume that $f_2 \neq 0$ and write

$$\phi_2 = -\frac{f_1}{f_2}\psi_1 + \frac{1}{f_2}\psi_2 - \frac{f_3}{f_2}\phi_3 - \cdots - \frac{f_r}{f_2}\phi_r \,. \tag{8.33}$$

As before, we can show that

$$\mathscr{B}_2 := \{\psi_1, \psi_2, \phi_3, \ldots, \phi_r\} \tag{8.34}$$

is a basis. In fact, by means of (8.33), any linear combination of $\psi_1, \phi_2, \ldots, \phi_r$ may be expressed as a linear combination of $\psi_1, \psi_2, \phi_3, \ldots, \phi_r$. Let

$$g_1\psi_1 + g_2\psi_2 + g_3\phi_3 + \cdots + g_r\phi_r = \theta \,, \tag{8.35}$$

If $g_2 \neq 0$, this equation can be solved to express ψ_2 as a linear combination of $\psi_1, \phi_3, \ldots, \phi_r$, which is in odds with (8.32) with $f_2 \neq 0$. So, $g_2 = 0$. It follows that $g_1, g_3, \ldots, g_r$ are all zero since $\psi_1, \phi_3, \ldots, \phi_r$ are linearly independent as we have already established below (8.30).
Proceeding in this manner, we may establish that

$$\mathscr{B}_r := \{\psi_1, \ldots, \psi_r\} \tag{8.36}$$

is a basis. Thus, $\psi_{r+1}, \ldots, \psi_m$ can be expressed as linear combinations of $\psi_1, \ldots, \psi_r$, which contradicts our assumption that $\psi_1, \ldots, \psi_m$ are linearly independent. □

b. This time, we suppose that $m < r$ but that

$$V = \text{span}(\{\psi_1, \ldots, \psi_m\}) \,. \tag{8.37}$$

If $\psi_1, \ldots, \psi_m$ are linearly independent, then by definition, $\{\psi_1, \ldots, \psi_m\}$ is a basis of V. But, then according to part a of the current theorem, $\phi_1, \ldots, \phi_r$ with $r > m$ cannot be linearly independent, contrary to the supposition that $\mathscr{B}$ is a basis. If $\psi_1, \ldots, \psi_m$ are linearly dependent, then

$$c_1\psi_1 + \cdots + c_m\psi_m = \theta \tag{8.38}$$

and at least one of $c_1, \ldots, c_m$ is nonzero. To be concrete, suppose that $c_1 \neq 0$. Then,

$$\psi_1 = -\frac{c_2}{c_1}\psi_2 - \cdots - \frac{c_m}{c_1}\psi_m \,, \tag{8.39}$$

in terms of which any $x \in V$ may be expressed as a linear combination of $\psi_2, \ldots, \psi_m$:

$$x = d_1\psi_1 + \cdots + d_m\psi_m = \left(d_2 - \frac{d_1 c_2}{c_1}\right)\psi_2 + \cdots + \left(d_m - \frac{d_1 c_m}{c_1}\right)\psi_m \,. \quad (8.40)$$

In this way, we can remove ψ_i's one by one until the remaining ψ_i's are linearly independent. But, the span of the set consisting of the remaining ψ_i's is still V. As we saw above, this leads to a contradiction. □

8.1.5 Scalar Product Space

Let V be a complex vector space. If there is a rule that assigns a *number* $(x,y) \in \mathbb{C}$ for each pair of $x, y \in V$ and if the rule satisfies the following properties, then we call (x,y) the **scalar product** of x and y and the vector space V a **scalar product space**:

S1: $(x+y,z) = (x,z) + (y,z)$ for all $x,y,z \in V$.
S2: $(y,x) = (x,y)^*$ for all $x,y \in V$, where * denotes the complex conjugation.
S3: $(\alpha x, y) = \alpha(x,y)$ for all $x,y \in V$ and for all $\alpha \in \mathbb{C}$.
S4: $(x,x) \geq 0$ for all $x \in V$.
S5: $(x,x) = 0$ if and only if $x = \theta$.

According to S2, (x,x) is a real number. Because of S4, it is also nonnegative. Thus, $\|x\| := \sqrt{(x,x)}$, called the **norm** of the vector x, is a real number. We say that $x \neq \theta$ and $y \neq \theta$ are **orthogonal** if $(x,y) = 0$. If V is a real vector space, we demand that $(x,y) \in \mathbb{R}$ instead and replace S2 and S3 as follows:

S2′ $(y,x) = (x,y)$ for all $x,y \in V$.
S3′ $(\alpha x, y) = \alpha(x,y)$ for all $x,y \in V$ and for all $\alpha \in \mathbb{R}$.

Example 8.6. Three-dimensional vectors: The vector space V of ordinary three-dimensional vectors is a scalar product space if

$$(\boldsymbol{a}, \boldsymbol{b}) := \boldsymbol{a} \cdot \boldsymbol{b} \,. \quad (8.41)$$

With this definition, the norm of $\boldsymbol{a}$ is just the length of $\boldsymbol{a}$, which we have been denoting by $\|\boldsymbol{a}\|$.

The following example illustrates a scalar product defined on a complex vector space.

Example 8.7. Complex-valued functions: Let V be a complex vector space of complex-valued functions defined on the interval $[a,b] \subset \mathbb{R}$. If the membership of V is such that the integral

$$(f,g) := \int_a^b g^*(x) f(x) \mathrm{d}x \tag{8.42}$$

exists for all $f,g \in V$, then V is a scalar product space.

Exercise 8.2. Let V be a scalar product space. Show that

a.
$$(x, \alpha y) = \alpha^*(x,y) \text{ for any } \alpha \in \mathbb{C}. \tag{8.43}$$

b.
$$(x,\theta) = (\theta,x) = 0\,. \tag{8.44}$$

c.
$$(x, y+z) = (x,y) + (x,z)\,. \tag{8.45}$$

///

From (8.43) and (8.45), it follows that

$$(x, \alpha y + \beta z) = \alpha^*(x,y) + \beta^*(x,z)\,. \tag{8.46}$$

We refer to this result by saying that the scalar product is **anti-linear** in the second argument.

8.1.6 Orthonormal Basis

A basis $\{b_1, \ldots, b_r\}$ is said to form an **orthonormal basis** if

$$(b_i, b_j) = \delta_{ij} \text{ for all } i,j = 1,\ldots,r\,, \tag{8.47}$$

where δ_{ij} is the Kronecker delta.

Provided that r is finite, an orthonormal basis can be constructed from any basis. Given an arbitrary basis $\{\phi_1, \ldots, \phi_r\}$, we proceed as

$$\begin{aligned}
b_1 &:= \phi_1 / \|\phi_1\|\,, \\
b_2 &:= b_2' / \|b_2'\|\,,\ b_2' := \phi_2 - (\phi_2, b_1) b_1\,, \\
b_3 &:= b_3' / \|b_3'\|\,,\ b_3' := \phi_3 - (\phi_3, b_1) b_1 - (\phi_3, b_2) b_2\,, \\
&\ldots \\
b_r &:= b_r' / \|b_r'\|\,,\ b_r' := \phi_r - \sum_{j=1}^{r-1} (\phi_r, b_j) b_j\,.
\end{aligned} \tag{8.48}$$

The newly constructed basis vectors clearly have the unit norm. This construction is know as the **Gram–Schmidt orthogonalization**.

Exercise 8.3. Check the orthogonality of the basis $\{b_1, \ldots, b_r\}$. ///

8.1.7 Functions

A complex vector space plays an essential role in quantum mechanics. Accordingly, we shall restrict ourselves to such spaces in what follows.

Consider two sets V and W. A rule that assigns for each $x \in V$ precisely one element $y \in W$ is called a function. That F is such a function is indicated by the symbol $F : V \longrightarrow W$. The element in W assigned to $x \in V$ by F is denoted by $F(x)$. Now, let $F : V \longrightarrow W$ and $G : V \longrightarrow W$. If $F(x) = G(x)$ for each $x \in V$, then F and G refer to one and the same rule of assignment. The functions F and G are said to be identical.

8.1.8 Linear Functional

The function $F : V \longrightarrow \mathbb{C}$ on a vector space V is a **linear functional** if it satisfies:

F1: $F(x+y) = F(x) + F(y)$
F2: $F(\alpha x) = \alpha F(x)$

for all $x, y \in V$ and all $\alpha \in \mathbb{C}$. A linear functional is determined completely once the value of $F(x)$ is given for each $x \in V$. Because any x can be expressed in terms of an orthonormal basis $\{b_1, \ldots, b_r\}$ of V as[46]

$$x = c_1 b_1 + \cdots + c_r b_r \,, \tag{8.49}$$

we have

$$F(x) = c_1 F(b_1) + \cdots + c_r F(b_r) \,. \tag{8.50}$$

Thus, F is completely determined by specifying r complex numbers $F(b_1)$, …, $F(b_r)$.

A scalar product is an example of a linear functional. Let us pick a particular $f \in V$ and define F by

$$F(x) := (x, f) \text{ for all } x \in V \,. \tag{8.51}$$

Clearly, $F(x)$ satisfies the properties F1 and F2, and hence is a linear functional. It follows that any vector in V defines a linear functional.

The converse is also true. That is, to any linear functional F, there corresponds a *unique* vector $f \in V$ such that $F(x) = (x, f)$ for all $x \in V$. In fact, consider

$$f = F(b_1)^* b_1 + \cdots + F(b_r)^* b_r \,. \tag{8.52}$$

That is, given F, we first compute r complex numbers $F(b_1), \ldots, F(b_r)$. We then take their complex conjugate and construct f according to (8.52). The resulting vector f satisfies $(x, f) = F(x)$ for all $x \in V$.

Exercise 8.4. Verify this assertion. ///

Exercise 8.5. Given a linear functional $F(x)$, show that there is only one vector $f \in V$ that satisfies (8.51). ///

We have just seen that there is a one-to-one correspondence between a vector in V and a linear functional on V. Let $\widetilde{V}$ denote a set of all linear functionals on V and define the addition of two linear functionals and the multiplication of a linear functional by a number as follows:

D1: $(F+G)(x) := F(x) + G(x)$ for all $F, G \in \widetilde{V}$.
D2: $(\alpha F)(x) := \alpha F(x)$ for all $F \in \widetilde{V}$ and for all $\alpha \in \mathbb{C}$.

Then, $\widetilde{V}$ is called the **dual space** of V. We now show that $\widetilde{V}$ is a vector space. First, we need to ensure that the following two properties are satisfied by our definitions of the addition (D1) and the multiplication by a number (D2):

a. $F + G \in \widetilde{V}$. That is, $F + G$ is a linear functional, and hence

$$(F+G)(x+y) = (F+G)(x) + (F+G)(y) \tag{8.53}$$

and

$$(F+G)(\alpha x) = \alpha (F+G)(x) \tag{8.54}$$

hold for any $x, y \in V$ and $\alpha \in \mathbb{C}$.

b. $\alpha F \in \widetilde{V}$. That is, αF is a linear functional, and hence

$$(\alpha F)(x+y) = (\alpha F)(x) + (\alpha F)(y) \tag{8.55}$$

and

$$(\alpha F)(\beta x) = \beta (\alpha F)(x) \tag{8.56}$$

hold for any $x, y \in V$ and $\beta \in \mathbb{C}$.

The content of (8.54), which may seem a little obscure at first sight, is that applying $F + G$ on αx produces the same result as applying $F + G$ on x first and then multiplying the result by α. Equation (8.56) indicates that applying the function αF on βx gives the same result as applying αF on x first and then multiplying the result by β.

Exercise 8.6. Prove (8.53) and (8.54) from D1 and (8.55) and (8.56) from D2. ///

Now we have to show that D1 and D2 are compatible with V1–V8. For this purpose, let f and g in V correspond to F and G in $\widetilde{V}$, respectively. As we have just established, $F + G$ and αF are elements of $\widetilde{V}$. What are their corresponding vectors in V? From D1 and (8.51), we see that

$$(F+G)(x) = F(x) + G(x) = (x, f) + (x, g) = (x, f+g)\,, \tag{8.57}$$

where the last step follows from (8.45). Thus, $F+G \in \widetilde{V}$ corresponds $f+g \in V$. Similarly, using D2 and (8.51),

$$(\alpha F)(x) = \alpha F(x) = \alpha(x,f) = (x,\alpha^* f)\,, \tag{8.58}$$

where the last step follows from (8.43). So, $\alpha F \in \widetilde{V}$ corresponds $\alpha^* f \in V$. It is worth emphasizing that the corresponding vector is $\alpha^* f$ and *not* αf.

Exercise 8.7. Show that the dual space is a vector space. ///

8.1.9 Linear Operator

Let both V and W be vector spaces. The function $T : V \longrightarrow W$ is a **linear operator** if T satisfies:

L1: $T(x+y) = T(x)+T(y)$ for all $x,y \in V$.
L2: $T(\alpha x) = \alpha T(x)$ for all $\alpha \in \mathbb{C}$ and for all $x \in V$.

We often write $T(x)$ as Tx.

Example 8.8. Linear operators: The function $f(x) = ax$, where $a,x \in \mathbb{C}$, is a linear operator from $\mathbb{C}$ to $\mathbb{C}$. In contrast, $g(x) = ax+b$, where $b \neq 0$, is not.

Let V be a set of continuously differentiable functions. That is, if $f \in V$, then $\mathrm{d}f/\mathrm{d}x$ exists and $\mathrm{d}f/\mathrm{d}x$ is continuous. If W is the set of all continuous functions, then the derivative $\mathrm{d}/\mathrm{d}x$ is a linear operator from V to W.

Exercise 8.8. Let $T : V \longrightarrow W$ be a linear operator. Show that

a. $T(\theta_V) = \theta_W$, where θ_V is the zero vector in V and θ_W is the zero vector in W.
b. $T(x-y) = T(x)-T(y)$. ///

8.2 Kets, Bras, and Operators

In classical mechanics, a microstate of a system is specified by (q^f, p^f). In quantum mechanics, a microstate is specified by a **ket**, a vector in a complex vector space called a **ket space** V_k. As in Sect. 8.1.8, we can construct a dual space of V_k through the scalar product defined on V_k. The dual space $V_b = \widetilde{V}_k$ so constructed is referred to as a **bra space**, a member of which is a **bra**. These vectors and the spaces they belong to constitute a fundamental construct of quantum mechanics. In this section, we develop rules of computing with bras and kets.

8.2.1 Bra–Ket

A ket is denoted by a symbol $|\cdots\rangle$ replacing "$\cdots$" by an appropriate label. For a given $|f\rangle \in V_k$,

$$F(|x\rangle) := (|x\rangle, |f\rangle) \text{ for all } |x\rangle \in V_k \tag{8.59}$$

defines a linear functional F. When this is done for all $|f\rangle \in V_k$, we have a set V_b of linear functionals. A member of V_b is called a bra and is denoted by $\langle\cdots|$ with "$\cdots$" replaced by an appropriate label. In this notation, the linear functional F derived from $|f\rangle$ is simply written as $\langle f|$. Thus, in place of (8.59), we write

$$\langle f|x\rangle := (|x\rangle, |f\rangle) \text{ for all } |x\rangle \in V_k. \tag{8.60}$$

As we saw in Sect. 8.1.8, this establishes the one-to-one correspondence between $|f\rangle \in V_k$ and $\langle f| \in V_b$. Pictorially, we write

$$|f\rangle \in V_k \overset{\text{DC}}{\longleftrightarrow} \langle f| \in V_b\,, \tag{8.61}$$

where "DC" stands for "dual-correspondence."

Now that the correspondence between kets in V_k and bras in V_b is established, we can turn V_b into a vector space by defining the addition of two bras and the multiplication of a bra by a number in accordance with D1 and D2 from Sect. 8.1.8. (See Exercise 8.7.) In terms of the bra–ket notation, they are given by

$$(\langle f| + \langle g|)|x\rangle = \langle f|x\rangle + \langle g|x\rangle \text{ for all } \langle f|\,, \langle g| \in V_b, \tag{8.62}$$

and

$$(c\langle f|)|x\rangle = c\langle f|x\rangle \text{ for all } \langle f| \in V_b \text{ and for all } c \in \mathbb{C}, \tag{8.63}$$

respectively. Equation (8.57) now reads

$$(\langle f| + \langle g|)|x\rangle = (|x\rangle, |f\rangle + |g\rangle)\,, \tag{8.64}$$

from which we conclude

$$|f\rangle + |g\rangle \in V_k \overset{\text{DC}}{\longleftrightarrow} \langle f| + \langle g| \in V_b\,. \tag{8.65}$$

Similarly, we see from (8.58) that

$$(c\langle f|)|x\rangle = (|x\rangle, c^*|f\rangle)\,, \tag{8.66}$$

and hence

$$c^*|f\rangle \in V_k \overset{\text{DC}}{\longleftrightarrow} c\langle f| \in V_b\,. \tag{8.67}$$

It is often helpful to think of V_b as some sort of a mirror image of V_k with the rule of finding a mirror image of a particular object in V_k given by (8.61), (8.65), and (8.67). Additional rules will be derived in what follows.

Exercise 8.9. Let $|\theta\rangle$ be the zero vector in a ket space V_k. Show that the corresponding bra $\langle\theta|$ is the zero vector in the corresponding bra space V_b. ///

Because $\langle x|y\rangle$ is a scalar product, it satisfies the properties S1–S5. In the current notation, they are:

$S1_{bk}$: $\langle z|(|x\rangle + |y\rangle) = \langle z|x\rangle + \langle z|y\rangle$.
$S2_{bk}$: $\langle x|y\rangle = \langle y|x\rangle^*$.
$S3_{bk}$: $\langle y|(c|x\rangle) = c\langle y|x\rangle$ for all $c \in \mathbb{C}$.
$S4_{bk}$: $\langle x|x\rangle \geq 0$.
$S5_{bk}$: $\langle x|x\rangle = 0$ if and only if $|x\rangle = |\theta\rangle$.

As with any scalar product space, two kets $|x\rangle \neq |\theta\rangle$ and $|y\rangle \neq |\theta\rangle$ in V_k are said to be **orthogonal** if $\langle x|y\rangle = 0$. The quantity $\sqrt{\langle x|x\rangle}$ is the **norm** of $|x\rangle$:

$$\||x\rangle\| := \sqrt{\langle x|x\rangle}\,. \tag{8.68}$$

At this point, our terminology "bra" and "ket" should appear appropriate:

$$\underset{\text{bra(c)ket}}{\langle x \mid y\rangle}\,. \tag{8.69}$$

This "bra–ket" notation is due to P. A. M. Dirac and allows us to manipulate various quantities arising in quantum mechanics rather mechanically.

Exercise 8.10. Let $c \in \mathbb{C}$. Show that

$$\langle y|(c|x\rangle) = (\langle y|c)|x\rangle\,, \tag{8.70}$$

which indicates that brackets are unnecessary and we shall write the product simply as $\langle y|c|x\rangle$. ///

Equations (8.62), (8.63), and those occurring $S1_{bk}$–$S5_{bk}$ are simply a matter of definitions. However, (8.62) and $S1_{bk}$ taken together indicates that multiplication between a ket and a bra is *distributive* over addition. According to (8.70), the multiplication of a bra, a number, and a ket, written in this order, is *associative*. $S3_{bk}$ then allows us to pull out the complex number placed between a bra and a ket, leaving an uninterrupted bra–ket. Without the benefit of the bra–ket notation, we have to express this fact as

$$(c|x\rangle, |y\rangle) = (|x\rangle, c^*|y\rangle) = c(|x\rangle, |y\rangle)\,. \tag{8.71}$$

Dirac's bra–ket notation expresses various definitions and properties pertaining to a scalar product space and its dual space through intuitively transparent distributive and associative laws. We shall encounter a few more such instances that illustrates the power of his notation.

8.2.2 Operator and Adjoint

Let $\hat{X}$ denote a linear operator from V_k to V_k. Then, we may write

$$|\beta\rangle = \hat{X}|\alpha\rangle \, . \tag{8.72}$$

to express the fact that, when the operator $\hat{X}$ acts on $|\alpha\rangle$, it produces another ket $|\beta\rangle$. Even though $\hat{X}$ is a linear operator on V_k, it is actually possible to define its action on a bra and hence think of $\hat{X}$ as a linear operator from V_b to V_b.

By convention, an operator acting on a bra is written to the right of the bra as in $\langle\alpha|\hat{X}$. This is a new construct, to which we have complete freedom to assign any meaning we wish. Taking advantage of this freedom, we declare that $\langle\alpha|\hat{X}$ be a functional on V_k, that is, it acts on a ket in V_k and produces a complex number. But, a functional is defined completely once its action on every $|\beta\rangle \in V_k$ is specified. Accordingly, we *define* $\langle\alpha|\hat{X}$ by

$$(\langle\alpha|\hat{X})|\beta\rangle := \langle\alpha|(\hat{X}|\beta\rangle) \text{ for all } |\beta\rangle \in V_k \, . \tag{8.73}$$

Since $\hat{X}$, when regarded as acting on V_k, is a linear operator, this definition implies that the functional $\langle\alpha|\hat{X}$ is linear. In fact,

$$\begin{aligned}(\langle\alpha|\hat{X})(|\beta\rangle + |\gamma\rangle) &= \langle\alpha|[\hat{X}(|\beta\rangle + |\gamma\rangle)] = \langle\alpha|(\hat{X}|\beta\rangle + \hat{X}|\gamma\rangle) \\ &= \langle\alpha|(\hat{X}|\beta\rangle) + \langle\alpha|(\hat{X}|\gamma\rangle) = (\langle\alpha|\hat{X})|\beta\rangle + (\langle\alpha|\hat{X})|\gamma\rangle \, ,\end{aligned} \tag{8.74}$$

where the third equality follows from $\mathrm{S1_{bk}}$. Also,

$$(\langle\alpha|\hat{X})(c|\beta\rangle) = \langle\alpha|[\hat{X}(c|\beta\rangle)] = \langle\alpha|(c\hat{X}|\beta\rangle) = c\langle\alpha|(\hat{X}|\beta\rangle) = c(\langle\alpha|\hat{X})|\beta\rangle \, , \tag{8.75}$$

where we used $\mathrm{S3_{bk}}$ in the third equality.

Since $\langle\alpha|\hat{X}$ is a linear functional on V_k, it is nothing but a bra in V_b. So, we are justified in writing

$$\langle\alpha|\hat{X} = \langle\gamma_\alpha| \in V_b \, , \tag{8.76}$$

where we use the subscript α to indicate the $\langle\alpha|$ dependence of the resulting bra. That is, $\hat{X}$ acts on a particular bra $\langle\alpha|$ and produces yet another bra $\langle\gamma_\alpha|$, which in turn is completely specified by means of (8.73). If $\langle\alpha|\hat{X}$ is defined for all $\langle\alpha| \in V_b$ by (8.73), the end result is a function $\hat{X}$ on V_b.

We note that $\hat{X}$ acting on V_b is a linear operator, that is,

$$(\langle\alpha| + \langle\beta|)\hat{X} = \langle\alpha|\hat{X} + \langle\beta|\hat{X} \, , \tag{8.77}$$

and

$$(c\langle\alpha|)\hat{X} = c(\langle\alpha|\hat{X}) \text{ for all } c \in \mathbb{C}. \tag{8.78}$$

Exercise 8.11. Prove (8.77) and (8.78). ///

We recall from Sect. 8.1.8 that, for each bra in V_b, there corresponds a ket in V_k. We denote the ket corresponding to $\langle\alpha|\hat{X}$ by $\hat{X}^\dagger|\alpha\rangle$:

$$\langle\alpha|\hat{X} \overset{\text{DC}}{\longleftrightarrow} \hat{X}^\dagger|\alpha\rangle \,. \tag{8.79}$$

As a consequence of (8.77) and (8.78), $\hat{X}^\dagger$ acting on V_k is a linear operator. To see this, let $|\gamma\rangle = |\alpha\rangle + |\beta\rangle$. Then,

$$\hat{X}^\dagger(|\alpha\rangle + |\beta\rangle) = \hat{X}^\dagger|\gamma\rangle \overset{\text{DC}}{\longleftrightarrow} \langle\gamma|\hat{X} = (\langle\alpha| + \langle\beta|)\hat{X} \,, \tag{8.80}$$

where we used (8.65) in the last step. On the other hand, (8.77) gives

$$(\langle\alpha| + \langle\beta|)\hat{X} = \langle\alpha|\hat{X} + \langle\beta|\hat{X} \overset{\text{DC}}{\longleftrightarrow} \hat{X}^\dagger|\alpha\rangle + \hat{X}^\dagger|\beta\rangle \,, \tag{8.81}$$

where the last step also follows from (8.65). Since dual correspondence is one-to-one, we conclude that

$$\hat{X}^\dagger(|\alpha\rangle + |\beta\rangle) = \hat{X}^\dagger|\alpha\rangle + \hat{X}^\dagger|\beta\rangle \,. \tag{8.82}$$

Exercise 8.12. Show that

$$\hat{X}^\dagger(c|\alpha\rangle) = c(\hat{X}^\dagger|\alpha\rangle) \text{ for all } c \in \mathbb{C}, \tag{8.83}$$

thus completing the demonstration that $\hat{X}^\dagger$ on V_k is linear. We have already seen how the linearity of $\hat{X}$ on V_k implied that of $\hat{X}$ on V_b. It follows that $\hat{X}^\dagger$ on V_b is also linear. ///

In view of (8.73), the multiplication of a bra, an operator, and a ket, written in this order, is associative. Thus, the brackets are unnecessary and we may simply write $\langle\alpha|\hat{X}|\beta\rangle$ for the product. The content of (8.73) is far from trivial. Written by means of the earlier notation for a scalar product, it reads

$$(|\beta\rangle, \hat{X}^\dagger|\alpha\rangle) = (\hat{X}|\beta\rangle, |\alpha\rangle) \,. \tag{8.84}$$

Dirac's notation allows us to perform complex operations such as this one effortlessly and is extremely useful for computation.

