Some finer points of quantum mechanics

7

In this chapter, central concepts in quantum mechanics such as states and operators are reconsidered and defined more carefully. This reveals a large arbitrariness in the choice of representations for states and operators, related to each other by unitary transformations. Even the way in which the time evolution of the systems is described can be changed so that, for example, the time dependence is carried by operators (the Heisenberg picture) rather than by state vectors (the Schrödinger picture). Also, an important generalization of the concept of the quantum state itself is introduced, which allows us to describe more general physical situations where a complete quantum-mechanical description is unavailable. The "state" in such a case is known as a *mixed state* and is described by a *density matrix*, as compared to a *pure state* represented by a wave function.

7.1 Representations 151
7.2 States and operators 155
7.3 Unbounded operators 158
7.4 Unitary transformations 167
7.5 The Heisenberg picture 169
7.6 The uncertainty principle 172
7.7 Mixed states and the density matrix 173
7.8 Quantization in general coordinates 178
Further reading 182
Guide to the Supplements 182
Problems 182

7.1 Representations

There is a wide freedom in the way states and dynamical variables are described in quantum mechanics; the physical results must be independent of the particular language used, called the *representation*. A somewhat similar arbitrariness in the choice of language also exists in classical mechanics, such as the wide class of possible choices of canonical variables, which are related to each other by canonical transformations. However, the possibility of using different representations according to the problem considered appears to have much more profound consequences and implications in quantum mechanics.

Such a freedom of choice of language means that central concepts such as states and operators, time evolution of the system, etc., must be defined in a more general and abstract way than has been done so far in this book (the so-called Schrödinger representation). Different descriptions are usually related by a *unitary transformation*. The theory of unitary transformations, besides providing us with a conceptual clarification of the whole construction of quantum mechanics, often offers a powerful method of solution.

The principles of quantum mechanics developed so far may be summarized as follows:

P1 A separable Hilbert space $\mathscr{H}$ is associated with each quantum-

mechanical system. A quantum state is described by the ray in this space, i.e., by a unit-norm vector (wave function), modulo the phase.

P2 A Hermitian (self-adjoint) operator A in $\mathscr{H}$ is associated with each observable $\mathcal{A}$.

P3 The expectation value of an observable $\mathcal{A}$ in a state is given by $\langle\psi|A|\psi\rangle$. If a_k stands for an eigenvalue of A and $|k,\alpha\rangle$ the corresponding eigenvector (the index α labels possible degenerate states), the probability of finding the value a_k in a measurement of $\mathcal{A}$ in the state $|\psi\rangle$ is given by $\sum_\alpha |\langle k,\alpha|\psi\rangle|^2$.

P4 There exists a Hermitian (self-adjoint) operator H, the Hamiltonian, which determines the time evolution according to the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi\rangle\,.$$

P5 Quantum operators Q, P are associated with the classical canonical variables q, p in any cartesian coordinates. They obey Heisenberg's commutation relation,

$$[Q,P] = i\hbar.$$

The aim of this chapter is to make these rules more precise, qualifying them a little more carefully. Another, important issue in this chapter is the *generalization* of the notion of the quantum state itself, point [P1] above, to include the cases of mixed states.

7.1.1 Coordinate and momentum representations

We begin with an example of the change of representation, which will illustrate well the points made in the introduction to this chapter.

In Schrödinger's approach the wave function $\psi(x,t)$ represents a quantum state. Here x plays the role of the *parameter*, in the sense that it helps us to specify the probability amplitude of finding the particle at various spatial points, but the *state* is really represented by the whole function ψ. But x plays also another role: it is the (improper) eigenvalue of the position operator $\hat{q}$ defined by

$$\hat{q}\,\psi(q) = q\,\psi(q)\,.$$

This operator has the eigenfunctions $f_q(x) = \delta(q-x)$

$$f_q(x) = \delta(q-x)\,, \qquad \hat{q}\,\delta(q-x) = x\,\delta(q-x) = q\,\delta(q-x)\,,$$

as we have seen already.