We now show that

$$\langle\alpha|\hat{X}^\dagger|\beta\rangle = \langle\beta|\hat{X}|\alpha\rangle^* \,. \tag{8.85}$$

From what was just said about (8.73),

$$\langle\alpha|\hat{X}^\dagger|\beta\rangle = \langle\alpha|(\hat{X}^\dagger|\beta\rangle) \,. \tag{8.86}$$

Using (8.60) and S2,

$$\langle\alpha|(\hat{X}^\dagger|\beta\rangle) = (\hat{X}^\dagger|\beta\rangle, |\alpha\rangle) = (|\alpha\rangle, \hat{X}^\dagger|\beta\rangle)^* . \tag{8.87}$$

But, since the bra corresponding to the ket $\hat{X}^\dagger|\beta\rangle$ is just $\langle\beta|\hat{X}$, we have

$$(|\alpha\rangle, \hat{X}^\dagger|\beta\rangle) = (\langle\beta|\hat{X})|\alpha\rangle = \langle\beta|\hat{X}|\alpha\rangle . \tag{8.88}$$

Combining everything, we arrive at (8.85).

Exercise 8.13. Show that

$$(\hat{X}^\dagger)^\dagger = \hat{X} . \tag{8.89}$$

///

Exercise 8.14. Show that

$$(c\hat{X})^\dagger = c^*\hat{X}^\dagger , \tag{8.90}$$

where

$$(c\hat{X})|\alpha\rangle := c(\hat{X}|\alpha\rangle) \text{ for all } c \in \mathbb{C}. \tag{8.91}$$

///

The linear operator $\hat{X}^\dagger$ is called the **adjoint** of $\hat{X}$. In general,

$$\hat{X}^\dagger|\alpha\rangle \neq \hat{X}|\alpha\rangle , \tag{8.92}$$

and hence $\hat{X}^\dagger \neq \hat{X}$. *If* $\hat{X}^\dagger = \hat{X}$, then the operator $\hat{X}$ is said to be **Hermitian**.

8.2.3 Addition and Multiplication of Operators

Addition of two operators is defined by

$$(\hat{X} + \hat{Y})|\alpha\rangle := \hat{X}|\alpha\rangle + \hat{Y}|\alpha\rangle \text{ for all } |\alpha\rangle \in V_k, \tag{8.93}$$

while multiplication of two operators is defined by

$$(\hat{X}\hat{Y})|\alpha\rangle := \hat{X}(\hat{Y}|\alpha\rangle) \text{ for all } |\alpha\rangle \in V_k. \tag{8.94}$$

Repeated application of (8.73) combined with (8.94) yields

$$\langle\beta|(\hat{X}\hat{Y}) = (\langle\beta|\hat{X})\hat{Y} . \tag{8.95}$$

In fact, for any $|\alpha\rangle \in V_k$,

$$[\langle\beta|(\hat{X}\hat{Y})]|\alpha\rangle = \langle\beta|[(\hat{X}\hat{Y})|\alpha\rangle] = \langle\beta|[\hat{X}(\hat{Y}|\alpha\rangle)] = (\langle\beta|\hat{X})(\hat{Y}|\alpha\rangle) = [(\langle\beta|\hat{X})\hat{Y}]|\alpha\rangle . \tag{8.96}$$

Since this holds for any $|\alpha\rangle$, we arrive at (8.95).

Exercise 8.15. Prove the following relations:

a.
$$\hat{X}(\hat{Y}+\hat{Z}) = \hat{X}\hat{Y} + \hat{Y}\hat{Z}\,. \tag{8.97}$$

b.
$$(\hat{X}\hat{Y})\hat{Z} = \hat{X}(\hat{Y}\hat{Z})\,, \tag{8.98}$$

c.
$$(\hat{X}\hat{Y})^\dagger = \hat{Y}^\dagger \hat{X}^\dagger\,. \tag{8.99}$$

///

Equation (8.98) indicates that a multiplication of operators is associative, and hence the product may be written simply as $\hat{X}\hat{Y}\hat{Z}$. Multiplication, however, is not necessarily commutative:

$$\hat{X}\hat{Y} \neq \hat{Y}\hat{X}. \tag{8.100}$$

The difference between these two products defines a **commutator** between $\hat{X}$ and $\hat{Y}$:

$$[\hat{X},\hat{Y}] := \hat{X}\hat{Y} - \hat{Y}\hat{X}\,, \tag{8.101}$$

which is another operator. The following set of identities can be verified easily:

$$[\hat{X},\hat{X}] = 0\,, \tag{8.102}$$
$$[c,\hat{X}] = 0\,, \quad c \in \mathbb{C} \tag{8.103}$$
$$[\hat{X},\hat{Y}+\hat{Z}] = [\hat{X},\hat{Y}] + [\hat{X},\hat{Z}]\,, \tag{8.104}$$
$$[\hat{X},\hat{Y}\hat{Z}] = \hat{Y}[\hat{X},\hat{Z}] + [\hat{X},\hat{Y}]\hat{Z}\,, \tag{8.105}$$

and

$$[\hat{X},[\hat{Y},\hat{Z}]] + [\hat{Y},[\hat{Z},\hat{X}]] + [\hat{Z},[\hat{X},\hat{Y}]] = 0\,. \tag{8.106}$$

The last equation is the quantum mechanical version of **Jacobi's identity** to be compared with (1.196).

8.2.4 Unitary Operator

Let $\hat{I}$ denote the **unit operator** defined by

$$\hat{I}|\alpha\rangle = |\alpha\rangle \text{ for all } |\alpha\rangle \in V_k\,. \tag{8.107}$$

An operator $\hat{U}$ satisfying

$$\hat{U}\hat{U}^\dagger = \hat{U}^\dagger\hat{U} = \hat{I} \tag{8.108}$$

is called a **unitary operator**.

Since $(\hat{U}^\dagger)^\dagger = \hat{U}$, (8.108) indicates that $\hat{V} = \hat{U}^\dagger$ is unitary if $\hat{U}$ is. Unitary operators do not affect the norm of a ket:

$$\|\hat{U}|\alpha\rangle\| = \||\alpha\rangle\| \,. \tag{8.109}$$

This follows immediately from (8.108):

$$\|\hat{U}|\alpha\rangle\| = \sqrt{\langle\alpha|\hat{U}^\dagger\hat{U}|\alpha\rangle} = \sqrt{\langle\alpha|\hat{I}|\alpha\rangle} = \sqrt{\langle\alpha|\alpha\rangle} \,. \tag{8.110}$$

Exercise 8.16. Show that $\hat{X}\hat{I} = \hat{I}\hat{X} = \hat{X}$. ///

8.2.5 Outer Product

As we have seen, the scalar product $\langle\alpha|\beta\rangle$ is a complex number. Here, we define the so-called **outer product** of $|\alpha\rangle$ and $\langle\beta|$ by the equation

$$(|\alpha\rangle\langle\beta|)|\gamma\rangle := |\alpha\rangle(\langle\beta|\gamma\rangle) \text{ for all } |\gamma\rangle \in V_k \,. \tag{8.111}$$

On the right-hand side, we have a complex number multiplying a ket, the result being another ket in V_k. That is, the outer product $|\alpha\rangle\langle\beta|$ acting on a ket produces a new ket proportional to $|\alpha\rangle$. Because of the linearity of the scalar product, the outer product is a linear operator. As shown in Exercise 8.17, however, the outer product $|\alpha\rangle\langle\beta|$ is *not* a Hermitian operator unless $|\alpha\rangle = |\beta\rangle$. Finally, (8.111) indicates that the multiplication of a ket, a bra, and a ket, written in this order, is associative. Thus, the brackets are unnecessary and we shall write the product simply as $|\alpha\rangle\langle\beta|\gamma\rangle$.

Exercise 8.17.

a. Let $\hat{X}$=$|\alpha\rangle\langle\beta|$. Show that $\hat{X}^\dagger = |\beta\rangle\langle\alpha|$.
b. Let $\hat{Y} = -i|\alpha\rangle\langle\beta|$ and find $\hat{Y}^\dagger$. ///

If a given pair of expressions, each including bras, kets, operators, or complex numbers, are in dual correspondence with each other, they are said to be the **adjoint** of each other. Table 8.2.5 summarizes useful rules and formulae for finding the adjoint of a given expression.

8.3 Eigenkets and Eigenvalues

Eigenvalues of operators and their corresponding kets play a crucial role in quantum mechanics. Accordingly, we shall summarize their properties in this section.

Dual correspondence	Equation	Identity	Equation
$\lvert f\rangle \overset{\text{DC}}{\longleftrightarrow} \langle f\rvert$	(8.61)	$\langle\alpha\rvert\hat{X}^\dagger\lvert\beta\rangle = \langle\beta\rvert\hat{X}\lvert\alpha\rangle^*$	(8.85)
$\lvert f\rangle + \lvert g\rangle \overset{\text{DC}}{\longleftrightarrow} \langle f\rvert + \langle g\rvert$	(8.65)	$(\hat{X}^\dagger)^\dagger = \hat{X}$	(8.89)
$c^*\lvert f\rangle \overset{\text{DC}}{\longleftrightarrow} c\langle f\rvert$	(8.67)	$(c\hat{X})^\dagger = c^*\hat{X}^\dagger$	(8.90)
$\hat{X}^\dagger\lvert f\rangle \overset{\text{DC}}{\longleftrightarrow} \langle f\rvert\hat{X}$	(8.79)	$(\hat{X}\hat{Y})^\dagger = \hat{Y}^\dagger\hat{X}^\dagger$	(8.99)
$c\lvert\alpha\rangle\langle\beta\rvert \overset{\text{DC}}{\longleftrightarrow} c^*\lvert\beta\rangle\langle\alpha\rvert$	Exercise 8.17		

Table 8.1 A list of useful rules and formulae for finding the adjoint of a given expression.

8.3.1 Definition

In general, an operator $\hat{X}$ acting on a ket produces another ket different from the original one. However, there may be a ket $|\alpha\rangle$ for which

$$\hat{X}|\alpha\rangle = \lambda|\alpha\rangle\ ,\ \lambda \in \mathbb{C} \tag{8.112}$$

holds. If this happens, λ and $|\alpha\rangle$ are called, respectively, an **eigenvalue** of $\hat{X}$ and the **eigenket** corresponding (or belonging) to λ. The set of all eigenvalues is called the **spectrum** of $\hat{X}$.

In view of Exercise 8.8a,

$$\hat{X}|\theta\rangle = |\theta\rangle = \lambda|\theta\rangle \text{ for any } \lambda \in \mathbb{C} \tag{8.113}$$

for any linear operator $\hat{X}$. That is, $|\theta\rangle$ satisfies (8.112) for arbitrary $\lambda \in \mathbb{C}$. By convention, however, the zero vector is not an eigenket. According to the definition, if $|\alpha\rangle$ is an eigenket, so is $c|\alpha\rangle$, where c is a nonzero complex number.

For a given eigenvalue, there may be multiple corresponding *linearly independent* eigenkets. Such an eigenvalue is said to be **degenerate**. If the number of linearly independent eigenkets corresponding to this eigenvalue is s, we say that the eigenvalue is s-fold degenerate or that its **degeneracy** is s.

Suppose that the eigenvalue λ of $\hat{X}$ is s-fold degenerate and denotes the corresponding s linearly independent eigenkets by $|\alpha_1\rangle, \ldots, |\alpha_s\rangle$. Then, their linear combination

$$|\phi\rangle := \sum_{k=1}^{s} c_k|\alpha_k\rangle \tag{8.114}$$

is also an eigenket of $\hat{X}$ corresponding to λ. In fact,

$$\hat{X}|\phi\rangle = \sum_{k=1}^{s} c_k\hat{X}|\alpha_k\rangle = \lambda\sum_{k=1}^{s} c_k|\alpha_k\rangle = \lambda|\phi\rangle\ . \tag{8.115}$$

The set of vectors $\{|\alpha_1\rangle, \ldots, |\alpha_s\rangle\}$ spans a s-dimensional vector space, called the **eigensubspace** of the eigenvalue λ.

Example 8.9. Rotation: If we take an ordinary three-dimensional vector $\boldsymbol{a}$ and rotate it around a given axis by an angle θ ($0 < \theta < 2\pi$), the result is another vector $\boldsymbol{b}$. The rotation is a linear operator, which we denote by $\hat{R}_\theta$. While $\boldsymbol{b} \neq \boldsymbol{a}$ in general,

$$\hat{R}_\theta \boldsymbol{a} = \lambda \boldsymbol{a} \ , \ \lambda \in \mathbb{R} \tag{8.116}$$

holds with $\lambda = 1$ for vectors that are either parallel or antiparallel to the axis of rotation. Then, $\lambda = 1$ is a nondegenerate eigenvalue of $\hat{R}_\theta$ with the said vectors all belonging to this eigenvalue. Geometrically, it is clear that there is no other eigenvalue for a general value of θ. However, $\hat{R}_\pi$ has another eigenvalue $\lambda = -1$. The vectors perpendicular to the axis of rotation are the corresponding eigenvectors. Because these vectors lie on a plane, the degeneracy of $\lambda = -1$ is two.

8.3.2 Closure relation

A Hermitian operator $\hat{X}$ defined on a ket space V_k is said to be an **observable** if a set of all linearly independent eigenkets of $\hat{X}$ forms a basis of V_k. It can be shown that a Hermitian operator always is an observable if the dimension of V_k is finite. In what follows, we denote generic observables by the symbols $\hat{A}$, $\hat{B}$, $\hat{C}$, and so on to distinguish them clearly from more general linear operators, which we continue to indicate by $\hat{X}$, $\hat{Y}$, $\hat{Z}$, etc.

Let $\hat{A}$ be an observable and denote its eigenvalues by $a_1, \ldots, a_r$. We label an eigenket by the eigenvalue to which it belongs. For example,

$$\hat{A}|a_1\rangle = a_1|a_1\rangle \ , \ldots , \ \hat{A}|a_r\rangle = a_r|a_r\rangle \ . \tag{8.117}$$

If there is a degenerate eigenvalue, some of a_i's are equal. For example, if $a_1 = 5$ is twofold degenerate, $a_1 = a_2 = 5$ and both $|a_1\rangle$ and $|a_2\rangle$ would become $|5\rangle$, requiring additional label to distinguish them. This may be the case especially during a solution of a concrete numerical problem. However, since we do not need to refer to actual numerical values of a_i's in what follows, the notation we have just adopted will be sufficient.

Theorem 8.3. *Let $\hat{A}$ be an observable. Then,*

a. The eigenvalues of $\hat{A}$ are real.
b. The eigenkets of $\hat{A}$ corresponding to different eigenvalues are orthogonal.

Proof. Let

$$\hat{A}|a_i\rangle = a_i|a_i\rangle \quad \text{and} \quad \hat{A}|a_j\rangle = a_j|a_j\rangle \ . \tag{8.118}$$

Multiplying the first equation by $\langle a_j|$ from the left,

$$\langle a_j|\hat{A}|a_i\rangle = a_i\langle a_j|a_i\rangle \ . \tag{8.119}$$

Taking the adjoint of the second equation and multiplying the resulting equation by $|a_i\rangle$ from the right,

$$\langle a_j|\hat{A}|a_i\rangle = a_j{}^*\langle a_j|a_i\rangle \,, \tag{8.120}$$

where we note that $\hat{A}^\dagger = \hat{A}$. Subtracting (8.120) from (8.119),

$$(a_i - a_j{}^*)\langle a_j|a_i\rangle = 0 \,. \tag{8.121}$$

Let $a_i = a_j$. Since $|a_i\rangle \neq |\theta\rangle$ by convention, (8.121) implies that $a_i = a_i{}^*$. This proves a. Because of a, (8.121) becomes

$$(a_i - a_j)\langle a_j|a_i\rangle = 0 \,. \tag{8.122}$$

If $a_i \neq a_j$, then $\langle a_j|a_i\rangle = 0$, thus establishing b. □

What happens if an eigenvalue is s-fold degenerate? According to this theorem, these s eigenkets are certainly orthogonal to eigenkets belonging to *other* eigenvalues. However, the theorem does not tell us if they are orthogonal among themselves. Fortunately, the Gram–Schmidt orthogonalization we saw in Sect. 8.1.6 allows us to construct s mutually orthogonal eigenkets as linear combinations of the original eigenkets. With a proper normalization, therefore, a set of all linearly independent eigenkets of $\hat{A}$ can be made orthonormal:

$$\langle a_i|a_j\rangle = \delta_{ij} \,. \tag{8.123}$$

In what follows, we shall assume that this is done.

Using an orthonormal basis $\{|a_1\rangle, \ldots, |a_r\rangle\}$ so constructed, we may write any ket $|\alpha\rangle \in V_k$ as

$$|\alpha\rangle = \sum_{i=1}^{r} C_{a_i}|a_i\rangle \,. \tag{8.124}$$

Multiplying (8.124) by $\langle a_j|$, we find

$$\langle a_j|\alpha\rangle = \sum_{i=1}^{r} C_{a_i}\langle a_j|a_i\rangle = \sum_{i=1}^{r} C_{a_i}\delta_{ji} = C_{a_j} \,. \tag{8.125}$$

When this is substituted back into (8.124), we obtain

$$|\alpha\rangle = \sum_{i=1}^{r} (\langle a_i|\alpha\rangle)\,|a_i\rangle = \sum_{i=1}^{r} |a_i\rangle\,(\langle a_i|\alpha\rangle) = \sum_{i=1}^{r} (|a_i\rangle\langle a_i|)\,|\alpha\rangle = \left[\sum_{i=1}^{r} |a_i\rangle\langle a_i|\right]|\alpha\rangle \,. \tag{8.126}$$

Since this holds for any $|\alpha\rangle \in V_k$, we conclude that

$$\sum_{i=1}^{r} |a_i\rangle\langle a_i| = \hat{I} \,, \tag{8.127}$$

which is to be compared with (A.21). Equation (8.127) is known as the **closure relation** and will be in frequent use in what follows.

Exercise 8.18. Rewrite (8.52) using the bra–ket notation. ///

Exercise 8.19. Show that

$$\hat{A} = \sum_{i=1}^{r} |a_i\rangle a_i \langle a_i| \,. \tag{8.128}$$

///

A function $f(\hat{A})$ of the operator $\hat{A}$ is a yet another operator and is defined by its Maclaurin expansion. (See Sect. B.1.) That is,

$$f(\hat{A}) = f(0)\hat{I} + f'(0)\hat{A} + \frac{1}{2}f''(0)\hat{A}^2 + \frac{1}{3}f'''(0)\hat{A}^3 + \cdots \,. \tag{8.129}$$

Exercise 8.20. Prove the following relations:

a.

$$f(\hat{A})|a_i\rangle = f(a_i)|a_i\rangle \,. \tag{8.130}$$

b.

$$f(\hat{A}) = \sum_{i=1}^{r} |a_i\rangle f(a_i) \langle a_i| \,. \tag{8.131}$$

///

For certain functions, (8.129) fails to define $f(\hat{A})$. For example, $\sqrt{x}$ does not have the Maclaurin expansion and $\sqrt{\hat{A}}$ cannot be defined by means of (8.129). More generally, we can define a function of an operator by (8.130). In fact, for any $|\alpha\rangle \in V_k$, (8.130) assigns a unique ket in V_k:

$$f(\hat{A})|\alpha\rangle = \sum_{i=1}^{r} f(\hat{A})|a_i\rangle\langle a_i|\alpha\rangle = \sum_{i=1}^{r} f(a_i)|a_i\rangle\langle a_i|\alpha\rangle \,, \tag{8.132}$$

where we used the closure relation in the first equality.

8.3.3 Matrix Representation

Using the closure relation twice, we have

$$\hat{X} = \sum_i \sum_j |a_i\rangle\langle a_i|\hat{X}|a_j\rangle\langle a_j| \,, \tag{8.133}$$

where $\langle a_i|\hat{X}|a_j\rangle$ is called a **matrix element** of $\hat{X}$ in the A-representation with A referring to the fact that we have used the set of all eigenkets $\{|a_1\rangle, \ldots, |a_r\rangle\}$ of

$\hat{A}$ as the basis. Evidently, numerical values of the matrix elements depend on the representation, that is, the choice of the basis kets.

Let's consider the product $\hat{Z} = \hat{X}\hat{Y}$:

$$\langle a_i|\hat{Z}|a_j\rangle = \langle a_i|\hat{X}\hat{Y}|a_j\rangle = \sum_k \langle a_i|\hat{X}|a_k\rangle\langle a_k|\hat{Y}|a_j\rangle \,. \tag{8.134}$$

With an abbreviation $X_{ij} := \langle a_i|\hat{X}|a_j\rangle$, this equation may be written as

$$Z_{ij} = \sum_k X_{ik}Y_{kj} \,, \tag{8.135}$$

or more explicitly,

$$\begin{pmatrix} Z_{11} & Z_{12} & \cdots \\ Z_{21} & Z_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots \\ X_{21} & X_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} Y_{11} & Y_{12} & \cdots \\ Y_{21} & Y_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \,. \tag{8.136}$$

The term "matrix element" is seen to be quite appropriate.

We now turn to the equation:

$$|\beta\rangle = \hat{X}|\alpha\rangle \,. \tag{8.137}$$

Multiplying this equation by $\langle a_i|$ from the left, we find

$$\langle a_i|\beta\rangle = \langle a_i|\hat{X}|\alpha\rangle = \sum_{j=1}^{r} \langle a_i|\hat{X}|a_j\rangle\langle a_j|\alpha\rangle \,, \quad i = 1,\ldots,r \,, \tag{8.138}$$

which may be written as

$$\begin{pmatrix} \langle a_1|\beta\rangle \\ \langle a_2|\beta\rangle \\ \vdots \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots \\ X_{21} & X_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \langle a_1|\alpha\rangle \\ \langle a_2|\alpha\rangle \\ \vdots \end{pmatrix} \,. \tag{8.139}$$

So, $|\alpha\rangle$ can be represented by a column matrix:

$$|\alpha\rangle \doteq \begin{pmatrix} \langle a_1|\alpha\rangle \\ \langle a_2|\alpha\rangle \\ \vdots \end{pmatrix} \,. \tag{8.140}$$

Let us see what happens with

$$\langle\beta| = \langle\alpha|\hat{X} \,. \tag{8.141}$$

Multiplying this equation by $|a_i\rangle$ from the right, we find

$$\langle\beta|a_i\rangle = \langle\alpha|\hat{X}|a_i\rangle = \sum_{j=1}^{r} \langle\alpha|a_j\rangle\langle a_j|\hat{X}|a_i\rangle \,, \quad i = 1,\ldots,r \,. \tag{8.142}$$

This may be written as

$$\big(\langle\beta|a_1\rangle\ \langle\beta|a_2\rangle\ \cdots\big) = \big(\langle\alpha|a_1\rangle\ \langle\alpha|a_2\rangle\ \cdots\big)\begin{pmatrix} X_{11} & X_{12} & \cdots \\ X_{21} & X_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \tag{8.143}$$

indicating that $\langle\alpha|$ can be represented by a row matrix:

$$\langle\alpha| \doteq \big(\langle\alpha|a_1\rangle \quad \langle\alpha|a_2\rangle\cdots\big) = \big(\langle a_1|\alpha\rangle^* \quad \langle a_2|\alpha\rangle^*\cdots\big)\ . \tag{8.144}$$

Exercise 8.21. Express $\langle\alpha|\beta\rangle$ using matrices. ///

The **trace** of an operator $\hat{A}$ is defined as the sum of all diagonal elements of the matrix that represents $\hat{A}$:

$$\mathrm{Tr}\{\hat{A}\} := \sum_{i=1}^{r}\langle a_i|\hat{A}|a_i\rangle\ . \tag{8.145}$$

$\mathrm{Tr}\{\hat{A}\}$ is independent of the basis. To see this, let $\{|a_1\rangle,\ldots,|a_r\rangle\}$ and $\{|a'_1\rangle,\ldots,|a'_r\rangle\}$ be distinct bases. Then,

$$\mathrm{Tr}\{\hat{A}\} = \sum_i\langle a_i|\hat{A}|a_i\rangle = \sum_{i,j}\langle a_i|a'_j\rangle\langle a'_j|\hat{A}|a_i\rangle = \sum_{i,j}\langle a'_j|\hat{A}|a_i\rangle\langle a_i|a'_j\rangle = \sum_j\langle a'_j|\hat{A}|a'_j\rangle\ . \tag{8.146}$$

Exercise 8.22. Show that

$$\mathrm{Tr}\{\hat{A}\hat{B}\} = \mathrm{Tr}\{\hat{B}\hat{A}\}\ . \tag{8.147}$$

///

8.3.4 Commuting Observables

An observable $\hat{A}$ is *diagonal* in the A-representation. That is,

$$\langle a_i|\hat{A}|a_j\rangle = \langle a_i|a_j|a_j\rangle = a_j\langle a_i|a_j\rangle = a_j\delta_{ij}\ , \tag{8.148}$$

where the last step follows from (8.123). Because the basis $\{|a_1\rangle,\ldots,|a_r\rangle\}$ is orthonormal by construction, this equation holds even if some of the eigenvalues are degenerate. What can we say about the A-representation of some other observable $\hat{B}$ that acts in V_k? The following theorem provides a partial answer to this question.

Theorem 8.4. *Let $\hat{A}$ and $\hat{B}$ be observables and suppose that none of the eigenvalues of $\hat{A}$ is degenerate. If $\hat{A}$ and $\hat{B}$ commute, that is, if $[\hat{A},\hat{B}] = 0$, $\hat{B}$ is diagonal in the A-representation*

Proof.

$$0 = \langle a_i|[\hat{A},\hat{B}]|a_j\rangle = \langle a_i|(\hat{A}\hat{B}-\hat{B}\hat{A})|a_j\rangle = \langle a_i|\hat{A}\hat{B}|a_j\rangle - \langle a_i|\hat{B}\hat{A}|a_j\rangle\ . \tag{8.149}$$

We recall that $|a_i\rangle$ is an eigenket of $\hat{A}$ corresponding to the eigenvalue a_i:

$$\hat{A}|a_i\rangle = a_i|a_i\rangle \ , \tag{8.150}$$

the adjoint of which reads

$$\langle a_i|\hat{A} = a_i\langle a_i| \ , \tag{8.151}$$

where we used the fact that $\hat{A}$ is Hermitian and that a_j is a real number. It follows that

$$0 = \langle a_i|[\hat{A},\hat{B}]|a_j\rangle = a_i\langle a_i|\hat{B}|a_j\rangle - \langle a_i|\hat{B}|a_j\rangle a_j = (a_i - a_j)\langle a_i|\hat{B}|a_j\rangle \ . \tag{8.152}$$

Because the eigenvalues of $\hat{A}$ are nondegenerate by assumption, $i \neq j$ implies $a_i \neq a_j$ and hence $\langle a_i|\hat{B}|a_j\rangle = 0$. □

But, if $\hat{B}$ is diagonal in the A-representation, the effect of $\hat{B}$ on a basis ket $|a_i\rangle$ should be just a multiplication by a number, and hence we have the following theorem:

Theorem 8.5. *Let $\hat{A}$ and $\hat{B}$ be observables satisfying $[\hat{A},\hat{B}] = 0$ and suppose that the eigenvalues of $\hat{A}$ are all nondegenerate. Then, $|a_i\rangle$ is also an eigenket of $\hat{B}$.*

Proof. Using the closure relation,

$$\hat{B}|a_i\rangle = \sum_j |a_j\rangle\langle a_j|\hat{B}|a_i\rangle \ . \tag{8.153}$$

Since the eigenvalues of $\hat{A}$ are nondegenerate, Theorem 8.4 applies and we obtain

$$\hat{B}|a_i\rangle = |a_i\rangle\langle a_i|\hat{B}|a_i\rangle \ . \tag{8.154}$$

That is, $|a_i\rangle$ is an eigenket of $\hat{B}$ corresponding to the eigenvalue $\langle a_i|\hat{B}|a_i\rangle$. □

Let us summarize the content of Theorems 8.3–8.5 here. *In quantum mechanics, we are interested in an observable whose eigenkets span the vector space in which it acts. The eigenkets may be made orthonormal using Gram–Schmidt orthogonalization scheme if necessary.* If *none of the eigenvalues of $\hat{A}$ are degenerate, the basis kets so constructed are simultaneous eigenkets of $\hat{A}$ and all other operators that commute with $\hat{A}$. These operators, including $\hat{A}$, are diagonal in the A-representation.*

What happens to this very nice result if there is a degeneracy? Even then, by means of what is called a **complete set of commuting observables**, the simultaneous eigenkets can still be found and the orthonormal basis can still be constructed using them. The observable in the set are all diagonal in this basis. This is the topic of next optional subsection. No harm will come from omitting it if you choose to accept the claim just made.

8.3.5 †*Degenerate Eigenvalues*

Now, suppose that $\hat{A}$ has an s-fold degenerate eigenvalue a_1 and that other eigenvalues $a_{s+1}, \ldots, a_r$ are nondegenerate. A generalization to cases with multiple degenerate eigenvalues will be trivial. If $[\hat{A}, \hat{B}] = 0$, how does the matrix of $\hat{B}$ look like in the A-representation? We label the eigenkets corresponding to a_1 as $|a^1\rangle, \ldots, |a^s\rangle$ and those corresponding to $a_{s+1}, \ldots, a_r$ as $|a_{s+1}\rangle, \ldots, |a_r\rangle$. We suppose that these kets are, *by construction*, orthonormal. Looking at the last step in the proof of Theorem 8.4, we see that

$$\langle a_i|\hat{B}|a_j\rangle = 0 \text{ if } i \neq j \,, \; \langle a_i|\hat{B}|a^j\rangle = 0 \,, \text{ and } \langle a^i|\hat{B}|a_j\rangle = 0 \,. \tag{8.155}$$

However, the theorem is silent about $\langle a^i|\hat{B}|a^j\rangle$.

For various computations, it would be very convenient if $\hat{B}$ is fully diagonal in the A-representation. Can we somehow accomplish this by properly choosing an orthonormal basis? We would still like to keep using the eigenkets of $\hat{A}$ as the basis kets, of course, since at least $\hat{A}$ is diagonal then.

We observe that (8.155) still holds even if we replace $|a^1\rangle, \ldots, |a^s\rangle$ by any of their linear combinations. By Theorem 8.3b, $|a_j\rangle$ is orthogonal to any linear combination of $|a^1\rangle, \ldots, |a^s\rangle$. If we choose appropriate linear combinations and use them with $|a_{s+1}\rangle, \ldots, |a_r\rangle$ as the basis kets, $\hat{B}$ might become fully diagonal.

To examine this possibility, we confine our attention to the eigensubspace V_{1k} of a_1, that is, the vector space spanned by eigenkets $|a^1\rangle, \ldots, |a^s\rangle$ corresponding to the eigenvalue a_1. We recall that, by construction, the set $\{|a^1\rangle, \ldots, |a^s\rangle\}$ is an orthonormal basis in V_{k1}. As we just saw, we are free to choose any other orthonormal basis without affecting (8.155). Let us denote this new orthonormal basis in V_{k1} by $\{|a_1\rangle, \ldots, |a_s\rangle\}$. Our goal is to construct this new basis from the old. But, this can be accomplished by a linear operator $\hat{U}$ acting in V_{k1}:

$$\hat{U} := \sum_{k=1}^{s} |a_k\rangle\langle a^k| \,. \tag{8.156}$$

In fact,

$$\hat{U}|a^i\rangle = \sum_{k=1}^{s} |a_k\rangle\langle a^k|a^i\rangle = \sum_{k=1}^{s} |a_k\rangle\delta_{ki} = |a_i\rangle \tag{8.157}$$

as desired.

Because we insist that new and old bases are both orthonormal, $\hat{U}$ satisfies (8.108), and hence is a unitary operator. We can easily confirm this:

$$\hat{U}\hat{U}^\dagger = \sum_{i,j=1}^{s} |a_i\rangle\langle a^i|a^j\rangle\langle a_j| = \sum_{i,j=1}^{s} |a_i\rangle\delta_{ij}\langle a_j| = \sum_{i=1}^{s} |a_i\rangle\langle a_i| = \hat{I} \,, \tag{8.158}$$

where we used the closure relation in V_{k1}. Similarly for $\hat{U}^\dagger\hat{U} = \hat{I}$.

Of course, $|a_k\rangle$ occurring in the definition for $\hat{U}$ *is* the very unknown we are trying to find. So, (8.156) is not an explicit prescription for finding $\hat{U}$.

How do we find $\hat{U}$ then? What we wish to accomplish with this $\hat{U}$ is to find the new basis in which $\hat{B}$ is diagonal:

$$\langle a_i|\hat{B}|a_j\rangle = \lambda\,\delta_{ij} = \lambda\,\langle a_i|a_j\rangle\ , \quad i,j = 1,\ldots,s\ . \tag{8.159}$$

in which λ is also unknown at this point. Because this equation must hold for any $\langle a_i|$, our goal is equivalent to ensuring

$$\hat{B}|a_j\rangle = \lambda\,|a_j\rangle\ . \tag{8.160}$$

That is, the new basis kets we seek are all eigenvectors of $\hat{B}$.

The method for finding eigenvalues and the corresponding eigenkets are as follows. We multiply (8.160) by $\langle a^i|$ from the left and rewrite the resulting equation using the closure relations:

$$\sum_{k=1}^{s} \langle a^i|\hat{B}|a^k\rangle\langle a^k|a_j\rangle = \lambda\,\langle a^i|a_j\rangle\ . \tag{8.161}$$

Written in terms of matrices, this reads

$$\begin{pmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & \ddots & \vdots \\ B_{s1} & \cdots & B_{ss} \end{pmatrix} \begin{pmatrix} \langle a^1|a_j\rangle \\ \vdots \\ \langle a^s|a_j\rangle \end{pmatrix} = \lambda \begin{pmatrix} \langle a^1|a_j\rangle \\ \vdots \\ \langle a^s|a_j\rangle \end{pmatrix}\ . \tag{8.162}$$

So that this equation has the nontrivial solutions, that is, for the solutions other than $\langle a^i|a_j\rangle = 0$ for all i to exist, the determinant of the operator $\hat{B} - \lambda\hat{I}$ must vanish:

$$\begin{vmatrix} B_{11}-\lambda & \cdots & B_{1s} \\ \vdots & \ddots & \vdots \\ B_{s1} & \cdots & B_{ss}-\lambda \end{vmatrix} = 0\ . \tag{8.163}$$

The determinant is the sth order polynomial of λ, from which we find s roots $\lambda = b_1,\ldots,b_s$. For each $\lambda = b_j$ so determined, we solve (8.162) to find the eigenvector

$$\begin{pmatrix} \langle a^1|a_j\rangle \\ \vdots \\ \langle a^s|a_j\rangle \end{pmatrix}\ . \tag{8.164}$$

When this is completed for each b_j ($j = 1,\ldots,s$), we have the matrix representation of $\hat{U}$:

$$\langle a^i|\hat{U}|a^j\rangle = \sum_{k=1}^{s} \langle a^i|a_k\rangle\langle a^k|a^j\rangle\ . \tag{8.165}$$

We recall that the newly determined basis vectors span V_{k1} and they are all eigenvectors of $\hat{A}$ corresponding to a_1. Now, by construction, $|a_j\rangle \in V_{k1}$ is also an eigenvector of $\hat{B}$ corresponding to b_j. If s solutions of (8.163) are all distinct, then $|a_1\rangle$, $\ldots$, $|a_s\rangle$ are orthogonal by Theorem 8.3b applied to $\hat{B}$. Under the same condition, the pair of numbers a_1 and b_j, taken together, specifies the eigenket completely, which is then denoted by $|a_1 b_j\rangle$. We say that $\hat{B}$ *resolves* the s-fold degeneracy of $\hat{A}$ completely.

The eigenkets $|a_{s+1}\rangle, \ldots, |a_r\rangle$ are also eigenkets of $\hat{B}$ by Theorem 8.5 and may be denoted in the same manner, even though the label b_j $(j = s+1, \ldots, r)$ will be redundant for them. Thus, when s-fold degeneracy of $\hat{A}$ is completely resolved by $\hat{B}$, we have the orthonormal basis

$$\{|a_1 b_1\rangle, \ldots, |a_1 b_s\rangle, |a_{s+1} b_{s+1}\rangle, \ldots, |a_r b_r\rangle\} \tag{8.166}$$

of the original vector space V_k that consists of the simultaneous eigenkets of $\hat{A}$ and $\hat{B}$.

If not all of s solutions of (8.163) are distinct, then there will be multiple eigenkets corresponding to a given pair of numbers, a_1 and one of $b_1, \ldots, b_s$. In this case, we can introduce additional operators that commute with both $\hat{A}$, $\hat{B}$, and with each other to further resolve the degeneracy. When all degeneracy is resolved in this manner, we have what is called a **complete set of commuting observables** $\hat{A}_1, \ldots, \hat{A}_c$, for which $[\hat{A}_i, \hat{A}_j] = 0$ for *all* pairs of $i, j = 1, \ldots, c$, and their simultaneous eigenkets form an orthonormal basis of V_k.

8.4 Postulates of Quantum Mechanics

Now we are *finally* ready to state the fundamental postulates of quantum mechanics.

Postulate 1: The state of the system is represented by the "direction" of a ket, that is, $|\phi\rangle$ and $c|\phi\rangle$ $(c \neq 0)$ represent the same state.

Postulate 2: A dynamical variable A, such as energy, position, and momentum, is *represented* by an observable $\hat{A}$.

Postulate 3: A measurement of A yields one of the eigenvalues a_1, a_2, $\cdots$, a_n, of $\hat{A}$. Equation (8.130) implies that, if a measurement of A yields a_i, then the measurement of $f(A)$ yields $f(a_i)$, which is a very natural thing to demand of our theory.

Postulate 4: If the system is in a state $|\phi\rangle$, the probability of obtaining the value a_i upon measuring A is given by

$$p(a_i) = |\langle a_i|\phi\rangle|^2 \tag{8.167}$$

provided that a_i is nondegenerate and $|\phi\rangle$ is normalized. If $|\phi\rangle$ is not normalized, this expression must be replaced by

$$p(a_i) = \frac{|\langle a_i|\phi\rangle|^2}{\langle\phi|\phi\rangle}\,. \tag{8.168}$$

Note that the value of $p(a_i)$ is independent of whether $|\phi\rangle$ is normalized or not, as is demanded by Postulate 1. In what follows, state kets are assumed to be normalized.

If a_i is s-fold degenerate, then

$$p(a_i) = \sum_{k=1}^{s} |\langle a_i, k|\phi\rangle|^2 \ , \tag{8.169}$$

where $|a_i, k\rangle$ is the kth eigenket corresponding to a_i.

Example 8.10. Eigenstate and the outcome of a measurement: Suppose that the system is in state $|\phi\rangle$. A measurement of a dynamical variable A of the system yields a_i with certainty *if and only if* $|\phi\rangle = |a_i\rangle$.

To see this, let us suppose first that $|\phi\rangle = |a_i\rangle$ in (8.167). Then,

$$p(a_i) = |\langle a_i|a_i\rangle|^2 = 1 \tag{8.170}$$

since the eigenkets are normalized. On the other hand, if $|\phi\rangle \neq |a_i\rangle$, then

$$|\phi\rangle = C_i|a_i\rangle + \sum_{j\neq i} C_j|a_j\rangle \ , \tag{8.171}$$

where $C_j \neq 0$ for at least one $C_{j\neq i}$. Since $|\phi\rangle$ is normalized,

$$\langle\phi|\phi\rangle = |C_i|^2 + \sum_{j\neq i} |C_j|^2 = 1 \ . \tag{8.172}$$

Thus,

$$p(a_i) = |C_i|^2 = 1 - \sum_{j\neq i} |C_j|^2 < 1 \ . \tag{8.173}$$

This proves the assertion.

Now, let us suppose we measured A of a system in state $|\phi\rangle$ and found the value a_j. If we wait around too long, the state of the system may evolve with time. But, if we measure A *immediately* after the first measurement, the system has no time to evolve. In this case, it seems natural to expect that we should obtain the same result a_j *with certainty*. It follows from Example 8.10 that the state of the system immediately after the measurement which yielded a_j must be $|a_j\rangle$. One may say that the state of the system has changed *abruptly* from $|\phi\rangle$ to $|a_j\rangle$ by the very act of measuring A. This is the so-called collapse postulate of quantum mechanics. For alternative views, see Refs. [1, 8].

Exercise 8.23. Suppose that we measure A right after preparing the system in the state $|\phi\rangle$. If we repeat this procedure many times, including the preparation of the

system in the state $|\phi\rangle$, we can find the average value of our many measurements, which we denote by $\langle A\rangle_{\rm QM}$. Show that

$$\langle A\rangle_{\rm QM} = \langle\phi|\hat{A}|\phi\rangle\ , \tag{8.174}$$

where $|\phi\rangle$ is normalized. ///

8.5 ‡Uncertainty Principle

Let us consider two observables $\hat{A}$ and $\hat{C}$. We assume that $\hat{A}$ and $\hat{C}$ commute, that is, $[\hat{A},\hat{C}] = 0$ and that $\hat{A}$ is nondegenerate. As we saw in Sect. 8.3.4, $\hat{A}$ and $\hat{C}$ has the simultaneous eigenkets that form an orthonormal basis $\{|a_1c_1\rangle,\ldots,|a_rc_r\rangle\}$ of the vector space V_k.

The state ket $|\phi\rangle \in V_k$ of the system can be written using this basis as

$$|\phi\rangle = \sum_{i=1}^{r} \lambda_i |a_ic_i\rangle. \tag{8.175}$$

What would the outcome be of a measurement of the dynamical variable A? According to the postulate of quantum mechanics, the outcome is one of a_i $(i = 1,\ldots,r)$ with probability $|\lambda_i|^2$. The state ket right after the measurement that gave a_i is $|a_ic_i\rangle$. Thus, if we measure C now, the outcome is c_i with certainty and the state ket after this second measurement again is $|a_ic_i\rangle$. The subsequent measurements of $\hat{A}$ and $\hat{C}$ in any order will give a_i and c_i with certainty, provided that all measurements are done in rapid succession before the state ket has a time to evolve. We may say that the values of A and C can be determined simultaneously.

The situation is quite different if we consider the measurements of A and B when $[\hat{A},\hat{B}] \neq 0$. In this case, the measurement of A gives a_i and the state ket right after this measurement is still $|a_ic_i\rangle$ as before. If we measure A again, we are certain to get a_i. However, $|a_ic_i\rangle$ is *not* an eigenket of $\hat{B}$. Denoting the orthonormal basis consisting of the eigenkets of $\hat{B}$ by $\{|b_1\rangle,\ldots,|b_r\rangle\}$, we write

$$|a_ic_i\rangle = \sum_{i=1}^{r} \lambda_i' |b_i\rangle\ . \tag{8.176}$$

Thus, the subsequent measurement will give one of b_i $(i = 1,\ldots,r)$ with probability $|\lambda_i'|^2$. Still, the resulting ket $|b_i\rangle$, though an eigenket of $\hat{B}$, is *not* an eigenket of $\hat{A}$:

$$|b_i\rangle = \sum_{i=1}^{r} \lambda_i'' |a_ic_i\rangle \tag{8.177}$$

and we are no longer certain of the outcome of the future measurement of A. Thus, if $[\hat{A},\hat{B}] \neq 0$, no amount of repeated measurements of A and B will ever give us

any certainty regarding the future measurements of *both* of these quantities. That is, values of A and B cannot be determined simultaneously.

The measurements of A and C are said to be **compatible**, while those of A and B are **incompatible**. From many repeated measurements of A and those of B, all done on the identically prepared state $|\phi\rangle$, however, we can still compute the average values, $\langle A\rangle_{\rm QM}$ and $\langle B\rangle_{\rm QM}$, and the probability distributions of all possible outcomes around these averages.

To characterize the widths of these distributions, let us take $\hat{A}$ and define

$$\hat{A}_\Delta := \hat{A} - \langle A\rangle_{\rm QM}\hat{I}\,. \tag{8.178}$$

Then,

$$\hat{A}_\Delta^2 = \hat{A}^2 - 2\langle A\rangle_{\rm QM}\hat{A} + \langle A\rangle_{\rm QM}{}^2\hat{I}\,, \tag{8.179}$$

and hence

$$\langle {A_\Delta}^2\rangle_{\rm QM} = \langle\phi|{\hat{A}_\Delta}^2|\phi\rangle = \langle A^2\rangle_{\rm QM} - \langle A\rangle_{\rm QM}{}^2\,. \tag{8.180}$$

This quantity is called the **dispersion** of A and serves as a measure for the width of the distribution for A. Similarly for $\hat{B}$.

The dispersions of noncommuting observables are subject to a fundamental limitation imposed by the **Heisenberg uncertainty principle**:

$$\langle {A_\Delta}^2\rangle_{\rm QM}\langle {B_\Delta}^2\rangle_{\rm QM} \geq \frac{1}{4}|\langle[\hat{A},\hat{B}]\rangle_{\rm QM}|^2\,. \tag{8.181}$$

Equation (3.155) responsible for the $1/h^{3N}$ factor for a classical statistical mechanical partition function stems from this principle as we shall see in Sect. 8.7.2.

The proof of the uncertainty principle requires three ingredients:

a. The Schwarz inequality

$$|\langle\alpha|\beta\rangle|^2 \leq \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle\,, \tag{8.182}$$

the proof of which is similar to that given in Appendix A.5 for the case of ordinary vectors in three-dimensional space. In particular, let $|\gamma\rangle := |\alpha\rangle + \lambda|\beta\rangle$ and note that

$$\langle\gamma|\gamma\rangle = \langle\alpha|\alpha\rangle + \lambda\langle\alpha|\beta\rangle + \lambda^*\langle\beta|\alpha\rangle + \lambda\lambda^*\langle\beta|\beta\rangle \geq 0 \tag{8.183}$$

for any $\lambda \in \mathbb{C}$. Setting $\lambda = -\langle\beta|\alpha\rangle/\langle\beta|\beta\rangle$ we obtain the inequality in question.

b. The average of a Hermitian operator $\hat{A}$ is a real number as is obvious from Theorem 8.3.

c. An operator satisfying $\hat{X}^\dagger = -\hat{X}$ is said to be **anti-Hermitian**. The average of an anti-Hermitian operator is purely imaginary. In fact, using (8.85),

$$\langle\phi|\hat{X}|\phi\rangle^* = \langle\phi|\hat{X}^\dagger|\phi\rangle = -\langle\phi|\hat{X}|\phi\rangle\,. \tag{8.184}$$

Let $\alpha = \hat{A}_\Delta|\phi\rangle$ and $\beta = \hat{B}_\Delta|\phi\rangle$ in the Schwarz inequality. This gives

$$\langle {A_\Delta}^2\rangle_{\rm QM}\langle {B_\Delta}^2\rangle_{\rm QM} \geq \left|\langle\phi|\hat{A}_\Delta\hat{B}_\Delta|\phi\rangle\right|^2 . \tag{8.185}$$

To transform the right-hand side, we write

$$\hat{A}_\Delta\hat{B}_\Delta = \frac{1}{2}[\hat{A}_\Delta,\hat{B}_\Delta]_- + \frac{1}{2}[\hat{A}_\Delta,\hat{B}_\Delta]_+ , \tag{8.186}$$

where

$$[\hat{X},\hat{Y}]_\pm := \hat{X}\hat{Y} \pm \hat{Y}\hat{X} . \tag{8.187}$$

The subscript "$-$" temporarily denotes the now familiar commutator, while the subscript "+" defines the so-called **anti-commutator**.

It is straightforward to show that $[\hat{A}_\Delta,\hat{B}_\Delta]_- = [\hat{A},\hat{B}]_-$ is anti-Hermitian, while $[\hat{A}_\Delta,\hat{B}_\Delta]_+$ is Hermitian. In view of the items b and c above, we conclude that

$$\begin{aligned}\langle {A_\Delta}^2\rangle_{\rm QM}\langle {B_\Delta}^2\rangle_{\rm QM} &\geq \frac{1}{4}\left|\langle\phi|[\hat{A},\hat{B}]_-|\phi\rangle\right|^2 + \frac{1}{4}\left|\langle\phi|[\hat{A}_\Delta,\hat{B}_\Delta]_+|\phi\rangle\right|^2 \\ &\geq \frac{1}{4}\left|\langle\phi|[\hat{A},\hat{B}]_-|\phi\rangle\right|^2 ,\end{aligned} \tag{8.188}$$

which is just (8.181).

8.6 ‡Operator with Continuous Spectrum

Certain quantities, such as the position and the momentum of a particle, are expected to be capable of continuous variation. In accordance with the postulates of quantum mechanics, the observables corresponding to these quantities must possess continuous spectrum of eigenvalues. Let $\hat{\xi}$ denote such an observable. Then,

$$\hat{\xi}|\xi\rangle = \xi|\xi\rangle , \tag{8.189}$$

where ξ is the eigenvalue of $\hat{\xi}$ and $|\xi\rangle$ is the corresponding eigenket. As it turns out, the norm of $|\xi\rangle$ is infinite if ξ belongs to a continuous spectrum, and hence $|\xi\rangle$ cannot be normalized. Instead of now familiar

$$\langle a_i|a_j\rangle = \delta_{ij} \tag{8.190}$$

we had for eigenkets belonging to eigenvalues of discrete spectrum, the orthonormality condition becomes

$$\langle\xi'|\xi''\rangle = \delta(\xi' - \xi'') , \tag{8.191}$$

where δ is the Dirac δ-function. The closure relation must be replaced by

$$\int |\xi\rangle\langle\xi| \mathrm{d}\xi = \hat{I} \,. \tag{8.192}$$

In this book, we shall take these results for granted. See Ref. [1] for a brief discussion of a more satisfactory treatment.