It helps perhaps to draw an analogy with vectors in a finite-dimensional space: the components v_i of a vector can be defined, in terms of an orthonormal basis of unit vectors e_i, as various projections $v_i = (e_i, v)$ (here $(*,*)$ stands for the scalar product between two vectors), but naturally the vector v has a well-defined meaning independently of how they are expressed, using components referring to a particular basis.

The idea is similar here: we should really consider the "state vector" ψ or the scalar product of two of state vectors independent of the particular basis chosen. Just as the scalar product

$$(a, b) = a_i b_i$$

between two finite-dimensional vectors

$$(a, b) = a_i b_i$$

has a basis-independent meaning, the quantum-mechanical state vectors and their Hermitian scalar products are independent of the particular basis chosen.

Thus one can write

$$\psi(x) \quad = \quad \int dx' \, \delta(x' - x)\psi(x') = \int dx' \, f_x^*(x')\psi(x') \tag{7.1}$$

and interpret (7.1) as a scalar product

$$\langle x|\psi\rangle \,,$$

having defined the scalar product as

$$\int dx' \, \phi^*(x')\chi(x') \equiv \langle\phi|\chi\rangle, \tag{7.2}$$

where a notation due to Dirac has been introduced in eqn (7.2).

The state vector $|q\rangle$, the position eigenvector, plays the role of the basis of abstract vectors, whose components are $\langle x|q\rangle = f_q(x) = \delta(x-q)$. In order to have a complete analogy with finite-dimensional vectors, it is necessary that the basis vectors form an orthonormal and complete set (so that *any* state can be expressed in terms of their linear combination). The latter is just the eqn (7.1), while the first is

$$\langle q|q'\rangle = \int dx \, f_q^*(x) \, f_{q'}(x) = \int dx \, \delta(q - x) \, \delta(q' - x) = \, \delta(q - q') \,,$$

which is similar to $(e_i, e_j) = \delta_{ij}$.

The idea is then that the construction can be done by using a different variable than q (which is the case for the Schrödinger representation, or *coordinate representation*). In *momentum representation*, we first consider the eigenvectors of $\hat{p}$ in the x representation,

$$f_p(x) = \frac{1}{\sqrt{2\pi\hbar}} \, e^{ipx/\hbar}, \tag{7.3}$$

which satisfy the orthonormality condition

$$\langle p|p'\rangle = \delta(p - p') \,. \tag{7.4}$$

The abstract state vector (or "ket") $|p\rangle$ can be defined such that $\hat{p} \, |p\rangle = p \, |p\rangle$, whose x representation is given by eqn (7.3).

A generic state $|\psi\rangle$ can be written in the momentum representation as

$$\boxed{\tilde{\psi}(p) = \langle p|\psi\rangle = \int dx\, f_p(x)^* \psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int dx\, e^{-ipx/\hbar} \psi(x)\,.}$$

The "wave function" in the p representation is nothing but a Fourier component of the usual wave function. Going to the p representation from the x representation thus amounts to the Fourier transform. In particular, the position eigenstate, with eigenvalue x', is given in the p representation by

$$\langle p|x\rangle = \int dx'\, f_p^*(x')\delta(x - x') = f_p^*(x) = \frac{1}{\sqrt{2\pi\hbar}}\, e^{-ipx/\hbar} \tag{7.5}$$

and, by virtue of the property of a Hermitian scalar product, $\langle p|x\rangle = \langle x|p\rangle^*$.

As another example, consider the n-th energy eigenstate of a harmonic oscillator in the two representations. In the x representation we obtained earlier,

$$\langle x|n\rangle = \psi_n(x) = C_n\, H_n(\alpha\, x)\, e^{-\alpha^2 x^2/2} \tag{7.6}$$

(see Chapter 3). The same state is described in the p representation by the wave function

$$\psi(p) = \langle p|n\rangle = \int dx\, \langle p|x\rangle\langle x|n\rangle = \frac{C_n}{\alpha\, \hbar^{1/2}}\, (-i)^n\, H_n\left(\frac{p}{\alpha\hbar}\right) e^{-p^2/2\alpha^2\hbar^2}\,, \tag{7.7}$$

which is the Fourier transform of (7.6).