Using (8.192), we see immediately that

$$|\alpha\rangle = \int |\xi\rangle\langle\xi|\alpha\rangle \mathrm{d}\xi \,, \tag{8.193}$$

and that

$$\langle\alpha|\beta\rangle = \langle\alpha| \left[\int |\xi\rangle\langle\xi| d\xi\right] |\beta\rangle = \int \langle\alpha|\xi\rangle\langle\xi|\beta\rangle \mathrm{d}\xi \,. \tag{8.194}$$

From (8.189) and (8.191), we have

$$\langle\xi'|\hat{\xi}|\xi''\rangle = \xi'' \delta(\xi' - \xi'') \,. \tag{8.195}$$

In place of (8.167), we have

$$|\langle\xi|\phi\rangle|^2 \mathrm{d}\xi \tag{8.196}$$

as the probability that the measurement of ξ yields the value between ξ and $\xi + \mathrm{d}\xi$.

8.6.1 ‡*Position Operator and Position Eigenkets*

Consider a particle confined to a one-dimensional space and place the x-axis along this space. In accordance with the postulates of quantum mechanics, the position is represented by an observable, which we call a **position operator** $\hat{x}$. If the particle is exactly at x, the measurement of its position should give us x. Thus,

$$\hat{x}|x\rangle = x|x\rangle \,. \tag{8.197}$$

This concept of position operator and its eigenkets may be extended to a three-dimensional space. In doing so, we assume the existence of a *simultaneous* eigenket $|xyz\rangle$ of position operators $\hat{x}$, $\hat{y}$, and $\hat{z}$ so that three components of the position vector of the particle can be determined simultaneously. (See Sect. 8.5.)

In what follows, it proves convenient to use x_1, x_2, x_3 in place of more familiar x, y, z. In this new notation, we have

$$[\hat{x}_i, \hat{x}_j] = 0 \quad i, j = 1, 2, 3 \tag{8.198}$$

and

$$\hat{x}_1|x_1x_2x_3\rangle = x_1|x_1x_2x_3\rangle \,, \hat{x}_2|x_1x_2x_3\rangle = x_2|x_1x_2x_3\rangle \,, \hat{x}_3|x_1x_2x_3\rangle = x_3|x_1x_2x_3\rangle \,. \tag{8.199}$$

These three equations may be written more compactly as

$$\hat{\boldsymbol{x}}|\boldsymbol{x}\rangle = \boldsymbol{x}|\boldsymbol{x}\rangle \ . \tag{8.200}$$

Because $\hat{\boldsymbol{x}}$ is assumed to be an observable, $\{|\boldsymbol{x}\rangle\}$ forms a basis and hence the closure relation

$$\int |\boldsymbol{x}\rangle\langle\boldsymbol{x}|\mathrm{d}x = \hat{I} \tag{8.201}$$

holds. As a result, any state ket can be expressed as a linear combination of $|\boldsymbol{x}\rangle$:

$$|\alpha\rangle = \int |\boldsymbol{x}\rangle\langle\boldsymbol{x}|\alpha\rangle\mathrm{d}\boldsymbol{x} \ , \tag{8.202}$$

in which the complex-valued function $\phi_\alpha(\boldsymbol{x}) := \langle\boldsymbol{x}|\alpha\rangle$ of $\boldsymbol{x}$ is called the **wave function**. A scalar product between two kets is now expressed as

$$\langle\beta|\alpha\rangle = \int \langle\beta|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\alpha\rangle\mathrm{d}\boldsymbol{x} = \int \phi^*_\beta(\boldsymbol{x})\phi_\alpha(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \ . \tag{8.203}$$

This is the scalar product between two complex-valued function we saw in Example 8.7.

8.7 †Linear Translation

Suppose that a system is in the state represented by $|\alpha\rangle \in V_k$. The system may be just a particle, for example. If we take this system and move it by $\Delta\boldsymbol{x}$ as a whole, the resulting ket, let us call it $|\alpha'\rangle \in V_k$, in general will be different from the original one. Insofar as we may perform such operation for any $|\alpha\rangle \in V_k$, there should be an operator $\hat{T}_{\Delta\boldsymbol{x}}$ representing this linear translation. Thus,

$$|\alpha'\rangle = \hat{T}_{\Delta\boldsymbol{x}}|\alpha\rangle \ . \tag{8.204}$$

We call $\hat{T}_{\Delta\boldsymbol{x}}$ the **translation operator** and demand that it is linear. In this section, we seek for the explicit expression for this operator and study its connection to $\hat{\boldsymbol{x}}$.

8.7.1 †Properties of Linear Translation

On physical grounds, we expect $\hat{T}_{\Delta\boldsymbol{x}}$ to satisfy the following conditions:

a. $\hat{T}_{\Delta\boldsymbol{x}}$ preserves the norm of $|\alpha\rangle$. Thus,

$$\langle\alpha'|\alpha'\rangle = \langle\alpha|\hat{T}^\dagger_{\Delta\boldsymbol{x}}\hat{T}_{\Delta\boldsymbol{x}}|\alpha\rangle = \langle\alpha|\alpha\rangle \ , \tag{8.205}$$

for which it is sufficient that

$$\hat{T}^{\dagger}_{\Delta \boldsymbol{x}} \hat{T}_{\Delta \boldsymbol{x}} = \hat{I} \,. \tag{8.206}$$

To understand the reason for demanding this property, recall that

$$\langle \alpha | \alpha \rangle = \sum_{i=1}^{r} \langle \alpha | a_i \rangle \langle a_i | \alpha \rangle = \sum_{i=1}^{r} |\langle a_i | \alpha \rangle|^2 \tag{8.207}$$

is the probability that the measurement of some dynamical variable A gives any of its possible values as an outcome. This probability has to be one and should not change upon linear translation of the system.

b. Two successive linear translations, $\Delta \boldsymbol{x}_1$ followed by $\Delta \boldsymbol{x}_2$, should produce a state identical to the one produced by a single linear translation by $\Delta \boldsymbol{x}_1 + \Delta \boldsymbol{x}_2$. Thus,

$$\hat{T}_{\Delta \boldsymbol{x}_2} \hat{T}_{\Delta \boldsymbol{x}_1} = \hat{T}_{\Delta \boldsymbol{x}_1 + \Delta \boldsymbol{x}_2} \,. \tag{8.208}$$

c. The state of the system will not change at all if no translation is performed. We also do not expect any discontinuity to be present at $\Delta \boldsymbol{x} = \boldsymbol{0}$. Thus,

$$\lim_{\Delta \boldsymbol{x} \to \boldsymbol{0}} \hat{T}_{\Delta \boldsymbol{x}} = \hat{T}_{\Delta \boldsymbol{x} = \boldsymbol{0}} = \hat{I} \,. \tag{8.209}$$

Let us explore some immediate consequence of these requirements. Since $\Delta \boldsymbol{x}_1 + \Delta \boldsymbol{x}_2 = \Delta \boldsymbol{x}_2 + \Delta \boldsymbol{x}_1$, (8.208) implies that

$$\hat{T}_{\Delta \boldsymbol{x}_2} \hat{T}_{\Delta \boldsymbol{x}_1} = \hat{T}_{\Delta \boldsymbol{x}_1} \hat{T}_{\Delta \boldsymbol{x}_2} \,, \tag{8.210}$$

that is,

$$[\hat{T}_{\Delta \boldsymbol{x}_1}, \hat{T}_{\Delta \boldsymbol{x}_2}] = 0 \,. \tag{8.211}$$

Let $\Delta \boldsymbol{x}_2 = \Delta \boldsymbol{x}$ and $\Delta \boldsymbol{x}_1 = -\Delta \boldsymbol{x}$ in (8.208), and apply (8.209). We find

$$\hat{T}_{\Delta \boldsymbol{x}} \hat{T}_{-\Delta \boldsymbol{x}} = \hat{I} \,. \tag{8.212}$$

Multiplying this equation by $\hat{T}^{\dagger}_{\Delta \boldsymbol{x}}$ from the left and using (8.206), we arrive at

$$\hat{T}_{-\Delta \boldsymbol{x}} = \hat{T}^{\dagger}_{\Delta \boldsymbol{x}} \,. \tag{8.213}$$

These properties of $\hat{T}_{\Delta \boldsymbol{x}}$ guide us in our search for a reasonable guess on its form.

To avoid unnecessary complications, let us confine our attention to an infinitesimal linear translation by $\delta \boldsymbol{x}$ so that higher order terms in $\delta \boldsymbol{x}$ can be safely ignored. Then, all of the properties we listed for the linear translation operator are satisfied by the following form of $\hat{T}_{\delta \boldsymbol{x}}$:

$$\hat{T}_{\delta \boldsymbol{x}} = \hat{I} - i \hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x} \,, \tag{8.214}$$

where $\hat{\boldsymbol{K}}$ is a Hermitian operator and

$$\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x} := \hat{K}_1 \delta x_1 + \hat{K}_2 \delta x_2 + \hat{K}_3 \delta x_3 \,. \tag{8.215}$$

We note that $\hat{T}_{\delta \boldsymbol{x}}$ should not change the dimension of a ket it is acting on. Consequently, $\hat{\boldsymbol{K}}$ has the dimension of inverse length.

Exercise 8.24. Prove that (8.214) satisfies (8.206), (8.208), and (8.209) to the first order of $\delta \boldsymbol{x}$. ///

8.7.2 †Commutation Relations

What effect would the linear translation have on the position eigenket $|\boldsymbol{x}\rangle$? By answering this question, we will be lead to discover the commutation relation between the position operator $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{K}}$. Let $|\alpha'\rangle$ denote the state ket obtained by linearly translating $|\alpha\rangle$ by $\delta \boldsymbol{x}$. According to the definition given in Sect. 8.6.1, the corresponding wave functions are given by

$$\phi_\alpha(\boldsymbol{x}) = \langle \boldsymbol{x}|\alpha\rangle \quad \text{and} \quad \phi_{\alpha'}(\boldsymbol{x}) = \langle \boldsymbol{x}|\alpha'\rangle = \langle \boldsymbol{x}|\hat{T}_{\delta \boldsymbol{x}}|\alpha\rangle \,, \tag{8.216}$$

respectively. Because the two states, $|\alpha\rangle$ and $|\alpha'\rangle$, are related by the displacement $\delta \boldsymbol{x}$, it seems utterly natural to expect that their corresponding wave functions should also be so related. That is, the value of $\phi_{\alpha'}$ at $\boldsymbol{x}$ is equal to that of ϕ_α at $\boldsymbol{x} - \delta \boldsymbol{x}$:

$$\phi_{\alpha'}(\boldsymbol{x}) = \phi_\alpha(\boldsymbol{x} - \delta \boldsymbol{x}) \,. \tag{8.217}$$

Using (8.216), we rewrite this equation as

$$\langle \boldsymbol{x}|\hat{T}_{\delta \boldsymbol{x}}|\alpha\rangle = \langle \boldsymbol{x} - \delta \boldsymbol{x}|\alpha\rangle \,. \tag{8.218}$$

Since $|\alpha\rangle \in V_k$ is arbitrary, we conclude that

$$\langle \boldsymbol{x}|\hat{T}_{\delta \boldsymbol{x}} = \langle \boldsymbol{x} - \delta \boldsymbol{x}| \,. \tag{8.219}$$

Taking the adjoint,

$$\hat{T}^\dagger_{\delta \boldsymbol{x}}|\boldsymbol{x}\rangle = |\boldsymbol{x} - \delta \boldsymbol{x}\rangle \,. \tag{8.220}$$

Recalling (8.213),

$$\hat{T}_{-\delta \boldsymbol{x}}|\boldsymbol{x}\rangle = |\boldsymbol{x} - \delta \boldsymbol{x}\rangle \,. \tag{8.221}$$

But because this equality holds for any $\delta \boldsymbol{x}$, it should hold when we replace $\delta \boldsymbol{x}$ by $-\delta \boldsymbol{x}$:

$$\hat{T}_{\delta \boldsymbol{x}}|\boldsymbol{x}\rangle = |\boldsymbol{x} + \delta \boldsymbol{x}\rangle \,, \tag{8.222}$$

Multiplying both sides of this equation by $\hat{\boldsymbol{x}}$, we find

$$\hat{\boldsymbol{x}}\hat{T}_{\delta \boldsymbol{x}}|\boldsymbol{x}\rangle = \hat{\boldsymbol{x}}|\boldsymbol{x} + \delta \boldsymbol{x}\rangle = (\boldsymbol{x} + \delta \boldsymbol{x})|\boldsymbol{x} + \delta \boldsymbol{x}\rangle \,, \tag{8.223}$$

where we used (8.200). On the other hand,

$$\hat{T}_{\delta \boldsymbol{x}}\hat{\boldsymbol{x}}|\boldsymbol{x}\rangle = \hat{T}_{\delta \boldsymbol{x}}\boldsymbol{x}|\boldsymbol{x}\rangle = \boldsymbol{x}\hat{T}_{\delta \boldsymbol{x}}|\boldsymbol{x}\rangle = \boldsymbol{x}|\boldsymbol{x} + \delta \boldsymbol{x}\rangle \,. \tag{8.224}$$

The second equality holds because $\boldsymbol{x}$ is simply a vector having real numbers as its components. The last step follows from (8.222). The last two equations indicate

$$[\hat{\boldsymbol{x}}, \hat{T}_{\delta \boldsymbol{x}}]|\boldsymbol{x}\rangle = \delta \boldsymbol{x}|\boldsymbol{x} + \delta \boldsymbol{x}\rangle \,. \tag{8.225}$$

Expanding $|\boldsymbol{x} + \delta \boldsymbol{x}\rangle$ around $\delta \boldsymbol{x} = \boldsymbol{0}$, and retaining the leading term only, we find

$$[\hat{\boldsymbol{x}}, \hat{T}_{\delta \boldsymbol{x}}]|\boldsymbol{x}\rangle = \delta \boldsymbol{x}|\boldsymbol{x}\rangle \,. \tag{8.226}$$

If a Taylor series expansion of a ket leaves you uncomfortable, we can proceed as follows:

$$\langle \phi|[\hat{\boldsymbol{x}}, \hat{T}_{\delta \boldsymbol{x}}]|\boldsymbol{x}\rangle = \langle \phi|\delta \boldsymbol{x}|\boldsymbol{x} + \delta \boldsymbol{x}\rangle = \delta \boldsymbol{x}\phi(\boldsymbol{x} + \delta \boldsymbol{x}) \approx \delta \boldsymbol{x}\phi(\boldsymbol{x}) = \delta \boldsymbol{x}\langle \phi|\boldsymbol{x}\rangle \,. \tag{8.227}$$

But since ϕ in this equation is arbitrary, this result implies (8.226). Furthermore, since $\{|\boldsymbol{x}\rangle\}$ forms a basis, (8.226) holds even if $|\boldsymbol{x}\rangle$ is replaced by an arbitrary ket. So, we conclude that

$$[\hat{\boldsymbol{x}}, \hat{T}_{\delta \boldsymbol{x}}] = \delta \boldsymbol{x} \,. \tag{8.228}$$

Using (8.214), we transform this equation as follows:

$$\begin{aligned} \hat{\boldsymbol{x}}(\hat{I} - i\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x}) - (\hat{I} - i\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x})\hat{\boldsymbol{x}} &= \delta \boldsymbol{x} \\ -i\hat{\boldsymbol{x}}(\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x}) + i(\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x})\hat{\boldsymbol{x}} &= \delta \boldsymbol{x} \\ \hat{\boldsymbol{x}}(\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x}) - (\hat{\boldsymbol{K}} \cdot \delta \boldsymbol{x})\hat{\boldsymbol{x}} &= i\delta \boldsymbol{x} \,. \end{aligned} \tag{8.229}$$

The first (or the x_1-) component of this equation reads

$$\hat{x}_1(\hat{K}_1 \delta x_1 + \hat{K}_2 \delta x_2 + \hat{K}_3 \delta x_3) - (\hat{K}_1 \delta x_1 + \hat{K}_2 \delta x_2 + \hat{K}_3 \delta x_3)\hat{x}_1 = i\delta x_1 \,. \tag{8.230}$$

Since $\delta \boldsymbol{x}$ is arbitrary, we conclude that

$$\hat{x}_1 \hat{K}_1 - \hat{K}_1 \hat{x}_1 = i \,, \quad \hat{x}_1 \hat{K}_2 - \hat{K}_2 \hat{x}_1 = 0 \,, \quad \text{and} \quad \hat{x}_1 \hat{K}_3 - \hat{K}_3 \hat{x}_1 = 0 \,. \tag{8.231}$$

Six additional equations follow from the other components of (8.229). The resulting (total of nine) equations can be summarized compactly as

$$[\hat{x}_i, \hat{K}_j] = i\delta_{ij} \,. \tag{8.232}$$

We introduce a Hermitian operator defined by

$$\hat{\boldsymbol{p}} = \hbar \hat{\boldsymbol{K}} \tag{8.233}$$

and call it the **momentum operator**, whose eigenvalue, by *definition*, is the **momentum**.

Justification of this seemingly arbitrary definition of momentum must be sought among the physical attributes of the eigenvalues predicted by the theory. See Sect. 8.15, for example. Here, we simply remark that $\hat{\boldsymbol{p}}$ does have the correct dimen-

sion of momentum because of the factor $\hbar := h/2\pi$, in which h is the Planck constant.

Now, we rewrite (8.232) in terms of the momentum operator as

$$[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij} \,. \tag{8.234}$$

The appearance of δ_{ij} is quite reasonable on physical grounds if we note that $\hat{\boldsymbol{p}}$ is a part of the linear translation operator. When you measure the x_i coordinate of a particle, it certainly matters whether or not you move the particle in the x_i-direction prior to the measurement, while moving it in the x_j-direction ($i \neq j$) should not have any effect. Substituting (8.234) into (8.181), we arrive at

$$\sqrt{\langle {x_{i\Delta}}^2 \rangle_{\mathrm{QM}} \langle {p_{j\Delta}}^2 \rangle_{\mathrm{QM}}} \geq \frac{\hbar}{2}\delta_{ij} \,, \tag{8.235}$$

which is essentially (3.155).

The commutation relations among $\hat{p}$'s are obtained from (8.211) and (8.214):

$$[\hat{p}_i, \hat{p}_j] = 0 \,. \tag{8.236}$$

Equations (8.198), (8.234), and (8.236) should be contrasted against their classical counterparts given in Exercise 1.17.

8.7.3 †*Momentum Eigenket*

Combining (8.214), (8.216), (8.233), we find

$$\phi_{\alpha'}(\boldsymbol{x}) = \left\langle \boldsymbol{x} \left| \hat{I} - \frac{i}{\hbar}\hat{\boldsymbol{p}} \cdot \delta\boldsymbol{x} \right| \alpha \right\rangle = \phi_\alpha(\boldsymbol{x}) - \frac{i}{\hbar}\delta\boldsymbol{x} \cdot \langle \boldsymbol{x}|\hat{\boldsymbol{p}}|\alpha\rangle \,. \tag{8.237}$$

Recall that we kept only up to the first order terms of $\delta\boldsymbol{x}$ in arriving at this equation. Under the same approximation, (8.217) gives

$$\phi_{\alpha'}(\boldsymbol{x}) = \phi_\alpha(\boldsymbol{x}) - \delta\boldsymbol{x} \cdot \nabla\phi_\alpha(\boldsymbol{x}) \,. \tag{8.238}$$

Comparing these two equations, and noting that $\delta\boldsymbol{x}$ is arbitrary, we find that

$$\langle \boldsymbol{x}|\hat{\boldsymbol{p}}|\alpha\rangle = -i\hbar\nabla\langle \boldsymbol{x}|\alpha\rangle \,. \tag{8.239}$$

On the right-hand side, $\langle \boldsymbol{x}|\alpha\rangle$ is the wave function for state $|\alpha\rangle$. On the left, we have the wave function corresponding to the ket $\hat{\boldsymbol{p}}|\alpha\rangle$, that is, the state $|\alpha\rangle$ after it is modified by the operator $\hat{\boldsymbol{p}}$. Equation (8.239) indicates that the effect of the momentum operator $\hat{\boldsymbol{p}}$ in the $\boldsymbol{x}$-representation is to take the derivative of the wave function. More pictorially,

$$\hat{\boldsymbol{p}} : \phi_\alpha(\boldsymbol{x}) \longrightarrow -i\hbar\nabla\phi_\alpha(\boldsymbol{x}) \,. \tag{8.240}$$

Using (8.239), we have

$$\langle\beta|\hat{\boldsymbol{p}}|\alpha\rangle = \int \langle\beta|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\boldsymbol{p}}|\alpha\rangle \mathrm{d}\boldsymbol{x} = \int \phi_\beta^*(\boldsymbol{x})(-i\hbar\nabla)\phi_\alpha(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\,, \tag{8.241}$$

which is one of the key equations in quantum mechanics.

Other useful results follow from (8.239). Because $|\alpha\rangle$ is arbitrary, we can substitute any ket in its place. In particular, let us use $|\boldsymbol{p}\rangle$, a momentum eigenket corresponding to the eigenvalue $\boldsymbol{p}$:

$$\hat{\boldsymbol{p}}|\boldsymbol{p}\rangle = \boldsymbol{p}|\boldsymbol{p}\rangle\,. \tag{8.242}$$

This gives

$$\boldsymbol{p}\langle\boldsymbol{x}|\boldsymbol{p}\rangle = -i\hbar\nabla\langle\boldsymbol{x}|\boldsymbol{p}\rangle\,. \tag{8.243}$$

But, this is just a differential equation for $\langle\boldsymbol{x}|\boldsymbol{p}\rangle$, with the solution

$$\langle\boldsymbol{x}|\boldsymbol{p}\rangle = ce^{\frac{i}{\hbar}\boldsymbol{p}\cdot\boldsymbol{x}}\,. \tag{8.244}$$

So, the momentum eigenket in the $\boldsymbol{x}$-representation is a plane wave.

Exercise 8.25. Verify that (8.244) satisfies (8.243). ///

How do we determine the constant c? As with $\hat{\boldsymbol{x}}$, we assume that $\hat{\boldsymbol{p}}$ is an observable and hence $\{|\boldsymbol{p}\rangle\}$ forms a basis, that is,

$$|\alpha\rangle = \int |\boldsymbol{p}\rangle\langle\boldsymbol{p}|\alpha\rangle \mathrm{d}\boldsymbol{p} \tag{8.245}$$

for any $|\alpha\rangle$. From (D.53), we see that

$$\int e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\mathrm{d}\boldsymbol{x} = (2\pi)^3\delta(\boldsymbol{k})\,, \tag{8.246}$$

where $\delta(\boldsymbol{k}) := \delta(k_1)\delta(k_2)\delta(k_3)$ by definition. By means of this identity, we find

$$\delta(\boldsymbol{x}-\boldsymbol{x}') = \langle\boldsymbol{x}|\boldsymbol{x}'\rangle = \int \langle\boldsymbol{x}|\boldsymbol{p}\rangle\langle\boldsymbol{p}|\boldsymbol{x}'\rangle \mathrm{d}\boldsymbol{p} = |c|^2(2\pi\hbar)^3\delta(\boldsymbol{x}-\boldsymbol{x}')\,. \tag{8.247}$$

While c can be any complex number satisfying this equation, the convention dictates that we take the real positive root. Thus,

$$c = \frac{1}{(2\pi\hbar)^{3/2}}\,. \tag{8.248}$$

Now that we have a complete expression for $\langle\boldsymbol{x}|\boldsymbol{p}\rangle$, we can immediately deduce the following relationship between the **coordinate-space wave function** $\phi_\alpha(\boldsymbol{x}) = \langle\boldsymbol{x}|\alpha\rangle$ and the **momentum–space wave function** $\phi_\alpha(\boldsymbol{p}) := \langle\boldsymbol{p}|\alpha\rangle$:

$$\phi_\alpha(\boldsymbol{x}) = \langle\boldsymbol{x}|\alpha\rangle = \int \langle\boldsymbol{x}|\boldsymbol{p}\rangle\langle\boldsymbol{p}|\alpha\rangle \mathrm{d}\boldsymbol{p} = \frac{1}{(2\pi\hbar)^{3/2}}\int \phi_\alpha(\boldsymbol{p})e^{\frac{i}{\hbar}\boldsymbol{p}\cdot\boldsymbol{x}}\mathrm{d}\boldsymbol{p} \tag{8.249}$$

and

$$\phi_\alpha(\boldsymbol{p}) = \langle \boldsymbol{p}|\alpha\rangle = \int \langle \boldsymbol{p}|\boldsymbol{x}\rangle\langle \boldsymbol{x}|\alpha\rangle \mathrm{d}\boldsymbol{x} = \frac{1}{(2\pi\hbar)^{3/2}} \int \phi_\alpha(\boldsymbol{x}) e^{-\frac{i}{\hbar}\boldsymbol{p}\cdot\boldsymbol{x}} \mathrm{d}\boldsymbol{x} \,. \tag{8.250}$$

In other words, $\phi_\alpha(\boldsymbol{x})$ and $\phi_\alpha(\boldsymbol{p})$ are the Fourier transform of each other.

Finally, by setting $|\alpha\rangle = |\boldsymbol{x}'\rangle$ in (8.239), we obtain

$$\langle \boldsymbol{x}|\hat{\boldsymbol{p}}|\boldsymbol{x}'\rangle = -i\hbar\nabla\langle \boldsymbol{x}|\boldsymbol{x}'\rangle = -i\hbar\nabla\delta(\boldsymbol{x}-\boldsymbol{x}') \,, \tag{8.251}$$

where we note that the derivative is with respect to $\boldsymbol{x}$ and not $\boldsymbol{x}'$. This is the matrix element of $\hat{\boldsymbol{p}}$ in the $\boldsymbol{x}$-representation.

Exercise 8.26. Show that

$$\hat{x}|p\rangle = -i\hbar\frac{\mathrm{d}|p\rangle}{\mathrm{d}p} \tag{8.252}$$

for a particle in a one-dimensional space. ///

8.8 †Time Evolution Operator

Now, we have to figure out how the state of a given system evolves with time. What we need is a linear operator that acts on a state ket $|\phi,t\rangle$ at time t and converts it to a new one $|\phi,t+\delta t\rangle$ at a later time $t+\delta t$. As with the linear translation, it is sufficient to consider an infinitesimally small interval of time. That is, we wish to find the **time evolution operator** $\hat{U}_{\delta t}(t)$ satisfying

$$|\phi,t+\delta t\rangle = \hat{U}_{\delta t}(t)|\phi,t\rangle \,, \tag{8.253}$$

where we allowed for the explicit dependence of $\hat{U}_{\delta t}(t)$ on t. As in the case of the linear translation, we list a few reasonable demands to be imposed on $\hat{U}_{\delta t}(t)$:

a. $\hat{U}_{\delta t}(t)$ preserves the norm of $|\phi,t\rangle$:

$$\langle \phi,t+\delta t|\phi,t+\delta t\rangle = \langle \phi,t|\hat{U}_{\delta t}(t)^\dagger \hat{U}_{\delta t}(t)|\phi,t\rangle = \langle \phi,t|\phi,t\rangle \,, \tag{8.254}$$

for which it is sufficient that

$$\hat{U}_{\delta t}(t)^\dagger \hat{U}_{\delta t}(t) = \hat{I} \,. \tag{8.255}$$

b. Two successive time translations, δt_1 followed by another δt_2, should produce a state identical to that produced by a single time translation by $\delta t_1 + \delta t_2$. Thus,

$$\hat{U}_{\delta t_2}(t+\delta t_1)\hat{U}_{\delta t_1}(t) = \hat{U}_{\delta t_1+\delta t_2}(t) \,. \tag{8.256}$$

c. The state of the system will not change at all if no time has passed. We also do not expect any discontinuity to be present at $\delta t = 0$. Thus,

$$\lim_{\delta t \to 0} \hat{U}_{\delta t} = \hat{U}_{\delta t=0} = \hat{I} \,. \tag{8.257}$$

Except for the use of δt in place of $\Delta \boldsymbol{x}$, these demands are identical to what we saw in Sect. 8.7.1. So, we can immediately write down the desired expression for $\hat{U}_{\delta t}$:

$$\hat{U}_{\delta t}(t) = \hat{I} - \frac{i}{\hbar}\hat{H}(t)\delta t \,, \tag{8.258}$$

in which $\hat{H}$, by definition, is called **Hamiltonian operator** and its eigenvalues are called **energy** of the system.

It is straightforward to verify that (8.258) is compatible with the demands we have listed. Unlike $\boldsymbol{x}$, there is no operator corresponding to t. As in classical mechanics, t is a label indicating when things are happening. So, there is no commutation relation to speak of here.

From here, the celebrated Schrödinger equation is just a stone's through away. Using (8.258) in (8.253),

$$|\phi, t+\delta t\rangle = \left[\hat{I} - \frac{i}{\hbar}\hat{H}(t)\delta t\right] |\phi, t\rangle \,. \tag{8.259}$$

Moving $\hat{I}|\phi,t\rangle = |\phi,t\rangle$ to the left and dividing the resulting expression by δt,

$$\frac{|\phi, t+\delta t\rangle - |\phi,t\rangle}{\delta t} = -\frac{i}{\hbar}\hat{H}(t)|\phi,t\rangle \,. \tag{8.260}$$

Taking the $\delta t \to 0$ limit, we find

$$i\hbar \frac{\mathrm{d}|\phi,t\rangle}{\mathrm{d}t} = \hat{H}(t)|\phi,t\rangle \,, \tag{8.261}$$

which is the **Schrödinger equation** written for a state ket. The form this equation takes in $\boldsymbol{r}$-representation is given in Sect. 8.13.

8.9 $\dagger\hat{U}_t$ is Unitary

From (8.258), we see immediately that the infinitesimal time evolution operator $\hat{U}_{\delta t}(t)$ is unitary:

$$\hat{U}^{\dagger}_{\delta t}(t)\hat{U}_{\delta t}(t) = \hat{U}_{\delta t}(t)\hat{U}^{\dagger}_{\delta t}(t) = \hat{I} \,. \tag{8.262}$$

Since any finite duration of time can be divided into many infinitesimal time intervals, (8.262) can be used to establish

$$\hat{U}_t^\dagger \hat{U}_t = \hat{U}_t \hat{U}_t^\dagger = \hat{I}\,, \tag{8.263}$$

where $\hat{U}_t := \hat{U}_t(0)$ is the time evolution operator that converts the state ket at $t = 0$ to that at time t. To see this more explicitly, let

$$\Delta t := \frac{t}{n} \quad \text{and} \quad t_i := i\Delta t\,, \quad i = 0, \ldots, n-1 \tag{8.264}$$

and write

$$\hat{U}_t = \hat{U}_{\Delta t}(t_{n-1})\hat{U}_{\Delta t}(t_{n-2}) \cdots \hat{U}_{\Delta t}(t_1)\hat{U}_{\Delta t}(t_0)\,. \tag{8.265}$$

Taking the adjoint of this equation, we obtain

$$\hat{U}_t^\dagger = \hat{U}_{\Delta t}^\dagger(t_0)\,, \hat{U}_{\Delta t}^\dagger(t_1) \cdots \hat{U}_{\Delta t}^\dagger(t_{n-2})\hat{U}_{\Delta t}^\dagger(t_{n-1})\,, \tag{8.266}$$

In the limit of $n \to \infty$, Δt becomes infinitesimally small and (8.262) applies:

$$\hat{U}_{\Delta t}^\dagger(t_i)\hat{U}_{\Delta t}(t_i) = \hat{U}_{\Delta t}(t_i)\hat{U}_{\Delta t}^\dagger(t_i) = \hat{I}\,, \quad i = 0, \ldots, n-1\,. \tag{8.267}$$

Combining (8.265), (8.266), and (8.267), we arrive at (8.263).

8.10 ‡Formal Solution of the Schrödinger Equation

In a purely formal manner, (8.261) can be solved for a general Hamiltonian $\hat{H}$. We say that the solution is only *formal* in the sense that it may be very difficult to evaluate the resulting expression for $|\phi, t\rangle$. Nevertheless, the formal solution is often very useful as it allows us to expresses $|\phi, t\rangle$ very compactly.

8.10.1 ‡Time-Independent $\hat{H}$

Let us first assume that $\hat{H}$ is independent of t. To find the formal solution of (8.261), we first expand $|\phi, t\rangle$ into the Taylor series around $t = 0$:

$$|\phi, t\rangle = |\phi, 0\rangle + \frac{\mathrm{d}}{\mathrm{d}t}|\phi, t\rangle\bigg|_{t=0} t + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}|\phi, t\rangle\bigg|_{t=0} t^2 + \cdots\,. \tag{8.268}$$

From (8.261),

$$\frac{\mathrm{d}}{\mathrm{d}t}|\phi, t\rangle = -\frac{i}{\hbar}\hat{H}|\phi, t\rangle. \tag{8.269}$$

Applying this formula to $\mathrm{d}|\phi,t\rangle/\mathrm{d}t$, which is just a ket, we find

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}|\phi,t\rangle = -\frac{i}{\hbar}\hat{H}\left[\frac{\mathrm{d}}{\mathrm{d}t}|\phi,t\rangle\right] = -\frac{i}{\hbar}\hat{H}\left[-\frac{i}{\hbar}\hat{H}|\phi,t\rangle\right] = \left(-\frac{i}{\hbar}\hat{H}\right)^2|\phi,t\rangle\ . \quad (8.270)$$

Continuing in this manner, we see that

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}|\phi,t\rangle = \left(-\frac{i}{\hbar}\hat{H}\right)^n|\phi,t\rangle\ , \quad n=1,2,\ldots \quad (8.271)$$

Setting $t=0$ in this equation and using the resulting expression in (8.268), we have

$$|\phi,t\rangle = \left[\sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{i}{\hbar}\hat{H}t\right)^n\right]|\phi,0\rangle\ , \quad (8.272)$$

which is the desired formal solution. This equation shows that

$$\hat{U}_t := \hat{U}_t(0) = \sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{i}{\hbar}\hat{H}t\right)^n \quad (8.273)$$

for a time-independent $\hat{H}$. Using this result, it is straightforward to show that

$$\frac{\mathrm{d}\hat{U}_t}{\mathrm{d}t} = -\frac{i}{\hbar}\hat{H}\hat{U}_t\ , \quad (8.274)$$

which is the Schrödinger equation written for $\hat{U}_t$. Multiplying both sides of (8.274) from the right by $|\phi,0\rangle$ and noting that this ket is independent of t, we recover (8.261).

Exercise 8.27. Derive (8.274) from (8.273). ///

As we saw in Sect. 8.3.2, a function of an operator is defined in terms of the Maclaurin expansion of the function. Recalling (B.5), we see that

$$\hat{U}_t = e^{-\frac{i}{\hbar}\hat{H}t}\ , \quad (8.275)$$

and hence

$$|\phi,t\rangle = e^{-\frac{i}{\hbar}\hat{H}t}|\phi,0\rangle\ . \quad (8.276)$$

8.10.2 †Time-Dependent $\hat{H}$

Equation (8.276) cannot be true if $\hat{H}$ depends on t. This becomes immediately clear if we ask ourselves at which instant $\hat{H}$ in the exponent should be evaluated. We must also allow for the possibility that $\hat{H}$ evaluated at different instances may not even commute.

To make a progress, we first integrate (8.261) from $t = 0$ to t:

$$|\phi,t\rangle = |\phi,0\rangle - \frac{i}{\hbar}\int_0^t \hat{H}(t_1)|\phi,t_1\rangle \mathrm{d}t_1 \,. \tag{8.277}$$

This is not a particularly useful form of expressing $|\phi,t\rangle$ since the right-hand side still contains the unknown ket $|\phi,t_1\rangle$. But, the ket can be expressed using (8.277) as

$$|\phi,t_1\rangle = |\phi,0\rangle - \frac{i}{\hbar}\int_0^{t_1} \hat{H}(t_2)|\phi,t_2\rangle \mathrm{d}t_2 \,. \tag{8.278}$$

Substituting this expression into (8.277), we obtain

$$|\phi,t\rangle = |\phi,0\rangle + \left[-\frac{i}{\hbar}\int_0^t \hat{H}(t_1)\mathrm{d}t_1\right]|\phi,0\rangle + \left(-\frac{i}{\hbar}\right)^2 \int_0^t \int_0^{t_1} \hat{H}(t_1)\hat{H}(t_2)|\phi,t_2\rangle \mathrm{d}t_2 \mathrm{d}t_1 \,, \tag{8.279}$$

where we pulled $|\phi,0\rangle$ out of the integral sign because it is independent of t. Now, we can use (8.277) once again to express $|\phi,t_2\rangle$ as in (8.278). Continuing in this manner, we see that the time evolution operator for a time-dependent $\hat{H}$ is given by

$$\hat{U}_t = \sum_{n=0}^{\infty}\left(-\frac{i}{\hbar}\right)^n \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} \hat{H}(t_1)\hat{H}(t_2)\cdots\hat{H}(t_n)\mathrm{d}t_n \cdots \mathrm{d}t_2 \mathrm{d}t_1 \,, \tag{8.280}$$

in which $n = 0$ term in the summation is the identity operator $\hat{I}$. This result is known as the **Neumann series**.

Iterated integrals in which a limit of an integral depends on the values of the other integration variables is rather tedious to handle. However, we can circumvent this difficulty through a clever trick. To this end, let us consider the integrals for the $n = 2$ term:

$$I_2 := \int_0^t \int_0^{t_1} \hat{H}(t_1)\hat{H}(t_2)\mathrm{d}t_2 \mathrm{d}t_1 \,. \tag{8.281}$$

Because t_1 and t_2 are simply the integration variables, we have

$$I_2 = \int_0^t \int_0^{t_2} \hat{H}(t_2)\hat{H}(t_1)\mathrm{d}t_1 \mathrm{d}t_2 \,. \tag{8.282}$$

In (8.281), the integral over t_2 runs from 0 to t_1. Thus, $0 \le t_2 \le t_1$. On the other hand, we have $0 \le t_1 \le t_2$ in (8.282).

Let us now define the **time-ordering operator** $\hat{\mathscr{T}}_>$ which acts on a *product of time-dependent operators* and rearranges them so that the time decreases from left to right. Thus, if $t_a < t_b$, we have

$$\hat{\mathscr{T}}_>\left[\hat{A}(t_a)\hat{B}(t_b)\right] = \hat{\mathscr{T}}_>\left[\hat{B}(t_b)\hat{A}(t_a)\right] = \hat{B}(t_b)\hat{A}(t_a) \,. \tag{8.283}$$

Similarly if we have a product of more than two time-dependent operators. By means of $\hat{\mathscr{T}}_>$, we can rewrite (8.281) and (8.282) as

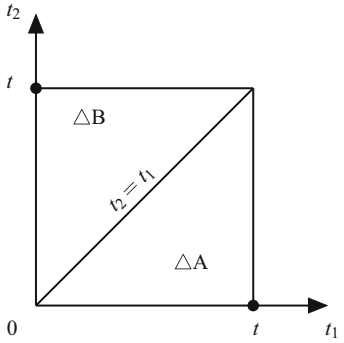

Fig. 8.1 The integral (8.284) is carried out over the *lower triangle* (△A) while (8.285) is over the *upper triangle* (△B). Taken together, they cover the square.

$$I_2 = \int_0^t \int_0^{t_1} \hat{\mathscr{T}}_> \left[\hat{H}(t_1)\hat{H}(t_2)\right] \mathrm{d}t_2 \mathrm{d}t_1 \tag{8.284}$$

and

$$I_2 = \int_0^t \int_0^{t_2} \hat{\mathscr{T}}_> \left[\hat{H}(t_1)\hat{H}(t_2)\right] \mathrm{d}t_1 \mathrm{d}t_2 \,, \tag{8.285}$$

respectively. These two integrals share the same integrand. From Fig. 8.1, we see that these two integrals, taken together, cover the square region defined by $0 \le t_1 \le t$ and $0 \le t_2 \le t$. Thus, adding (8.284) and (8.285), we arrive at

$$I_2 = \frac{1}{2} \int_0^t \int_0^t \hat{\mathscr{T}}_> \left[\hat{H}(t_1)\hat{H}(t_2)\right] \mathrm{d}t_1 \mathrm{d}t_2 \,. \tag{8.286}$$

In a similar manner, it is possible to show that the n-dimensional integral in (8.280) can be written as

$$I_n = \frac{1}{n!} \int_0^t \cdots \int_0^t \hat{\mathscr{T}}_> \left[\hat{H}(t_1) \cdots \hat{H}(t_n)\right] \mathrm{d}t_1 \cdots \mathrm{d}t_n \,. \tag{8.287}$$

In this way, we arrive at

$$\hat{U}_t = \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{i}{\hbar}\right)^n \int_0^t \cdots \int_0^t \hat{\mathscr{T}}_> \left[\hat{H}(t_1) \cdots \hat{H}(t_n)\right] \mathrm{d}t_1 \cdots \mathrm{d}t_n \,, \tag{8.288}$$

which is known as the **Dyson series**. As a short-hand notation, it is common to express this result in terms of the **time-ordered exponential** as

$$\hat{U}_t = \hat{\mathscr{T}}_> \exp\left[-\frac{i}{\hbar} \int_0^t \hat{H}(t')\mathrm{d}t'\right] \,. \tag{8.289}$$

Exercise 8.28. Show that $\hat{U}_t$ given by (8.288) satisfies the Schrödinger equation:

$$\frac{\mathrm{d}\hat{U}_t}{\mathrm{d}t} = -\frac{i}{\hbar}\hat{H}(t)\hat{U}_t \tag{8.290}$$

for the time evolution operator $\hat{U}_t$. ///

If $\hat{H}(t_a)$ and $\hat{H}(t_b)$ commutes for *any* t_a and t_b in the interval $[0,t]$, that is, if

$$\hat{H}(t_a)\hat{H}(t_b) = \hat{H}(t_b)\hat{H}(t_a) \quad \text{for any } t_1, t_2 \in [0,t] \tag{8.291}$$

then we have

$$\hat{\mathscr{T}}_{>}\left[\hat{H}(t_1)\cdots\hat{H}(t_n)\right] = \hat{H}(t_1)\cdots\hat{H}(t_n)\,, \tag{8.292}$$

that is, a product of $\hat{H}$'s evaluated at different instances defines an unique operator regardless of the order in which $\hat{H}$'s are multiplied. In this case, (8.287) reduces to

$$I_n = \frac{1}{n!}\int_0^t \cdots \int_0^t \hat{H}(t_1)\cdots\hat{H}(t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n = \frac{1}{n!}\left[\int_0^t \hat{H}(t')\mathrm{d}t'\right]^n \tag{8.293}$$

and we now have

$$\hat{U}_t = \exp\left[-\frac{i}{\hbar}\int_0^t \hat{H}(t')\mathrm{d}t'\right]\,. \tag{8.294}$$

in place of (8.289). If $\hat{H}$ is actually independent of t, this expression reduces to (8.275).

8.11 ‡Heisenberg's Equation of Motion

Thus far, we have taken a point of view that the state of a quantum mechanical system is represented by a state ket $|\phi,t\rangle$ with its time evolution dictated by the Schrödinger equation. The time dependence of the average value $\langle A\rangle_{\mathrm{QM}}(t)$ of a dynamical variable A is a result of this time evolution and a possible *explicit* time dependence of $\hat{A}$. This view point is called the **Schrödinger picture**.

An alternative view point emerges through the use of the time evolution operator $\hat{U}_t$:

$$\langle A\rangle_{\mathrm{QM}}(t) = \langle\phi,t|\hat{A}|\phi,t\rangle = \langle\phi,0|\hat{U}_t^{\dagger}\hat{A}\hat{U}_t|\phi,0\rangle\,, \tag{8.295}$$

where we used the formal solution,

$$|\phi,t\rangle = \hat{U}_t|\phi,0\rangle\,, \tag{8.296}$$

of the Schrödinger equation from Sect. 8.10 and and its adjoint:

$$\langle\phi,t| = \langle\phi,0|\hat{U}_t^{\dagger}\,. \tag{8.297}$$

We now interpret (8.295) as indicating that the time dependence of $\langle A\rangle_{\rm QM}(t)$ arises *entirely* from that of the *time-dependent* operator

$$\hat{A}_H(t) := \hat{U}_t^\dagger \hat{A} \hat{U}_t \,, \tag{8.298}$$

while the *state ket* $|\phi, 0\rangle$ remains frozen in the ket space. This point of view is called the **Heisenberg picture**. By construction, both pictures lead to the same prediction for $\langle A\rangle_{\rm QM}(t)$.

Because $\hat{U}_{t=0} = \hat{I}$, the newly defined operator $\hat{A}_H(t)$ satisfies the initial condition that

$$\hat{A}_H(0) = \hat{A} \,. \tag{8.299}$$

What is the differential equation governing its time evolution? To find out, let us take the time derivative of (8.298):

$$\frac{\mathrm{d}\hat{A}_H}{\mathrm{d}t} = \frac{\mathrm{d}\hat{U}_t^\dagger}{\mathrm{d}t} \hat{A} \hat{U}_t + \hat{U}_t^\dagger \hat{A} \frac{\mathrm{d}\hat{U}_t}{\mathrm{d}t} + \hat{U}_t^\dagger \frac{\partial \hat{A}}{\partial t} \hat{U}_t \,. \tag{8.300}$$

As we shall see below, the right-hand side can be rewritten to highlight the formal similarity between this equation from quantum mechanics and the equation of motion (1.185) from classical mechanics.

8.11.1 ‡*Time-Independent $\hat{H}$*

For time-independent Hamiltonian, (8.274) and its adjoint

$$\frac{\mathrm{d}\hat{U}_t^\dagger}{\mathrm{d}t} = \frac{i}{\hbar} \hat{U}_t^\dagger \hat{H} \tag{8.301}$$

give

$$\frac{\mathrm{d}\hat{A}_H}{\mathrm{d}t} = \frac{i}{\hbar} \hat{U}_t^\dagger \hat{H} \hat{A} \hat{U}_t - \frac{i}{\hbar} \hat{U}_t^\dagger \hat{A} \hat{H} \hat{U}_t + \hat{U}_t^\dagger \frac{\partial \hat{A}}{\partial t} \hat{U}_t \,. \tag{8.302}$$

Following (8.298), we write

$$\left(\frac{\partial \hat{A}}{\partial t} \right)_H := \hat{U}_t^\dagger \frac{\partial \hat{A}}{\partial t} \hat{U}_t \,. \tag{8.303}$$

Since $\hat{U}_t$ and $\hat{U}_t^\dagger$ are both functions of $\hat{H}$ and $\hat{H}$ is independent of time, we have

$$[\hat{H}, \hat{U}_t] = 0 \quad \text{and} \quad [\hat{H}, \hat{U}_t^\dagger] = 0 \,. \tag{8.304}$$

This allows us to rewrite (8.302) as

$$\frac{\mathrm{d}\hat{A}_H}{\mathrm{d}t} = \frac{i}{\hbar}\hat{H}\hat{A}_H(t) - \frac{i}{\hbar}\hat{A}_H(t)\hat{H} + \left(\frac{\partial\hat{A}}{\partial t}\right)_H = \frac{1}{i\hbar}[\hat{A}_H(t),\hat{H}] + \left(\frac{\partial\hat{A}}{\partial t}\right)_H . \quad (8.305)$$

This is **Heisenberg's equation of motion** to be contrasted against (1.185).

8.11.2 †*Time-Dependent* $\hat{H}$

Let us turn to a time-dependent $\hat{H}$ next. In this case, we must use (8.290) and its adjoint in place of (8.274) and (8.301). This gives (8.302) with $\hat{H}$ replaced by $\hat{H}(t)$. However, (8.304) can be used *only if* (8.291) holds. What do we do if it does not?

This more general situations can be handled by means of (8.263), which allows us to rewrite (8.302), with $\hat{H}$ replaced by $\hat{H}(t)$, as

$$\begin{aligned}\frac{\mathrm{d}\hat{A}_H}{\mathrm{d}t} &= \frac{i}{\hbar}\hat{U}_t^\dagger\hat{H}(t)\hat{U}_t\hat{U}_t^\dagger\hat{A}\hat{U}_t - \frac{i}{\hbar}\hat{U}_t^\dagger\hat{A}\hat{U}_t\hat{U}_t^\dagger\hat{H}(t)\hat{U}_t + \left(\frac{\partial\hat{A}}{\partial t}\right)_H \\ &= \frac{i}{\hbar}\hat{H}_H(t)\hat{A}_H(t) - \frac{i}{\hbar}\hat{A}_H(t)\hat{H}_H(t) + \left(\frac{\partial\hat{A}}{\partial t}\right)_H \\ &= \frac{1}{i\hbar}[\hat{A}_H(t),\hat{H}_H(t)] + \left(\frac{\partial\hat{A}}{\partial t}\right)_H , \end{aligned} \quad (8.306)$$

where, following (8.298), we defined

$$\hat{H}_H(t) := \hat{U}_t^\dagger\hat{H}(t)\hat{U}_t . \quad (8.307)$$

If (8.291) holds, then so does (8.304) and we obtain

$$\hat{H}_H(t) = \hat{H}(t)\hat{U}_t^\dagger\hat{U}_t = \hat{H}(t) . \quad (8.308)$$

In this case, (8.306) becomes

$$\frac{\mathrm{d}\hat{A}_H}{\mathrm{d}t} = \frac{1}{i\hbar}[\hat{A}_H(t),\hat{H}(t)] + \left(\frac{\partial\hat{A}}{\partial t}\right)_H , \quad (8.309)$$

which reduces to (8.305) *if* $\hat{H}(t)$ is actually independent of time.

8.12 Eigenstates of $\hat{H}$

As shown in Sect. 8.8, the time evolution of a state ket is determined by the Schrödinger equation (8.261), in which $\hat{H}$ is the Hamiltonian operator. An eigenvalue of the Hamiltonian and the corresponding eigenket are called, respectively,

the **energy eigenvalue** and the **energy eigenket**. The entire collection of eigenvalues may form either a continuum spectrum, a discrete spectrum, or a combination thereof. However, for a system of particles confined to a finite region of space, the spectrum is discrete. (See Example 8.13 and Sect. 8.14.) This is the case for systems of our primary interest and we shall pretend that the spectrum is discrete in what follows. Thus, let E_n and $|n\rangle$ denote an energy eigenvalue and the corresponding eigenket:

$$\hat{H}|n\rangle = E_n|n\rangle\,, \quad n = 0, 1, \cdots \tag{8.310}$$

with the label n distinguishing one eigenvalue (or eigenket) from others.