The effect of the change of basis can also be seen on operators. In the momentum representation, the momentum is represented by a multiplicative operator, $\hat{p} = p$, whereas the position operator becomes a differential operator

$$\hat{x} = i\hbar \frac{\partial}{\partial p}.$$

Note the (crucial) difference in the sign with respect to the more familiar expression $\hat{p} = -i\hbar\frac{\partial}{\partial x}$ of the momentum in the x representation. This is necessary so that the fundamental commutation relation

$$[\hat{x}, \hat{p}] = i\hbar$$

is valid in any representation.

We see that in particular the harmonic oscillator Hamiltonian in the p representation takes the form

$$H = -\frac{1}{2} m\, \omega^2\, \hbar^2\, \frac{\partial^2}{\partial p^2} + \frac{p^2}{2\, m},$$

which has the same form as the standard one in the x representation, after the replacement,

$$m \to M = \frac{1}{m\, \omega^2}, \qquad \omega \to \omega,$$

and this explains the simple result (7.7): it could have been obtained just by inspection, without really performing any Fourier transformation, from eqn (7.6)!

7.2 States and operators

7.2.1 Bra and ket; abstract Hilbert space

The preceding discussion clearly highlights a fact which is conceptually important: a quantum state is described by (the ray of) an abstract vector, called a *ket*,

$$|\psi\rangle.$$

It is also convenient to introduce a kind of conjugate, $\langle\psi|$, called the *bra* vector. These terminologies were introduced by Dirac, from the English word "bracket". The description of "ψ" in terms of a complex function (e.g., of x) is just a possible representation. The operators, the equation of motion, etc., must be defined in the space of such abstract state vectors. We shall study first the properties of this space, leaving the study of relations among different representations (unitary transformations) to subsequent sections.

The required properties of the space $\mathscr{H}$ of vectors representing possible quantum states of a system are:

[1] $\mathscr{H}$ is a vector space.

[2] A scalar product $\langle\chi|\phi\rangle$, which is a complex number, is defined for each pair of vectors in $\mathscr{H}$.

[3] $\mathscr{H}$ is complete.

[4] $\mathscr{H}$ is separable.

A space with these properties is known as a (separable) *Hilbert space*.[1]

[1] The concept of a Hilbert space was introduced by D. Hilbert ($\sim$ 1910), as generalizations of the finite-dimensional Euclidean space R^n (with elements $(x_1, x_2, \ldots, x_n)$) in the $n \to \infty$ limit. A Hilbert space inherits many of its properties from the latter, in a natural fashion.

[1] $\mathscr{H}$ is a vector space

This property means that, given two vectors $|\psi\rangle$, $|\phi\rangle$ in $\mathscr{H}$, any combination

$$c\,\psi + d\,\phi \in \mathscr{H}, \qquad c, d \in \mathbb{C} \tag{7.8}$$

belongs to $\mathscr{H}$ also. In other words, in $\mathscr{H}$ the sum among vectors and multiplication with complex numbers are defined, with the standard rules:

$$\psi + \phi = \phi + \psi\,; \qquad (\psi + \phi) + \chi = \psi + (\phi + \chi)\,;$$
$$c\,(\psi + \phi) = c\,\psi + c\,\phi\,; \qquad (c\,d)\,\psi = c\,(d\,\psi)\,;$$
$$0 \cdot \psi = \mathbf{0}\,; \qquad 1 \cdot \psi = \psi.$$

Note in particular that a null vector $\psi - \psi = \mathbf{0}$ exists in $\mathscr{H}$. The vectors $\psi_1, \psi_2, \ldots, \psi_k$ are *linearly independent* if

$$c_1\psi_1 + c_2\psi_2 + \ldots + c_k\psi_k = 0, \tag{7.9}$$

implies

$$c_1 = c_2 = \ldots = c_k = 0.$$

The maximum number of linearly independent vectors is the *dimension* of $\mathscr{H}$.

[2] A scalar product $\langle\chi|\phi\rangle$ is defined for each pair of vectors in $\mathscr{H}$

For each pair of vectors in $\mathscr{H}$, ψ, and ϕ, a (Hermitian) scalar product $\langle\phi|\psi\rangle \in \mathbb{C}$ (a complex number) is defined such that

$$\begin{aligned} \langle\phi|c_1\psi_1 + c_2\psi_2\rangle &= c_1\langle\phi|\psi_1\rangle + c_2\langle\phi|\psi_2\rangle; \\ \langle\phi|\psi\rangle^* &= \langle\psi|\phi\rangle; \\ \langle\psi|\psi\rangle &\geq 0, \quad (= 0, \quad \text{if and only if} \quad |\psi\rangle = 0\ ). \end{aligned} \tag{7.10}$$

The first two relations imply that[2]

$$\langle c_1\psi_1 + c_2\psi_2|\phi\rangle = c_1^*\langle\psi_1|\phi\rangle + c_2^*\langle\psi_2|\phi\rangle.$$

[2] Sometimes different notations are used in the literature, e.g. (ϕ, ψ) or (ψ, ϕ), instead of $\langle\phi|\psi\rangle$.

In coordinate representation the scalar product can be written explicitly as:

$$\langle\phi|\psi\rangle = \int dq\, \phi^*(q)\psi(q)\,,$$

where $dq \equiv d^n x$ if the coordinate degrees of freedom is n. For instance $n = 6$ for a two-particle system in three dimensions.

The positivity of the scalar product (7.10) allows us to introduce the *norm* of a state vector by

$$\|\psi\| \equiv \sqrt{\langle\psi|\psi\rangle}.$$

The introduction of the norm—the length of each vector—in $\mathscr{H}$ implies that the "distance" between any pair of states can be defined naturally as

$$\|\psi - \phi\| = \sqrt{\langle\psi - \phi|\psi - \phi\rangle}. \tag{7.11}$$

$\mathscr{H}$ is thus a *metric* space. In such a space one can introduce the concept of the limit of a sequence, $\{\psi_n\} = \psi_1, \psi_2, \ldots$, by Cauchy's criterion:

$$\lim_{n\to\infty} \psi_n = \psi \quad \text{exists if} \quad \forall\,\epsilon > 0,\ \exists\, N > 0 : \quad \forall n > N\,,\ \ ||\psi_n - \psi|| < \epsilon\,.$$

Any acceptable definition of a distance must be such that for any three points (chosen here as $\mathbf{0}$, ψ, and ϕ) the relation

$$\|\psi - \phi\| \leq \|\psi\| + \|\phi\| \tag{7.12}$$

holds (the triangular inequality). Equation (7.12) can also be written as (ϕ to $-\phi$):

$$\|\psi + \phi\| \leq \|\psi\| + \|\phi\|. \tag{7.13}$$

It is not difficult to show that eqn (7.13) is satisfied by definition (7.11). Note first that

$$\langle\psi+\phi|\psi+\phi\rangle = \|\psi\|^2 + \|\phi\|^2 + 2\mathrm{Re}\langle\phi|\psi\rangle.$$

But for any complex number the relation

$$\mathrm{Re}\langle\phi|\psi\rangle \leq |\langle\phi|\psi\rangle|$$

holds, so we need only to demonstrate that

$$|\langle\phi|\psi\rangle| \leq \|\phi\|\|\psi\|. \tag{7.14}$$

(the Schwarz inequality).

To show eqn (7.14), it suffices to consider a vector

$$\tilde{\phi} \equiv \phi - \psi\cdot\sqrt{\langle\psi|\phi\rangle}\|\psi\|^2.$$

Equation (7.14) is a simple consequence of the fact that $\tilde{\phi}$ is semi-positive definite. Note also that the equality holds if and only if $c_1\psi = c_2\phi, \quad c_1, c_2 \in \mathbb{C}$.