For our purposes, it is sufficient to consider a time-independent $\hat{H}$. Then, $|n\rangle$ is also independent of time. In this case, (8.276) applies and we have

$$|\phi,t\rangle = e^{-\frac{i}{\hbar}\hat{H}t}|\phi,0\rangle = \sum_n e^{-\frac{i}{\hbar}\hat{H}t}|n\rangle\langle n|\phi,0\rangle = \sum_n C_n e^{-\frac{i}{\hbar}E_n t}|n\rangle\,, \tag{8.311}$$

in which

$$C_n := \langle n|\phi,0\rangle \tag{8.312}$$

is a constant determined by initial conditions. Equation (8.311) expresses the time dependent $|\phi,t\rangle$ as a linear combination of energy eigenkets. It follows that the time evolution of the state ket $|\phi,t\rangle$ is completely determined by the solution to the eigenvalue problem (8.310) and initial conditions.

Exercise 8.29. Suppose that the state ket $|\phi,t\rangle$ coincided with one of the energy eigenkets, say $|m\rangle$ at $t = 0$. Being an eigenket of a time-independent $\hat{H}$, $|m\rangle$ remains constant. But, $|\phi,t\rangle$ evolves with time. Show that the expectation value $\langle A\rangle_{\rm QM}$ of an observable $\hat{A}$ remains constant. For this reason, the energy eigenkets are called **stationary states**. Carry out the similar calculation for more general case, in which $|\phi\rangle$ is given by (8.311). Under what condition is $\langle A\rangle_{\rm QM}$ constant? ///

Exercise 8.30. Suppose that $\hat{H}$ of the system is time independent. By means of a direct calculation, show that the expectation value of the system energy $\langle\phi,t|\hat{H}|\phi,t\rangle$ remains constant despite the fact that the state of the system $|\phi,t\rangle$ evolves with time. ///

Example 8.11. Two state system: Let

$$\hat{H} = c[|\phi_1\rangle\langle\phi_2| + |\phi_2\rangle\langle\phi_1|]\,, \tag{8.313}$$

where c is a real nonzero constant having the dimension of energy and $\{|\phi_1\rangle, |\phi_2\rangle\}$ is an orthonormal basis. Let us find

a. The eigenvalues of $\hat{H}$.
b. The corresponding eigenkets.
c. The time evolution of a stateket $|\phi,t\rangle$ if $|\phi,0\rangle = |\phi_1\rangle$.

a. Let $\hat{H}|\alpha\rangle = \lambda|\alpha\rangle$. Using the closure relation,

$$\sum_j \hat{H}|\phi_j\rangle\langle\phi_j|\alpha\rangle = \lambda|\alpha\rangle\,. \tag{8.314}$$

Multiplying this equation by $\langle\phi_i|$, we find

$$\sum_j \langle\phi_i|\hat{H}|\phi_j\rangle\langle\phi_j|\alpha\rangle = \lambda\langle\phi_i|\alpha\rangle\,. \tag{8.315}$$

For $\hat{H}$ given by (8.313), we have

$$\langle\phi_1|\hat{H}|\phi_1\rangle = \langle\phi_2|\hat{H}|\phi_2\rangle = 0 \quad \text{and} \quad \langle\phi_1|\hat{H}|\phi_2\rangle = \langle\phi_2|\hat{H}|\phi_1\rangle = c\,. \tag{8.316}$$

Thus, (8.315) may be written more explicitly as

$$\begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \lambda\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}\,, \tag{8.317}$$

or,

$$\begin{pmatrix} -\lambda & c \\ c & -\lambda \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,, \tag{8.318}$$

where

$$\alpha_i := \langle\phi_i|\alpha\rangle\,. \tag{8.319}$$

Equation (8.318) has a nontrivial solution if and only if

$$\begin{vmatrix} -\lambda & c \\ c & -\lambda \end{vmatrix} = 0\,, \tag{8.320}$$

that is, $\lambda^2 - c^2 = 0$. So, the eigenvalues are $\lambda = \pm c$.

b. Let $\lambda = c$. Then, (8.318) becomes

$$\begin{pmatrix} -c & c \\ c & -c \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,, \tag{8.321}$$

from which we find $\alpha_1 = \alpha_2$. So,

$$|c\rangle = \sum_i |\phi_i\rangle\langle\phi_i|c\rangle = \alpha_1[|\phi_1\rangle + |\phi_2\rangle]\,. \tag{8.322}$$

Upon normalization, we find

$$|c\rangle = \frac{1}{\sqrt{2}}[|\phi_1\rangle + |\phi_2\rangle]\,. \tag{8.323}$$

Similarly, the eigenket corresponding to the eigenvalue $-c$ is found to be

$$|-c\rangle = \frac{1}{\sqrt{2}}[|\phi_1\rangle - |\phi_2\rangle] \,. \tag{8.324}$$

c. According to (8.311), the general solution of the Schrödinger equation is the linear combination of $|c\rangle$ and $|-c\rangle$:

$$|\phi,t\rangle = C_1 e^{-\frac{i}{\hbar}ct}|c\rangle + C_2 e^{\frac{i}{\hbar}ct}|-c\rangle \,. \tag{8.325}$$

From the initial condition, we have

$$|\phi,0\rangle = C_1|c\rangle + C_2|-c\rangle = |\phi_1\rangle \,. \tag{8.326}$$

Multiplying this equation by $\langle\phi_1|$ from the left, we find

$$\frac{1}{\sqrt{2}}(C_1 + C_2) = 1 \,. \tag{8.327}$$

Multiplying by $\langle\phi_2|$ from the left, we have

$$\frac{1}{\sqrt{2}}(C_1 - C_2) = 0 \,. \tag{8.328}$$

Thus, $C_1 = C_2 = 1/\sqrt{2}$ and

$$|\phi,t\rangle = \frac{1}{\sqrt{2}}\left[e^{-\frac{i}{\hbar}ct}|c\rangle + e^{\frac{i}{\hbar}ct}|-c\rangle\right] \,. \tag{8.329}$$

Exercise 8.31. Let

$$\hat{H} = a\,[|\phi_1\rangle\langle\phi_1| + i|\phi_1\rangle\langle\phi_2| + i|\phi_2\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2|] \,, \tag{8.330}$$

where a is a real nonzero constant having a dimension of energy. Can this be the Hamiltonian of some two-state system? Why or why not? ///

8.13 ‡Schrödinger Wave Equation

Given a physical system of interest, how do we find its Hamiltonian? As an example, let us consider a particle moving in a one-dimensional space.

Classical mechanically, the Hamiltonian of such a system is

$$H = \frac{p^2}{2m} + \psi(x) \,, \tag{8.331}$$

where m is the mass of the particle and ψ is an external field. In quantum mechanics, we simply adopt this Hamiltonian and replace p and x by the corresponding Hermitian operators:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \psi(\hat{x}) \,. \tag{8.332}$$

Accordingly, the Schrödinger equation (8.261) becomes

$$i\hbar \frac{\mathrm{d}|\phi,t\rangle}{\mathrm{d}t} = \left[\frac{\hat{p}^2}{2m} + \psi(\hat{x})\right] |\phi,t\rangle \,. \tag{8.333}$$

Let us multiply this equation by $\langle x|$ from the left and consider what happens to each terms. The end result is a differential equation for the wave function $\phi(x,t) := \langle x|\phi,t\rangle$.

Because $\langle x|$ is a *basis bra* in the $\boldsymbol{x}$-representation, it does not depend on time and hence

$$\left\langle x \left| i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \right| \phi,t \right\rangle = i\hbar \frac{\partial}{\partial t} \langle x|\phi,t\rangle = i\hbar \frac{\partial \phi(x,t)}{\partial t} \,. \tag{8.334}$$

Note that we replaced the total time derivative by the partial derivative since the wave function is a function of both x and t. Using (8.130), we find

$$\psi(\hat{x})|x\rangle = \psi(x)|x\rangle \,. \tag{8.335}$$

Since $\hat{H}$ is Hermitian, so is $\psi(\hat{x})$. Thus, the adjoint of this equation reads

$$\langle x|\psi(\hat{x}) = \langle x|\psi(x) \,, \tag{8.336}$$

and we arrive at

$$\langle x|\psi(\hat{x})|\phi,t\rangle = \langle x|\psi(x)|\phi,t\rangle = \psi(x)\langle x|\phi,t\rangle = \psi(x)\phi(x,t) \,. \tag{8.337}$$

The remaining term takes a little more effort. Using the closure relation (8.201) written for the one-dimensional case, we write

$$\langle x|\hat{p}^2|\phi,t\rangle = \int \langle x|\hat{p}|x'\rangle\langle x'|\hat{p}|\phi,t\rangle \mathrm{d}x' \,. \tag{8.338}$$

Using (8.239) and (8.251),

$$\langle x|\hat{p}^2|\phi,t\rangle = (-i\hbar)^2 \int \frac{\partial}{\partial x}\delta(x-x')\frac{\partial}{\partial x'}\phi(x',t)\mathrm{d}x' \,. \tag{8.339}$$

Since the integration is with respect to x', $\partial/\partial x$ can be pulled out of the integral sign, thus leading to

$$\langle x|\hat{p}^2|\phi,t\rangle = -\hbar^2 \frac{\partial}{\partial x} \int \delta(x-x')\frac{\partial}{\partial x'}\phi(x',t)\mathrm{d}x' = -\hbar^2 \frac{\partial^2}{\partial x^2}\phi(x,t) \,. \tag{8.340}$$

Combining everything, we arrive at

$$i\hbar\frac{\partial}{\partial t}\phi(x,t) = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \psi(x)\right]\phi(x,t)\,, \tag{8.341}$$

which is the famous **time-dependent Schrödinger wave equation** applied to the one-dimensional system.

As we have seen in Sect. 8.12, the problem of finding $|\phi,t\rangle$ reduces to that of solving the eigenvalue problem (8.310). Its $\boldsymbol{x}$-representation for the present problem is

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \psi(x)\right]u_n(x) = E_n u_n(x)\,, \tag{8.342}$$

where $u_n(x) := \langle x|n\rangle$ is the **energy eigenfunction** corresponding to the energy eigenvalue E_n. Equation (8.342) is the **time-independent Schrödinger wave equation**. The general solution of (8.341) is obtained by simply expressing (8.311) in the $\boldsymbol{x}$-representation:

$$\phi(x,t) = \sum_n C_n e^{-\frac{i}{\hbar}E_n t}u_n(x)\,. \tag{8.343}$$

As in Sect. 8.12, the solution to the eigenvalue problem (8.342) and initial conditions completely determine the time evolution of the wave function $\phi(x,t)$. One can also obtain (8.342) and (8.343) directly from (8.341) by means of separation of variables, in which one assumes that the solution is of the form of $\phi(x,t) = \theta(t)u(x)$.

In accordance with the postulates of quantum mechanics,

$$|\langle x|\phi,t\rangle|^2\mathrm{d}x = \phi^*(x,t)\phi(x,t)\mathrm{d}x \tag{8.344}$$

is the probability of finding the particle within the interval $\mathrm{d}x$ taken around the point x. Because of this probabilistic interpretation, $\phi(x,t)$ cannot diverge. We also expect $\phi(x,t)$ to be a continuous function. If the particle is confined to a finite region of space, these conditions demand that the energy eigenvalues form a discrete spectrum. For a general discussion on this point, see Chap. 3 of Ref. [6]. Specific examples are given in Example 8.13 and Sect. 8.14.

Example 8.12. Free particle: For a free particle moving in a one-dimensional space,

$$\hat{H} = \frac{\hat{p}^2}{2m}\,, \tag{8.345}$$

whose energy eigenfunction is the one-dimensional version of (8.244):

$$\langle x|p\rangle = Ce^{\frac{i}{\hbar}px}\,. \tag{8.346}$$

To see this, we note that

$$\frac{\mathrm{d}}{\mathrm{d}x}\langle x|p\rangle = \frac{i}{\hbar}pCe^{\frac{i}{\hbar}px} = \frac{i}{\hbar}p\langle x|p\rangle \tag{8.347}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\langle x|p\rangle = \left(\frac{i}{\hbar}p\right)^2\langle x|p\rangle = -\left(\frac{p}{\hbar}\right)^2\langle x|p\rangle\,. \tag{8.348}$$

Thus,

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\langle x|p\rangle = \frac{p^2}{2m}\langle x|p\rangle\,. \tag{8.349}$$

This is just (8.342) written for the free particle, for which $\psi(x) \equiv 0$. That is, $\langle x|p\rangle$ is an eigenfunction of the Hamiltonian (8.345) and belongs to the eigenvalue $p^2/2m$.

Insofar as we have called (though rather arbitrarily) the eigenvalue of $\hat{p}$ momentum in Sect. 8.7, it is quite appropriate that we have identified an eigenvalue E of $\hat{H}$ as the energy in Sect. 8.8. Finally, we note that (8.349) holds for any real number p. In the absence of an external field that confines the particle to a finite region of space, the energy of the system forms a continuous spectrum.

It is straightforward to generalize (8.341) and (8.342) for a system of N particles in a three-dimensional space. The results are

$$i\hbar\frac{\partial}{\partial t}\phi(\boldsymbol{r}^N,t) = \left[\sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}{\nabla_i}^2 + \psi(\boldsymbol{r}^N)\right]\phi(\boldsymbol{r}^N,t)\,, \tag{8.350}$$

where ∇_i acts only on the coordinates of the ith particle, and

$$\left[\sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}{\nabla_i}^2 + \psi(\boldsymbol{r}^N)\right]\phi(\boldsymbol{r}^N,t) = E\phi(\boldsymbol{r}^N,t)\,, \tag{8.351}$$

respectively.

Example 8.13. Particle in a box: Consider a particle of mass m confined to a rectangular box of dimension $L_x \times L_y \times L_z$. The classical mechanical Hamiltonian of the particle is given by (3.75), in which $\psi_w(\boldsymbol{r})$ is the wall potential specified by (3.76). The corresponding time-independent Schrödinger wave equation reads

$$-\frac{\hbar^2}{2m}\left[\frac{\partial^2 u(\boldsymbol{r})}{\partial x^2} + \frac{\partial^2 u(\boldsymbol{r})}{\partial y^2} + \frac{\partial^2 u(\boldsymbol{r})}{\partial z^2}\right] + \psi_w(\boldsymbol{r})u(\boldsymbol{r}) = Eu(\boldsymbol{r})\,. \tag{8.352}$$

Outside the box, ψ_w is ∞. Provided that E, u, and its second partial derivatives are all finite, we have $u = 0$ outside the box. So, let us focus on the region inside the box, in which $\psi_w(\boldsymbol{r}) = 0$, and hence

$$-\frac{\hbar^2}{2m}\left[\frac{\partial^2 u(\boldsymbol{r})}{\partial x^2}+\frac{\partial^2 u(\boldsymbol{r})}{\partial y^2}+\frac{\partial^2 u(\boldsymbol{r})}{\partial z^2}\right] = Eu(\boldsymbol{r})\,. \tag{8.353}$$

Suppose that $u(\boldsymbol{r}) = X(x)Y(y)Z(z)$. Substituting this expression into (8.353) and dividing the resulting equation by u, we find

$$\frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2}+\frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2}+\frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} = -\frac{2mE}{\hbar^2}\,. \tag{8.354}$$

By rewriting this as

$$\frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} = -\frac{2mE}{\hbar^2}-\frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2}-\frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} \tag{8.355}$$

we see that the left-hand side is a function only of x, while the right-hand side is a function only of y and z. Because the equality holds for any x, y, and z, they must equal a constant, which we call ${k_x}^2$.[47] This gives

$$\frac{\mathrm{d}^2X}{\mathrm{d}x^2}+{k_x}^2X = 0 \quad \text{and} \quad \frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2} = -\frac{2mE}{\hbar^2}-\frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2}+{k_x}^2\,. \tag{8.356}$$

Again, we see that the left-hand side of the second equation must be a constant, which we denote by $-{k_y}^2$ and obtain

$$\frac{\mathrm{d}^2Y}{\mathrm{d}y^2}+{k_y}^2Y = 0 \quad \text{and} \quad \frac{\mathrm{d}^2Z}{\mathrm{d}z^2}+{k_z}^2Z = 0\,, \tag{8.357}$$

where we defined k_z by

$${k_x}^2+{k_y}^2+{k_z}^2 = \frac{2mE}{\hbar^2}\,. \tag{8.358}$$

So far, we have not placed any restriction on k_x, k_y, or k_z. They may even be a complex number. The ordinary differential equation for X in (8.356) has the following solution:

$$X(x) = A_x\cos k_x x + B_x \sin k_x x\,, \tag{8.359}$$

where A_x and B_x are constants to be determined by the boundary conditions. Because $u(\boldsymbol{r})$ is zero outside the box, $u(\boldsymbol{r}) = 0$ if $x < 0$ or $x > L_x$. But, we also expect that $u(\boldsymbol{r})$ is continuous at the walls of the box. This means that

$X(0) = X(L_x) = 0$. The first equality gives $A_x = 0$, while the second one gives $B_x \sin k_x L_x = 0$. If $B_x = 0$, then $u(\boldsymbol{r}) \equiv 0$, which is not a solution we seek. So, $\sin k_x L_x = 0$, indicating that

$$k_x = \frac{n_x \pi}{L_x} , \quad n_x = 1, 2, 3, \ldots \tag{8.360}$$

We exclude $n_x = 0$, for which $u(\boldsymbol{r}) \equiv 0$. The negative integers ($n_x = -1, -2, -3, \ldots$) does not generate any new solution because $B_x \sin(-k_x x) = -B_x \sin(k_x x)$. A similar procedure leads to

$$Y(y) = B_y \sin k_y y \, , k_y = \frac{n_y \pi}{L_y} , \quad n_y = 1, 2, 3, \ldots \tag{8.361}$$

and

$$Z(z) = B_z \sin k_z z \, , k_z = \frac{n_z \pi}{L_z} , \quad n_z = 1, 2, 3, \ldots \tag{8.362}$$

From (8.358), we see that the energy eigenvalue E is determined by

$$\frac{8mE}{h^2} = \left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2 , \tag{8.363}$$

in which n_x, n_y, and n_z are positive integers and we used $h = 2\pi\hbar$. In contrast to the case of a free particle, energy spectrum is discrete. The corresponding energy eigenfunction is now labeled by three indices n_x, n_y, and n_z:

$$u_{n_x, n_y, n_z}(\boldsymbol{r}) = B \sin\left(\frac{\pi n_x}{L_x} x\right) \sin\left(\frac{\pi n_y}{L_y} y\right) \sin\left(\frac{\pi n_z}{L_z} z\right) . \tag{8.364}$$

The remaining constant B may be determined by the normalization condition. Recalling that $\sin^2 \theta = (1 - \cos 2\theta)/2$, we find

$$\int_0^{L_x} \sin^2\left(\frac{\pi n_x}{L_x} x\right) \mathrm{d}x = \frac{L_x}{\pi n_x} \int_0^{\pi n_x} \sin^2 \theta \mathrm{d}\theta = \frac{L_x}{2} . \tag{8.365}$$

Similarly for the other factors in (8.364). Thus,

$$u_{n_x, n_y, n_z}(\boldsymbol{r}) = \sqrt{\frac{8}{V}} \sin\left(\frac{\pi n_x}{L_x} x\right) \sin\left(\frac{\pi n_y}{L_y} y\right) \sin\left(\frac{\pi n_z}{L_z} z\right) , \tag{8.366}$$

where $V = L_x L_y L_z$ and we chose B to be a real positive number.

8.14 ‡Harmonic Oscillator

For a one-dimensional harmonic oscillator, the Hamiltonian is given by

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}k\hat{x}^2 \,. \tag{8.367}$$

The primary goal of this section is to find the expression for the energy eigenvalues for this Hamiltonian. We have already made use of the result in Example 4.2.

With $\hat{H}$ given by (8.367), the time-independent Schrödinger wave equation (8.342) reads

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}kx^2\right] u_n(x) = E_n u_n(x) \,, \tag{8.368}$$

which can be solved directly to yield both E_n and $u_n(x)$. See Ref. [6] for details. As an alternative, we shall follow an elegant operator method.

8.14.1 ‡Operator Method

Let $\omega := \sqrt{k/m}$ and rewrite (8.367) as

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2 = \frac{m\omega^2}{2}\left(\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2}\right) . \tag{8.369}$$

Because $\hat{x}$ and $\hat{p}$ do not commute, we *cannot* just factorize this as

$$\hbar\omega\sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} + \frac{i\hat{p}}{m\omega}\right)\sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} - \frac{i\hat{p}}{m\omega}\right) = \hbar\omega\hat{a}\hat{a}^\dagger \,, \tag{8.370}$$

where we defined

$$\hat{a} := \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} + \frac{i\hat{p}}{m\omega}\right) \tag{8.371}$$

and hence

$$\hat{a}^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} - \frac{i\hat{p}}{m\omega}\right) . \tag{8.372}$$

By means of direct computations, however, we obtain

$$\hat{a}\hat{a}^\dagger = \frac{m\omega}{2\hbar}\left(\hat{x}^2 - \frac{i}{m\omega}\hat{x}\hat{p} + \frac{i}{m\omega}\hat{p}\hat{x} + \frac{\hat{p}^2}{m^2\omega^2}\right) \tag{8.373}$$

and

$$\hat{a}^\dagger\hat{a} = \frac{m\omega}{2\hbar}\left(\hat{x}^2 + \frac{i}{m\omega}\hat{x}\hat{p} - \frac{i}{m\omega}\hat{p}\hat{x} + \frac{\hat{p}^2}{m^2\omega^2}\right) . \tag{8.374}$$

It follows that

$$[\hat{a}, \hat{a}^\dagger] = -\frac{i}{\hbar}[\hat{x}, \hat{p}] = 1 \tag{8.375}$$

where the last equality follows from (8.234). In contrast,

$$\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a} = \frac{m\omega}{\hbar}\left(\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2}\right) = \frac{2}{\hbar\omega}\hat{H} \,. \tag{8.376}$$

Eliminating $\hat{a}\hat{a}^\dagger$ in (8.376) by means of (8.375), we obtain

$$\hat{H} = \hbar\omega\left(\hat{N} + \frac{1}{2}\right) \,, \tag{8.377}$$

where

$$\hat{N} := \hat{a}^\dagger\hat{a} \tag{8.378}$$

is known as the **number operator**.

Let n denote the eigenvalue of $\hat{N}$ and write the corresponding eigenket as $|n\rangle$:

$$\hat{N}|n\rangle = n|n\rangle \,. \tag{8.379}$$

From (8.377), it is clear that

$$[\hat{H}, \hat{N}] = 0 \,. \tag{8.380}$$

By Theorem 8.5, the eigenkets of $\hat{N}$ are also energy eigenkets. But,

$$\hat{H}|n\rangle = \hbar\omega\left(\hat{N} + \frac{1}{2}\right)|n\rangle = \hbar\omega\left(n + \frac{1}{2}\right)|n\rangle \,. \tag{8.381}$$

So, $|n\rangle$ belongs to the energy eigenvalue $\hbar\omega(n + 1/2)$.

It might appear that (4.75) follows directly from (8.381). But, we have yet to show that n can only be a nonnegative integer. Since $\hat{N}$ is Hermitian, we know at least that n must be a real number. In addition, n must be nonnegative. This is because the norm of a ket cannot be negative:

$$0 \leq \|(\hat{a}|n\rangle)\|^2 = \langle n|\hat{a}^\dagger\hat{a}|n\rangle = n\langle n|n\rangle \,. \tag{8.382}$$

But, $|n\rangle$ is not a zero ket due to the convention mentioned in Sect. 8.3.1. So, $\langle n|n\rangle > 0$ and we conclude that $n \geq 0$.

In order to further narrow down the possible values n can take, we evaluate

$$\hat{N}\hat{a}|n\rangle = (\hat{a}^\dagger\hat{a})\hat{a}|n\rangle = (\hat{a}\hat{a}^\dagger - 1)\hat{a}|n\rangle = \hat{a}(\hat{a}^\dagger\hat{a} - 1)|n\rangle = (n-1)\hat{a}|n\rangle \,, \tag{8.383}$$

where we used (8.375) to transform the second expression to the third. Equation (8.383) indicates that $\hat{a}|n\rangle$ is an eigenket of $\hat{N}$ corresponding to the eigenvalue $n - 1$. That is,

$$\hat{a}|n\rangle = c|n-1\rangle \,. \tag{8.384}$$

The constant c may be determined by demanding the normalization of the energy eigenkets. Noting that

$$|c|^2\langle n-1|n-1\rangle = \langle n|\hat{a}^\dagger\hat{a}|n\rangle = n\langle n|n\rangle = n\,, \tag{8.385}$$

we have $|c|^2 = n$. By convention, we choose a real positive root for c, that is

$$c = \sqrt{n}\,. \tag{8.386}$$

and write (8.384) as

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle\,. \tag{8.387}$$

The operator $\hat{a}$, when acting on $|n\rangle$, reduces the value of n by one. Since this corresponds to the decrease in the energy eigenvalue by one quantum unit of energy $\hbar\omega$, $\hat{a}$ is known as the **annihilation operator**.

Repeated applications of $\hat{a}$ on $|n\rangle$ give the following sequence of kets:

$$\begin{aligned}\hat{a}|n\rangle &= \sqrt{n}|n-1\rangle\,,\\ \hat{a}^2|n\rangle &= \sqrt{n(n-1)}|n-2\rangle\,,\\ \hat{a}^3|n\rangle &= \sqrt{n(n-1)(n-2)}|n-3\rangle\,,\cdots\end{aligned} \tag{8.388}$$

As we have seen above, however, $n-m \geq 0$ for any eigenket $|n-m\rangle$ of $\hat{N}$. This means that the sequence (8.388) must terminate before $n-m$ becomes negative. This happens if n is a nonnegative integer because

$$\hat{a}^n|n\rangle = \sqrt{n(n-1)(n-2)\cdots 1}|0\rangle \tag{8.389}$$

and an additional application of $\hat{a}$ produces the zero ket $|\theta\rangle$, which will remain unchanged upon further applications of $\hat{a}$. (See Exercises 8.1b and 8.8a.) Thus, the energy eigenvalues form a discrete set:

$$E_n = \hbar\omega\left(n+\frac{1}{2}\right)\,,\quad n = 0,1,2\ldots\,, \tag{8.390}$$

which is (4.75) we encountered before.

It is important to distinguish $|\theta\rangle$ and $|0\rangle$. The former is the zero ket, the norm of which is zero. The latter, in contrast, is a simultaneous eigenket of $\hat{N}$ and $\hat{H}$ corresponding to the lowest possible energy $\hbar\omega/2$ the harmonic oscillator can take. In other words, $|0\rangle$ is the state ket of the so-called **ground state**. Classical mechanically, the lowest energy occurs when $(x,p) = (0,0)$. Quantum mechanically, such a state is in violation of the Heisenberg uncertainty principle as we saw in Sect. 8.5.

Given the ground state ket $|0\rangle$, $\hat{a}^\dagger$ allows us to systematically generate the other eigenstates of $\hat{N}$. To see this, we evaluate

$$\hat{N}\hat{a}^\dagger|n\rangle = \hat{a}^\dagger(\hat{a}\hat{a}^\dagger)|n\rangle = \hat{a}^\dagger(\hat{a}^\dagger\hat{a}+1)|n\rangle = \hat{a}^\dagger(n+1)|n\rangle = (n+1)\hat{a}^\dagger|n\rangle\,, \tag{8.391}$$

where we used (8.375). According to (8.391), $\hat{a}^\dagger|n\rangle$ is an eigenket of $\hat{N}$ corresponding to the eigenvalue $n+1$:

$$\hat{a}^\dagger|n\rangle = c|n+1\rangle \,, \tag{8.392}$$

where c is a constant. The effect of $\hat{a}^\dagger$ on $|n\rangle$ is to increase the value of n by one and hence the energy eigenvalue by one quantum unit of energy $\hbar\omega$. So, the operator $\hat{a}^\dagger$ is referred to as the **creation operator**.

Because the effect of $\hat{a}$ on $|n\rangle$ is known and $\hat{a}^\dagger$ is the adjoint of $\hat{a}$, the formalism we have developed so far should be sufficient to determine the value of c. In fact, multiplying (8.392) by $\hat{a}$ from the left, we find

$$\begin{aligned} \hat{a}\hat{a}^\dagger|n\rangle &= c\hat{a}|n+1\rangle \\ \left(\hat{a}^\dagger\hat{a}+1\right)|n\rangle &= c\sqrt{n+1}|n\rangle \\ (n+1)|n\rangle &= c\sqrt{n+1}|n\rangle \,. \end{aligned} \tag{8.393}$$

Multiplying both sides by $\langle n|$ from the left, we obtain $c = \sqrt{n+1}$. Thus,

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \,. \tag{8.394}$$

Repeated applications of $\hat{a}^\dagger$ on $|0\rangle$ give

$$\begin{aligned} \hat{a}^\dagger|0\rangle &= |1\rangle \\ (\hat{a}^\dagger)^2|0\rangle &= \hat{a}^\dagger|1\rangle = \sqrt{2}|2\rangle \\ (\hat{a}^\dagger)^3|0\rangle &= \hat{a}^\dagger\sqrt{2}|2\rangle = \sqrt{3\cdot 2}|3\rangle \\ (\hat{a}^\dagger)^4|0\rangle &= \hat{a}^\dagger\sqrt{3\cdot 2}|3\rangle = \sqrt{4\cdot 3\cdot 2}|4\rangle \,. \end{aligned} \tag{8.395}$$

Proceeding in this manner, we find

$$|n\rangle = \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}}|0\rangle \,. \tag{8.396}$$

8.14.2 ‡*Energy Eigenfunctions*

It is now straightforward to find the wave function $u_n(x) := \langle x|n\rangle$. We start from

$$\hat{a}|0\rangle = |\theta\rangle \,, \tag{8.397}$$

and hence

$$\langle x|\hat{a}|0\rangle = \langle x|\theta\rangle = 0 \,. \tag{8.398}$$

Using (8.371) and recalling (8.239), we have

$$\left(x + \frac{\hbar}{m\omega}\frac{\mathrm{d}}{\mathrm{d}x}\right) u_0 = 0 \,, \tag{8.399}$$

which can be solved to give

$$u_0(x) = ce^{-\frac{1}{2}\alpha x^2} , \tag{8.400}$$

where $\alpha = m\omega/\hbar$. Upon normalization under the convention that c be a real positive number, we have

$$u_0(x) = \left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\alpha x^2/2} . \tag{8.401}$$

Substituting this result into (8.368), we find that $E_0 = \hbar\omega/2$, which is in agreement with (8.390).

From (8.396),

$$u_n(x) = \frac{1}{\sqrt{n!}} \left\langle x \middle| (\hat{a}^\dagger)^n \middle| 0 \right\rangle . \tag{8.402}$$

Once again, we use (8.372) and (8.239) and find

$$\begin{aligned} u_n(x) &= \frac{1}{\sqrt{n!}} \left(\frac{\alpha}{2}\right)^{n/2} \left(x - \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n u_0(x) \\ &= \frac{1}{\sqrt{2^n n!}} \left(\sqrt{\alpha}x - \frac{1}{\sqrt{\alpha}}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \left(\frac{\alpha}{\pi}\right)^{1/4} \exp\left[-\frac{1}{2}\left(\sqrt{\alpha}x\right)^2\right] \\ &=: N_n \left(\xi - \frac{\mathrm{d}}{\mathrm{d}\xi}\right)^n e^{-\xi^2/2} , \end{aligned} \tag{8.403}$$

where $\xi := \sqrt{\alpha}x$ and

$$N_n := \frac{1}{\sqrt{2^n n!}} \left(\frac{\alpha}{\pi}\right)^{1/4} . \tag{8.404}$$

More explicitly, the energy eigenfunctions are given by

$$\begin{aligned} u_0(\xi) &= N_0 e^{-\xi^2/2} , \\ u_1(\xi) &= N_1 (2\xi) e^{-\xi^2/2} , \\ u_2(\xi) &= N_2 (4\xi^2 - 2) e^{-\xi^2/2} , \\ u_3(\xi) &= N_3 (8\xi^3 - 12\xi) e^{-\xi^2/2} , \\ u_4(\xi) &= N_4 (16\xi^4 - 48\xi^2 + 12) e^{-\xi^2/2} , \\ u_5(\xi) &= N_5 (32\xi^5 - 160\xi^3 + 120\xi) e^{-\xi^2/2} , \cdots . \end{aligned} \tag{8.405}$$

The polynomials in these equations, that is,

$$\begin{aligned} H_0(\xi) &= 1 , \\ H_1(\xi) &= 2\xi , \\ H_2(\xi) &= 4\xi^2 - 2 , \\ H_3(\xi) &= 8\xi^3 - 12\xi , \\ H_4(\xi) &= 16\xi^4 - 48\xi^2 + 12 , \\ H_5(\xi) &= 32\xi^5 - 160\xi^3 + 120\xi , \cdots . \end{aligned} \tag{8.406}$$

are known as the **Hermit polynomials**. For various properties of this mathematical object, see Ref. [6].

Classical mechanically, the energy E of the one-dimensional harmonic oscillator is given by

$$E = \frac{p^2}{2m} + \frac{1}{2}kx^2 = \frac{p^2}{2m} + \frac{1}{2}\hbar\omega\xi^2 \,. \tag{8.407}$$

Thus,

$$\frac{1}{2}\hbar\omega\xi^2 = E - \frac{p^2}{2m} \leq E \,, \tag{8.408}$$

and the motion of the particle is confined to the region defined by

$$|\xi| \leq \sqrt{\frac{2E}{\hbar\omega}} \,. \tag{8.409}$$

$$|\xi| \leq \sqrt{2n+1} \,, \quad n = 0, 1, 2, \ldots \tag{8.410}$$

Equation (8.405) indicates, however, that the probability, $u_n^*(x)u_n(x)\mathrm{d}x$, of finding a particle in the classical mechanically forbidden region ($|\xi| > \sqrt{2n+1}$) is nonzero. This is illustrated in Fig. 8.2 for several values of n.

8.15 †Ehrenfest's Theorem

In Sect. 8.11, we derived Heisenberg's equation of motion, which may be regarded as the quantum mechanical counterpart of (1.185) in classical mechanics. Insofar as (1.185) contains Hamilton's and therefore Newton's equations of motion, it is natural to ask if their quantum mechanical counterparts can be derived from Heisenberg's equation of motion. This is indeed the case as we shall see in this section.

For simplicity, we restrict ourselves to a one-dimensional system described by the Hamiltonian:

$$\hat{H}(t) = \frac{\hat{p}^2}{2m} + \psi(\hat{x}, t) \,. \tag{8.411}$$

Applying (8.306) to $\hat{x}_H(t)$ and noting that $\hat{x}$ (an operator in the Schrödinger picture) is independent of time,

$$\frac{\mathrm{d}\hat{x}_H}{\mathrm{d}t} = \frac{1}{i\hbar}[\hat{x}_H(t), \hat{H}_H(t)] = \frac{1}{i\hbar}\hat{U}_t^\dagger[\hat{x}, \hat{H}(t)]\hat{U}_t \,, \tag{8.412}$$

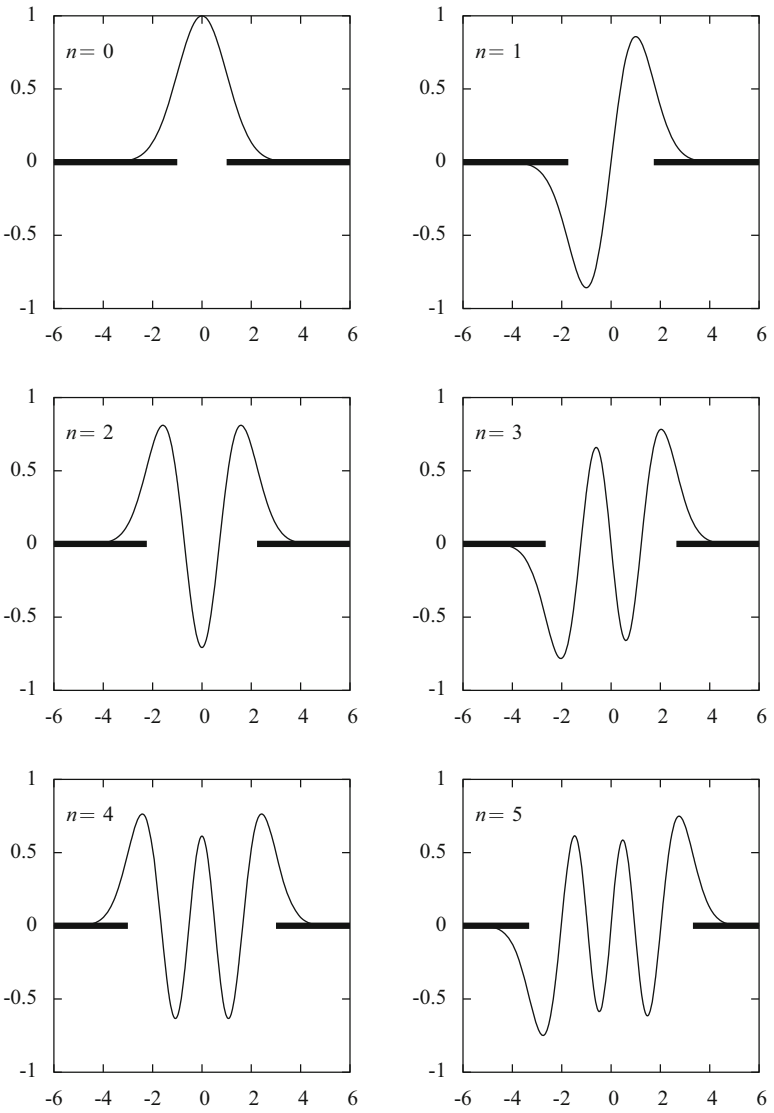

Fig. 8.2 Energy eigenfunctions $(\alpha/\pi)^{-1/4}u_n(\xi)$ (*solid line*) and classically forbidden regions (*thick horizontal lines*) for several values of n. The horizontal axis is ξ.

where we used (8.298) in the last step. But,

$$[\hat{x}, \hat{H}(t)] = \left[\hat{x}, \frac{\hat{p}^2}{2m} + \psi(\hat{x}, t)\right] = \frac{1}{2m}[\hat{x}, \hat{p}^2]\,. \tag{8.413}$$

Using (8.105) and (8.234),

$$[\hat{x},\hat{p}^2] = \hat{p}[\hat{x},\hat{p}] + [\hat{x},\hat{p}]\hat{p} = 2i\hbar\hat{p}\,. \tag{8.414}$$

Thus,

$$\frac{\mathrm{d}\hat{x}_H}{\mathrm{d}t} = \frac{1}{m}\hat{U}_t^\dagger \hat{p}\hat{U}_t = \frac{\hat{p}_H(t)}{m}\,. \tag{8.415}$$

This is to be contrasted with one of Hamilton's equations of motion:

$$\dot{x} = \frac{\partial H}{\partial p}\,. \tag{8.416}$$

For the classical mechanical Hamiltonian corresponding to (8.411), this gives

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{p}{m}\,. \tag{8.417}$$

In view of (8.415) and (8.417), the name momentum operator we have given to $\hat{p}$ is quite appropriate.

In a similar manner,

$$\frac{\mathrm{d}\hat{p}_H}{\mathrm{d}t} = \frac{1}{i\hbar}[\hat{p}_H(t),\hat{H}_H(t)] = \frac{1}{i\hbar}\hat{U}_t^\dagger[\hat{p},\hat{H}(t)]\hat{U}_t = \frac{1}{i\hbar}\hat{U}_t^\dagger[\hat{p},\psi(\hat{x},t)]\hat{U}_t\,. \tag{8.418}$$

If $\psi(x,t)$ has the Maclaurin expansion, the operator $\psi(\hat{x},t)$ is given by

$$\psi(\hat{x},t) = \psi(0,t) + \psi'(0,t)\hat{x} + \frac{1}{2}\psi''(0,t)\hat{x}^2 + \frac{1}{3!}\psi'''(0,t)\hat{x}^3 + \cdots\,, \tag{8.419}$$

where $'$ indicates the *partial* derivative with respect to x. With repeated applications of (8.105) and (8.234), we find

$$[\hat{p},1] = 0\,, \quad [\hat{p},\hat{x}] = -i\hbar\,, \quad [\hat{p},\hat{x}^2] = \hat{x}[\hat{p},\hat{x}] + [\hat{p},\hat{x}]\hat{x} = -2i\hbar\hat{x}\,, \tag{8.420}$$

and

$$[\hat{p},\hat{x}^3] = \hat{x}[\hat{p},\hat{x}^2] + [\hat{p},\hat{x}]\hat{x}^2 = -2i\hbar\hat{x}^2 - i\hbar\hat{x}^2 = -3i\hbar\hat{x}^2\,. \tag{8.421}$$

Proceeding in this manner, we obtain

$$[\hat{p},\hat{x}^n] = -i\hbar n\hat{x}^{n-1}\,, \quad n = 1,2,\ldots \tag{8.422}$$

Thus,

$$[\hat{p},\psi(\hat{x},t)] = -i\hbar\psi'(\hat{x},t)\,, \tag{8.423}$$

and hence

$$\frac{\mathrm{d}\hat{p}_H}{\mathrm{d}t} = -\hat{U}_t^\dagger\psi'(\hat{x},t)\hat{U}_t\,. \tag{8.424}$$

Noting that

$$\hat{U}_t^\dagger\hat{x}^n\hat{U}_t = [\hat{x}_H(t)]^n\,, \tag{8.425}$$

as can be seen by inserting $\hat{U}_t\hat{U}_t^\dagger = 1$ between $\hat{x}$'s, we arrive at

$$\frac{\mathrm{d}\hat{p}_H}{\mathrm{d}t} = -\psi'(\hat{x}_H(t),t)\,. \tag{8.426}$$

This result should be compared with its classical mechanical counterpart:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\psi'(x,t)\,, \tag{8.427}$$

which follows from the remaining half of Hamilton's equations of motion:

$$\dot{p} = -\frac{\partial H}{\partial x}\,. \tag{8.428}$$

Combining (8.415) and (8.426), we find

$$m\frac{\mathrm{d}^2\hat{x}_H}{\mathrm{d}t^2} = -\psi'(\hat{x}_H(t),t)\,. \tag{8.429}$$

This is the quantum mechanical counterpart of Newton's equation of motion.

As seen from (8.295) and (8.298), the expectation value of an operator in the Heisenberg picture, such as $\hat{x}_H$ and $\hat{p}_H$, must be evaluated using the state ket at $t = 0$. With (8.415), this leads to

$$\frac{\mathrm{d}\langle x\rangle_{\mathrm{QM}}}{\mathrm{d}t} = \frac{\langle p\rangle_{\mathrm{QM}}}{m}\,. \tag{8.430}$$

For the other half of the equations of motion, it is easier to work with (8.424) than with (8.426). The result is

$$\frac{\mathrm{d}\langle p\rangle_{\mathrm{QM}}}{\mathrm{d}t} = -\langle\psi'(\hat{x},t)\rangle_{\mathrm{QM}}\,, \tag{8.431}$$

If the particle is well localized, the approximate relation:

$$\langle\psi'(\hat{x},t)\rangle_{\mathrm{QM}} \approx \psi'(\langle x\rangle_{\mathrm{QM}},t) \tag{8.432}$$

holds. In fact, following Exercise 8.23, we have

$$\begin{aligned}\langle\psi'(\hat{x},t)\rangle_{\mathrm{QM}} &= \langle\phi|\psi'(\hat{x})|\phi\rangle = \int\langle\phi|\psi'(\hat{x})|x\rangle\langle x|\phi\rangle\mathrm{d}x = \int\psi'(x)\langle\phi|x\rangle\langle x|\phi\rangle\mathrm{d}x \\ &= \int\psi'(x)|\langle x|\phi\rangle|^2\mathrm{d}x\,,\end{aligned} \tag{8.433}$$

in which $|\langle x|\phi\rangle|^2 \mathrm{d}x$ is the probability that the particle is found between x and $x+\mathrm{d}x$ as seen from (8.344). If this probability is sharply peaked around the average $\langle x\rangle_{\mathrm{QM}}$, then

$$\langle \psi'(\hat{x},t)\rangle_{\mathrm{QM}} \approx \psi'(\langle x\rangle_{\mathrm{QM}}) \int |\langle x|\phi\rangle|^2 \mathrm{d}x = \psi'(\langle x\rangle_{\mathrm{QM}}) \,. \tag{8.434}$$

Within this approximation,

$$\frac{\mathrm{d}\langle p\rangle_{\mathrm{QM}}}{\mathrm{d}t} = -\psi'(\langle x\rangle_{\mathrm{QM}},t) \,. \tag{8.435}$$

Equations (8.430) and (8.435) are known as **Ehrenfest's theorem**.

Some functions, such as $\psi(x,t) = x^n$ $(n < 1)$, do not have the Maclaurin expansion. Even then, (8.423) still holds provided that $\psi'(x,t)$ exists. One way to see this is to expand $\psi(\hat{x},t)$ into a Taylor series around $x = a \neq 0$ and note that a, being a number, commutes with $\hat{p}$. Alternatively, we start by writing

$$\begin{aligned}[\hat{p},\psi(\hat{x})] &= \hat{p}\psi(\hat{x}) - \psi(\hat{x})\hat{p} = \int_{-\infty}^{\infty} [\hat{p}|x\rangle\langle x|\psi(\hat{x}) - \psi(\hat{x})|x\rangle\langle x|\hat{p}]\mathrm{d}x \\ &= \int_{-\infty}^{\infty} [\hat{p}|x\rangle\psi(x)\langle x| - |x\rangle\psi(x)\langle x|\hat{p}]\mathrm{d}x \,, \end{aligned} \tag{8.436}$$

where we used the closure relation and (8.130). Using the closure relation one more time,

$$\hat{p}|x\rangle = \int_{-\infty}^{\infty} |x_1\rangle\langle x_1|\hat{p}|x\rangle \mathrm{d}x_1 = \int_{-\infty}^{\infty} |x_1\rangle(-i\hbar)\frac{\mathrm{d}}{\mathrm{d}x_1}\delta(x-x_1)\mathrm{d}x_1 \,, \tag{8.437}$$

where we made use of (8.251). Let $X := x - x_1$ and use the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}x_1}\delta(x-x_1) = \frac{\mathrm{d}X}{\mathrm{d}x_1}\frac{\mathrm{d}}{\mathrm{d}X}\delta(X) = -\frac{\mathrm{d}}{\mathrm{d}X}\delta(X) = -\frac{\mathrm{d}}{\mathrm{d}x}\delta(x-x_1) \,, \tag{8.438}$$

Because x is not the integration variable, we have

$$\hat{p}|x\rangle = i\hbar\frac{\mathrm{d}}{\mathrm{d}x}\int |x_1\rangle\delta(x-x_1)\mathrm{d}x_1 = i\hbar\frac{\mathrm{d}|x\rangle}{\mathrm{d}x} \,. \tag{8.439}$$

By means of this equation and its adjoint, we may rewrite (8.436) as

$$\begin{aligned}[\hat{p},\psi(\hat{x})] &= i\hbar\int_{-\infty}^{\infty}\left[\frac{\mathrm{d}|x\rangle}{\mathrm{d}x}\psi(x)\langle x| + |x\rangle\psi(x)\frac{\mathrm{d}\langle x|}{\mathrm{d}x}\right]\mathrm{d}x \\ &= i\hbar\int_{-\infty}^{\infty}\left\{\frac{\mathrm{d}}{\mathrm{d}x}[|x\rangle\psi(x)\langle x|] - |x\rangle\psi'(x)\langle x|\right\}\mathrm{d}x \,, \end{aligned} \tag{8.440}$$

When we compute the expectation value of this expression with respect to the state ket $|\phi,t\rangle$, the first term in the integrand can be dropped provided that $\psi(x)$ becomes much larger than the system energy in the $|x| \to \infty$ limit. This is because $\langle\phi,t|x\rangle$ vanishes rapidly as $|x| \to \infty$ for such $\psi(x)$. (See Fig. 8.2 for example.) More explic-

itly,

$$\langle\phi,t|\left\{\int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}[|x\rangle\psi(x)\langle x|]\mathrm{d}x\right\}|\phi,t\rangle = \int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}[\langle\phi,t|x\rangle\psi(x)\langle x|\phi,t\rangle]\mathrm{d}x$$
$$= \langle\phi,t|x\rangle\psi(x)\langle x|\phi,t\rangle|_{-\infty}^{\infty} = 0\,. \quad (8.441)$$

As far as the expectation value is concerned, therefore, we may write

$$[\hat{p},\psi(\hat{x})] = -i\hbar\int_{-\infty}^{\infty}|x\rangle\psi'(x)\langle x|\mathrm{d}x = -i\hbar\psi'(\hat{x})\,, \quad (8.442)$$

where the last step follows from (8.131).

Finally, exchanging the role of $\hat{p}$ and $\hat{x}$ in (8.423) and noting that $[\hat{x},\hat{p}] = -[\hat{p},\hat{x}] = i\hbar$, we find

$$[\hat{x},f(\hat{p})] = i\hbar f'(\hat{p})\,. \quad (8.443)$$

This formula, too, can be extended for f that does not have the Maclaurin expansion.

Exercise 8.32. Using (8.252), show that (8.443) holds even for a function that does not have the Maclaurin expansion. Assume that

$$\lim_{p\to\pm\infty}\langle\phi,t|p\rangle f(p)\langle p|\phi,t\rangle = 0\,. \quad (8.444)$$

///

8.16 Quantum Statistical Mechanics

In this final section, we build on the key concepts of quantum mechanics we have reviewed so far and formulate statistical mechanics using the language of quantum mechanics.

8.16.1 Density Matrix

In classical mechanics, we considered the ensemble average given by (3.48):

$$\langle A\rangle = \int A(q^f,p^f)\rho(q^f,p^f)\mathrm{d}q^f\mathrm{d}p^f\,, \quad (8.445)$$

which was identified with the result of a measurement of some dynamical variable A. We suppose that an analogous relation holds for quantum mechanical systems and write

$$\langle A\rangle = \sum_\alpha \omega_\alpha\langle\phi_\alpha|\hat{A}|\phi_\alpha\rangle\,, \quad (8.446)$$

where $\langle\phi_\alpha|\hat{A}|\phi_\alpha\rangle$ is the expectation value for the outcome of the measurement of A when the system is in the microstate α for which the state ket is $|\phi_\alpha\rangle$. Note that this averaging is a part of the postulate of quantum mechanics and that it was absent in classical mechanics. The coefficient ω_α is the quantum counterpart of $\rho(q^f, p^f)$ in classical statistical mechanics and is the probability that the system is in the microstate α. Being a probability, ω_α must satisfy the normalization condition:

$$\sum_\alpha \omega_\alpha = 1 \,. \tag{8.447}$$

We emphasize that $|\phi_\alpha\rangle$ $(\alpha = 1, 2, \ldots)$ do not necessarily form a basis, nor are they orthogonal. In what follows, however, we assume that they are normalized. Using (8.128), we can rewrite (8.446) as

$$\begin{aligned}\langle A\rangle &= \sum_{\alpha,i} \omega_\alpha \langle\phi_\alpha|a_i\rangle a_i \langle a_i|\phi_\alpha\rangle = \sum_{\alpha,i} \langle a_i|\phi_\alpha\rangle \omega_\alpha \langle\phi_\alpha|a_i\rangle a_i = \sum_{\alpha,i} \langle a_i|\phi_\alpha\rangle \omega_\alpha \langle\phi_\alpha|\hat{A}|a_i\rangle \\ &= \sum_i \langle a_i| \left[\sum_\alpha |\phi_\alpha\rangle \omega_\alpha \langle\phi_\alpha|\right] \hat{A}|a_i\rangle =: \sum_i \langle a_i|\hat{\rho}_t \hat{A}|a_i\rangle = \mathrm{Tr}\{\hat{\rho}_t \hat{A}\} \,,\end{aligned} \tag{8.448}$$

where we defined the **density operator** $\hat{\rho}_t$ by

$$\hat{\rho}_t := \sum_\alpha |\phi_\alpha\rangle \omega_\alpha \langle\phi_\alpha| \,. \tag{8.449}$$

In terms of the density operator, (8.447) is written as

$$\mathrm{Tr}\{\hat{\rho}_t\} = 1 \,. \tag{8.450}$$

In fact,

$$\mathrm{Tr}\{\hat{\rho}_t\} = \sum_{i,\alpha} \langle a_i|\phi_\alpha\rangle \omega_\alpha \langle\phi_\alpha|a_i\rangle = \sum_{i,\alpha} \omega_\alpha \langle\phi_\alpha|a_i\rangle\langle a_i|\phi_\alpha\rangle = \sum_\alpha \omega_\alpha \langle\phi_\alpha|\phi_\alpha\rangle = \sum_\alpha \omega_\alpha \,, \tag{8.451}$$

where we used the closure relation and the fact that $|\phi_\alpha\rangle$ is normalized.

8.16.2 Statistical Equilibrium

As is the case with classical mechanics, we first define statistical equilibrium by demanding that

$$\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} \equiv 0 \tag{8.452}$$

hold for any observable $\hat{A}$ that is independent of time. To express this in a more informative fashion, let us first rewrite the trace using some *time-independent basis*

$\{|b_1\rangle, \ldots, |b_r\rangle\}$:

$$\langle A\rangle = \mathrm{Tr}\{\hat{\rho}_t \hat{A}\} = \sum_{i,j} \langle b_i|\hat{\rho}_t|b_j\rangle\langle b_j|\hat{A}|b_i\rangle \,. \tag{8.453}$$

Because $\hat{A}$ is time independent, the matrix element $\langle b_j|\hat{A}|b_i\rangle$ is independent of time and the only time dependence of $\langle A\rangle$ is in $\hat{\rho}_t$. Thus,

$$\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} = \sum_{i,j} \left\langle b_i \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_j \right\rangle \langle b_j|\hat{A}|b_i\rangle \,. \tag{8.454}$$

We see shortly that this expression vanishes for arbitrary $\hat{A}$ if and only if *all* the matrix elements of $\mathrm{d}\hat{\rho}_t/\mathrm{d}t$ vanish in the B-representation:

$$\left\langle b_i \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_j \right\rangle \equiv 0 \tag{8.455}$$

which is true if and only if

$$\frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \equiv 0 \,. \tag{8.456}$$

This is then the necessary and sufficient condition for statistical equilibrium.

Equation (8.455) is evidently sufficient for (8.452). To see its necessity, suppose that

$$\langle b_\beta|\hat{A}|b_\alpha\rangle \neq 0 \,. \tag{8.457}$$

Because $\hat{A}$ is Hermitian,

$$\langle b_\alpha|\hat{A}|b_\beta\rangle = \langle b_\beta|\hat{A}|b_\alpha\rangle^* \neq 0 \,. \tag{8.458}$$

If these are the *only* nonzero matrix elements of $\hat{A}$ in the B-representation, (8.454) becomes

$$\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} = \left\langle b_\alpha \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_\beta \right\rangle \langle b_\beta|\hat{A}|b_\alpha\rangle + \left\langle b_\beta \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_\alpha \right\rangle \langle b_\alpha|\hat{A}|b_\beta\rangle \,. \tag{8.459}$$

Because ω_α in (8.449) is a real number, we see that $\hat{\rho}_t$ is a Hermitian at all time. Thus,

$$\frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\hat{\rho}_{t+\Delta t} - \hat{\rho}_t}{\Delta t} \tag{8.460}$$

is also a Hermitian and we can write

$$\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} = \left\langle b_\alpha \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_\beta \right\rangle \langle b_\alpha|\hat{A}|b_\beta\rangle^* + \left\langle b_\alpha \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_\beta \right\rangle^* \langle b_\alpha|\hat{A}|b_\beta\rangle \,. \tag{8.461}$$

which is twice the real part of the expression

$$\left\langle b_\alpha \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_\beta \right\rangle \langle b_\alpha|\hat{A}|b_\beta\rangle^* \,. \tag{8.462}$$

Equation (8.452) thus demands that the expression (8.462) be purely imaginary. But, since $\hat{A}$ is arbitrary, and hence $\langle b_\alpha|\hat{A}|b_\beta\rangle$ is a complex number in general, this will be the case only if

$$\left\langle b_\alpha \left| \frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} \right| b_\beta \right\rangle = 0 \,. \tag{8.463}$$

8.16.3 Liouville's Theorem

Equation (8.456) can be transformed into a more revealing form by means of the quantum mechanical version of Liouville's theorem. The situation parallels what we saw in classical statistical mechanics.

Before proceeding to the proof of the theorem, we must point out that the ket $|\phi_\alpha\rangle$ in $\hat{\rho}_t$, in general, evolves with time. There is nothing in our definition of $\hat{\rho}_t$ to exclude the possibility that the ket evolves to become yet another ket already included in the summation over α or otherwise become something that was not in that summation initially. So, we need to be a little more precise about what is meant by the sum over α in (8.449). The density matrix is constructed at some fixed point in time, say at $t = 0$, using the *state kets* $|\phi_1,0\rangle$, $|\phi_2,0\rangle$,.... These kets may evolve with time and, at a later time t, become $|\phi_1,t\rangle$, $|\phi_2,t\rangle$,..., respectively. Then, at time t, our $\hat{\rho}_t$ will be given by

$$\hat{\rho}_t = \sum_\alpha |\phi_\alpha,t\rangle\omega_\alpha\langle\phi_\alpha,t| \,. \tag{8.464}$$

That is, ω_α does not change with time because this refers to the fraction of copies in microstate $|\phi_\alpha\rangle$ at time $t = 0$ in the statistical ensemble that was created at that time.

We now proceed to evaluate $\mathrm{d}\hat{\rho}_t/\mathrm{d}t$. Taking the time derivative of (8.464), we have

$$\frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} = \sum_\alpha \left[\frac{\mathrm{d}|\phi_\alpha,t\rangle}{\mathrm{d}t}\omega_\alpha\langle\phi_\alpha,t| + |\phi_\alpha,t\rangle\omega_\alpha\frac{\mathrm{d}\langle\phi_\alpha,t|}{\mathrm{d}t} \right] . \tag{8.465}$$

Recall the Schrödinger equation (8.261) and take its adjoint:

$$\frac{\mathrm{d}|\phi_\alpha,t\rangle}{\mathrm{d}t} = \frac{\hat{H}(t)}{i\hbar}|\phi_\alpha,t\rangle \quad \text{and} \quad \frac{\mathrm{d}\langle\phi_\alpha,t|}{\mathrm{d}t} = -\langle\phi_\alpha,t|\frac{\hat{H}(t)}{i\hbar} \,, \tag{8.466}$$

where we allowed for the time dependence of $\hat{H}$ for now. Using these equations in (8.465), we arrive at the quantum mechanical version of **Liouville's theorem**:

$$\frac{\mathrm{d}\hat{\rho}_t}{\mathrm{d}t} + \frac{1}{i\hbar}[\hat{\rho}_t,\hat{H}(t)] = 0 \,. \tag{8.467}$$

It should be emphasized that (8.467) is a condition any density operator must satisfy *regardless* of whether the system is in statistical equilibrium or not. This is quite analogous to what was said about (3.33) in classical statistical mechanics.

So, the necessary and sufficient condition of statistical equilibrium is that *both* (8.456) and (8.467) hold. But, this is equivalent to demanding both (8.456) and

$$[\hat{\rho}_t, \hat{H}(t)] = 0 \,. \tag{8.468}$$

This equation is the quantum mechanical version of (3.37) and, in view of Theorem 8.4, implies that $\hat{\rho}_t$ is diagonal in the H-representation. For this to be the case, it is sufficient that $\hat{\rho}_t$ be a function of $\hat{H}$ only:

$$\hat{\rho}_t = f(\hat{H}(t)) \,. \tag{8.469}$$

Equation (8.456) then demands that $\hat{H}$ be independent of time.

Exercise 8.33. If you have read Sects. 8.10 and 8.11, this exercise provides an alternative derivation of (8.467) and (8.468).

a. Show that

$$\hat{\rho}_t = \hat{U}_t \hat{\rho}_0 \hat{U}_t^\dagger \,. \tag{8.470}$$

b. Using (8.290) and (8.470), derive (8.467).
c. Recall (8.298) and show that

$$\mathrm{Tr}\{\hat{\rho}_t \hat{A}\} = \mathrm{Tr}\{\hat{\rho}_0 \hat{A}_H(t)\} \,. \tag{8.471}$$

d. Using (8.306), show that

$$\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} = \mathrm{Tr}\{[\hat{H}(t), \hat{\rho}_t]\hat{A}\} \,, \tag{8.472}$$

where $\hat{A}$ is a Hermitian operator that is independent of time. Demanding that (8.452) holds for any $\hat{A}$, we arrive at (8.468). ///

8.16.4 Canonical Ensemble

Restricting ourselves for systems with time-independent $\hat{H}$, we define quantum mechanical **canonical ensemble** by

$$\hat{\rho} = \frac{1}{Z} e^{-\beta \hat{H}} \,, \tag{8.473}$$

where we dropped the subscript t. Since any function of $\hat{H}$ commutes with $\hat{H}$, this expression clearly satisfies (8.468) and is an example of (8.469).

Using the orthonormal basis of energy eigenkets, that is, the H-representation, and recalling (8.131), we can express $\hat{\rho}$ as

$$\hat{\rho} = \frac{1}{Z} \sum_n |n\rangle e^{-\beta E_n} \langle n| \,. \tag{8.474}$$

Comparing this expression with (8.449), we see that the probability in this ensemble that a given copy is in the energy eigenstate $|n\rangle$ belonging to the eigenvalue E_n is $e^{-\beta E_n}/Z$.

The normalization constant Z is the **canonical partition function** and is given by

$$Z = \mathrm{Tr}\{e^{-\beta \hat{H}}\} \,. \tag{8.475}$$

In the H-representation, this reads

$$Z = \sum_n \langle n|e^{-\beta \hat{H}}|n\rangle = \sum_n e^{-\beta E_n} \,, \tag{8.476}$$

where the sum is over all energy eigenstates labeled by n. We can also write Z as

$$Z = \sum_{E_n} g(E_n) e^{-\beta E_n} \,, \tag{8.477}$$

where the sum is now over all distinct energy eigen*values* and $g(E_n)$ is the degeneracy of the eigenvalue E_n, that is, the number of linearly independent eigenkets corresponding to E_n.

Using the Dirac δ-function, we can rewrite (8.477) as

$$Z = \sum_{E_n} g(E_n) \int \delta(E - E_n) e^{-\beta E} \mathrm{d}E = \int \sum_{E_n} g(E_n) \delta(E - E_n) e^{-\beta E} \mathrm{d}E \,. \tag{8.478}$$

As we have seen in Chaps. 4 and 5,

$$\overline{\Omega}(E) := \sum_{E_n} g(E_n) \delta(E - E_n) \tag{8.479}$$

is nothing but the **density of states** when the possible values of the system energy E form a discrete spectrum. This observation allows us to write (8.477) as

$$Z = \int \overline{\Omega}(E) e^{-\beta E} \mathrm{d}E \,, \tag{8.480}$$

in agreement with our earlier result of classical statistical mechanics. See (4.72) in particular.

Exercise 8.34. The Hamiltonian of a two-state system is given by

$$\hat{H} = ia\,[|\phi_1\rangle\langle\phi_2| - |\phi_2\rangle\langle\phi_1|] \,, \tag{8.481}$$

where $\langle\phi_i|\phi_j\rangle = \delta_{ij}$ and a is a positive constant having the dimension of energy:

a. Find the energy eigenvalues and the corresponding eigenkets as linear combinations of $|\phi_1\rangle$ and $|\phi_2\rangle$.
b. Suppose that you have an ensemble of the identical systems. At some instant, 30 % of them are in the state $|\phi_1\rangle$ and the rest is in the state $|\phi_2\rangle$. Construct the density matrix and calculate the average energy for the ensemble at that instant.

c. Find the matrix representation of $\hat{\rho}$ in the H-representation.
d. The system is brought to contact with a thermal bath at temperature T, thereby changing $\hat{\rho}$. After a sufficient time has passed, the system has reached statistical equilibrium. Find the Helmholtz free energy and the internal energy of the system. ///

8.16.5 Ideal Gas and Classical Limit

Let us compute the canonical partition function for a particle of mass m confined to a rectangular box of dimension $L_x \times L_y \times L_z$. As shown in Example 8.13, the energy eigenvalues are determined by (8.363).

For a given value of E, (8.363) defines an ellipsoid with semi-principal axes of length $\sqrt{8mE/h^2}L_x$, $\sqrt{8mE/h^2}L_y$, and $\sqrt{8mE/h^2}L_z$ in the three-dimensional space defined by n_x-, n_y-, n_z-axes. (If $L_x = L_y = L_z = L$, the ellipsoid becomes a sphere of radius $\sqrt{8mE/h^2}L$.) Now, each point in the $n_x n_y n_z$-space *whose coordinates are positive integers* represents a distinct energy eigenstate. The number of such points on or inside the ellipsoid gives the number $\mathscr{W}(E)$ of distinct energy eigenstates whose energy is equal to or less than E. But, because each point occupies a unit volume, $\mathscr{W}(E)$ is approximately 1/8 of the volume of the ellipsoid. In the factor $1/8 = (1/2)^3$, the first factor of $1/2$ removes the half of the ellipsoid, which corresponds to $n_x \leq 0$. The second $1/2$ removes the half of the remaining portion of the ellipsoid, for which $n_y \leq 0$, and so on.

We note that the ellipsoid with semi-principal axes of length a, b, and c is obtained from the unit sphere by stretching it by factors of a, b, and c in x-, y-, and z-directions, respectively. So, the volume of the ellipsoid is $4\pi abc/3$. This leads to

$$\mathscr{W}(E) \approx \frac{\pi}{6}\left(\frac{8mE}{h^2}\right)^{3/2} L_x L_y L_z = \frac{\pi}{6}\left(\frac{8mE}{h^2}\right)^{3/2} V\,, \tag{8.482}$$

where $V = L_x L_y L_z$ is the volume of the box. We can now compute the density of states as

$$\overline{\Omega}(E) = \frac{\partial \mathscr{W}}{\partial E} \approx \frac{\pi}{4}\left(\frac{8m}{h^2}\right)^{3/2} V E^{1/2}\,. \tag{8.483}$$

Using this expression in (8.480) and recalling the formula (3.92), we arrive at

$$Z = \frac{V}{\Lambda^3} \tag{8.484}$$

in agreement with the prediction of the classical statistical mechanics, that is, (3.86) with the correction factor h^{-3}.

Suppose now that we have N noninteracting identical particles in the box. In this case, (8.363) is replaced by

$$\frac{8mE}{h^2} = \sum_{i=1}^{N} \left[\left(\frac{n_{xi}}{L_x} \right)^2 + \left(\frac{n_{yi}}{L_y} \right)^2 + \left(\frac{n_{zi}}{L_z} \right)^2 \right] \tag{8.485}$$

with subscript i labeling a particle. This equation defines an ellipsoid in a $3N$-dimensional space. Recalling (4.47), which gives the volume of a n-dimensional unit sphere, we find

$$\mathscr{W}(E) = \frac{1}{8^N N!} \frac{\pi^{3N/2}}{\Gamma(3N/2+1)} \left(\frac{8mE}{h^2} \right)^{3N/2} V^N \tag{8.486}$$

in place of (8.482). Here, the $1/8^N$ factor selects out the portion of the ellipsoid corresponding to the positive values of $n_{x1}, \ldots, n_{zN}$ and the $1/N!$ factor arises from the indistinguishability of identical particles. Thus,

$$\overline{\Omega}(E) = \frac{1}{8N!} \frac{\pi^{3N/2}}{\Gamma(3N/2+1)} \left(\frac{8m}{h^2} \right)^{3N/2} V^N \frac{3N}{2} E^{3N/2-1} . \tag{8.487}$$

With the aid of (4.42) and (4.43), this leads to

$$Z = \frac{V^N}{\Lambda^{3N} N!} \tag{8.488}$$

in agreement with what you saw in Exercise 3.16 for the classical mechanical system.

According to quantum mechanics, a particle is either a **fermion** following a **Fermi–Dirac statistics** or a **boson** following a **Bose–Einstein statistics**. No more than one fermion can occupy the same quantum state (specified by the triplet of numbers n_x, n_y, and n_z for the ideal gas under consideration) at any instant. No such restriction applies to bosons. Yet, (8.488) fails for bosons at high densities. This is because the $1/N!$ factor can no longer properly account for the indistinguishability of identical particles if more than one particle occupies a given quantum state. These are the results of the so-called **exchange effect**, which is present even if the Hamiltonian has no potential energy term due to interaction among particles, and lead to net repulsion among fermions and net attraction among bosons.

If the system is sufficiently dilute, the exchange effect can safely be ignored since there will be far more quantum states than there are particles and the probability that any given quantum state is occupied is very small. This is called the **classical limit**. Because our treatment of the ideal gas ignores the exchange effect, (8.488) is applicable only in this limit. Further details on Fermi–Dirac and Bose–Einstein statistics can be found in Refs. [2, 3, 5].

8.16.6 Microcanonical Ensemble

Following the approach we took in Chap. 4, we introduce a microcanonical ensemble as an approximation to a canonical ensemble when the system under consideration is macroscopic.

In a macroscopic system, the difference between two consecutive energy eigenvalues is very small. This being the case, it makes sense to lump together the densely distributed δ-functions in (8.479) and regard $\overline{\Omega}(E)$ as a continuous function of E. More explicitly, (8.479) may be replaced by the approximate expression

$$\overline{\Omega}(E) \approx \frac{1}{\delta E} \int_{E-\delta E}^{E} \sum_{E_n} g(E_n)\delta(E - E_n)\mathrm{d}E \,, \tag{8.489}$$

without incurring any sensible loss of accuracy in (8.480), where δE is a small interval of energy that is nevertheless much larger than the difference in two consecutive energy eigenvalues.

This modified density of states is usually a very rapidly increasing function of E. When such a function is multiplied by the very rapidly decreasing function $e^{-\beta E}$, where $\beta > 0$, the result is often a very sharply peaked function of E.[48] If this happens, we can take an small interval $(E - \Delta E, E]$ that contains the sharp peak and safely assume that essentially all of the copies in the canonical ensemble belongs to this narrow interval of the energy. For sufficiently small ΔE, all states in the interval $E - \Delta E < E_n \leq E$ may be regarded as equally probable. This then defines our **microcanonical ensemble**. In the H-representation, therefore,

$$\hat{\rho} = \frac{1}{C_M} \sum_n |n\rangle f(E_n) \langle n| \,, \tag{8.490}$$

where

$$f(E_n) = \begin{cases} 1 & \text{if } E - \Delta E < E_n \leq E \\ 0 & \text{otherwise.} \end{cases} \tag{8.491}$$

The quantity C_M is the normalization constant and is determined by the normalization condition (8.450), which now yields the **microcanonical partition function**

$$C_M = \sum_n f(E_n) \,, \tag{8.492}$$

where the summation is over all energy eigenstates. Because of (8.491), C_M is the total number of energy eigenstates corresponding to the energy eigenvalues that fall into the interval $E - \Delta E < E_n \leq E$. Thus, using ΔE for δE in (8.489),

$$C_M = \int_{E-\Delta E}^{E} \sum_{E_n} g(E_n)\delta(E - E_n)\mathrm{d}E = \overline{\Omega}(E)\Delta E \,, \tag{8.493}$$

which is nothing but (4.24).

As we saw in Chap. 4, other statistical ensembles can be derived from (8.493) and the physical interpretation given to $\overline{\Omega}(E)\Delta E$. Our derivation in that chapter made no reference to the underlying mechanics except in the case of the isothermal–isobaric ensemble.[49] Any difference between the classical and the quantum mechanical formulations of statistical mechanics is thus seen to arise in part from the form of $\overline{\Omega}(E)$ prescribed by the underlying mechanics. We have already encountered a profound implication of this difference in Example 4.2. The other difference stems from the exchange effect that impacts how a quantum mechanical system populates those microstates enumerated by $\overline{\Omega}(E)$. This is then a logical place for us to end this brief excursion into quantum mechanics and its application in formulating statistical mechanics.

8.17 Frequently Used Symbols

$\dagger$, adjoint.
c^* , complex conjugate of c.
$\langle A \rangle$, ensemble average of a dynamical variable A.
$\langle A \rangle_{\text{QM}}$, quantum mechanical average of A.
$|n\rangle$, energy eigenket corresponding to the $n$the energy eigenvalue E_n.
$|\boldsymbol{p}\rangle$, momentum eigenket.
$|\boldsymbol{x}\rangle$, position eigenket.
$[\hat{X}, \hat{Y}]$, commutator defined by $\hat{X}\hat{Y} - \hat{Y}\hat{X}$.
$[\hat{X}, \hat{Y}]_+$, anti-commutator defined by $\hat{X}\hat{Y} + \hat{Y}\hat{X}$.
$\text{Tr}\{\hat{A}\}$, trace of $\hat{A}$.

$\hat{a}$, annihilation operator.
$\hat{a}^\dagger$, creation operator.
$g(E_n)$, degeneracy of the energy eigenvalue E_n.
$\hbar$, $h/2\pi$, where $h = 6.626 \times 10^{-34}$ (J·s) is the Planck constant.
i , imaginary unit $\sqrt{-1}$.
m , mass of a particle.
$\hat{\boldsymbol{p}}$, momentum operator.
r , dimension of a ket space V_k.
s , degeneracy of an eigenvalue.
$\hat{\boldsymbol{x}}$, position operator.

$\hat{A}$, $\hat{B}$,..., observables.
C_M , microcanonical partition function.
E_n , the nth energy eigenvalue.
$\hat{H}$, Hamiltonian operator.
$\hat{I}$, identity operator.
$\hat{N}$, number operator.

$\hat{U}_t$, time evolution operator that advances the state ket from $t = 0$ to t.
$\hat{X}$, $\hat{Y}$,..., generic linear operator.
Z , canonical partition function.

$\delta(x)$, the Dirac δ-function.
$\hat{\rho}_t$, density operator at time t.
Ω , density of states.

References and Further Reading

1. Ballentine L E (1998) Quantum Mechanics: A Modern Development. World Scientific, River Edge, New Jersey
   For a more careful discussion of infinite dimensional vector space and operators with a continuous spectrum, see Sects. 1.3 and 1.4.
2. Goodstein D L (1985) States of matter. Dover, New York
   Chapter 2 provides a highly readable account of Fermi–Dirac and Bose–Einstein statistics.
3. Landau L D, Lifshitz E M (1980) Statistical physics: Part 1, 3rd edn. Pergamon Press, New York
   A detailed treatment of Fermi–Dirac and Bose–Einstein statistics is found in Chap. 5.
4. Macdonald A (2011) Linear and Geometric Algebra. Printed by CreateSpace
   The first four chapters offer a very compact and highly accessible introduction to linear algebra in just 70 pages.
5. Pathria R K (1972) Statistical Mechanics. Pergamon, Elmsford, New York
   For a detailed consideration of classical limit, see Sect. 5.5.
6. Pauling L, Wilson, Jr. E B (1985) Introduction to Quantum Mechanics with Applications to Chemistry. Dover, New York
   A very lucid and detailed treatment of the one-dimensional harmonic oscillator is found in Chap. 3.
7. Sakurai J J (1985) Modern Quantum Mechanics. Addison-Wesley, Reading, Massachusetts
   The need for defining a quantum mechanical state as a vector in a complex vector space is demonstrated using the Stern-Garlach experiment at the very beginning of the book. Our own outline of quantum mechanics followed the first two chapters of the book, but focusing only on those topics essential for our purpose. The density matrix is treated in some detail in Sect. 3.4 of the book.
8. Schlosshauer M (2007) Decoherence and the Quantum-to-Classical Transition. Springer-Verlag, Heidelberg
9. Tolman R C (1979) The principles of statistical mechanics. Dover, New York
   Chapters 7 and 8 provides a very thorough introduction to quantum mechanics. For a summary of relevant experimental discoveries that lead to quantum mechanics, see Sect. 52.