[3] $\mathscr{H}$ is complete

A Cauchy sequence is

$$\forall\,\epsilon>0\,,\,\exists N\;:\quad \forall\,n,m>N,\quad \|\psi_n-\psi_m\|<\epsilon\,.$$

The space $\mathscr{H}$ is complete if each Cauchy sequence converges in $\mathscr{H}$.

An example of a space which is not complete is the set of rational numbers $\mathbb{Q}$ which is a subspace of the real numbers $\mathbb{R}$.

A useful property of a complete space is that each closed subspace is complete. Recall that a closed space contains all the limit points. Then a closed vector subspace of $\mathscr{H}$ is complete, and is itself a Hibert space.

[4] $\mathscr{H}$ is separable

That is, there is a countable subensemble (base) $S \subset \mathscr{H}$, dense everywhere in $\mathscr{H}$. In other words, each vector $\psi \in \mathscr{H}$ is a limit of a sequence $\{\phi_n\}$ in S. (The set of rational numbers forms a numerable base and is dense everywhere in the space of real numbers, therefore $\mathbb{R}$ is separable.)

The most important consequence of [1]–[4] is the existence of a complete and orthonormal system of vectors in $\mathscr{H}$, $\{\psi_n\}$. Any vector in $\mathscr{H}$ can be written as

$$\psi = \lim_{N\to\infty}\sum_{n=0}^{N} c_n\psi_n \equiv \sum c_n\psi_n \qquad \text{where} \quad \langle\psi_i|\psi_j\rangle = \delta_{ij}\,. \tag{7.15}$$

The expansion coefficients c_n are given by

$$c_n = \langle\psi_n|\psi\rangle,$$

that is, for each vector the relation

$$|\psi\rangle = \sum_n |\psi_n\rangle\langle\psi_n|\psi\rangle$$

holds. This can be written symbolically as

$$\sum_n |\psi_n\rangle\langle\psi_n| = \mathbf{1},$$

which is the completeness relation seen already in Section 2.6.

We emphasize that not all linear combinations (7.15) define a vector in $\mathscr{H}$. The requirement of finite norm implies, by virtue of the orthogonality, that

$$\langle\psi|\psi\rangle = \sum_i |c_i|^2 < \infty\,.$$

In a Hilbert space of finite dimensions properties [3] and [4] are automatically satisfied. Vice versa, for infinite-dimensional Hilbert spaces these requirements are of fundamental importance.

7.3 Unbounded operators

The observables are described by operators acting in the abstract Hilbert space

$$A : \mathscr{H} \to \mathscr{H}. \tag{7.16}$$

We want the action of an operator to preserve the linear structure, so require it to be *linear*

$$A\ (\alpha_1|\varphi_1\rangle + \alpha_2|\varphi_2\rangle) = \alpha_1\, A\,|\varphi_1\rangle\ +\ \alpha_2\, A\,|\varphi_2\rangle\,.$$

If there exists a real positive number C such that

$$||A\,\psi|| < C\,||\psi||, \qquad \forall\psi \in \mathscr{H}, \tag{7.17}$$

the operator is said to be *bounded*; otherwise it is unbounded. The *norm* of the operator A, denoted as $||A||$, is defined as the lower limit of the constant C satisfying eqn (7.17), i.e.,

$$||A|| = \sup_{\psi\neq 0} ||A\,\psi||/||\psi||; \qquad \text{or} \qquad ||A|| = \sup_{||\psi||=1} ||A\,\psi||\,. \tag{7.18}$$

In other words, the operator is bounded if $||A|| < \infty$. It is easy to convince oneself that an operator A is continuous if and only if it is bounded. In a finite-dimensional Hilbert space all operators are bounded; in an infinite-dimensional space it is not the case.

We note that for linear operators the notion of continuity is global: if an operator is continuous in a point it is so everywhere in the Hilbert space. Vice versa, if it is unbounded it is discontinuous everywhere. This can be understood, simplifying a little, by noting that the condition of continuity at ψ, given by

$$\forall\,\epsilon > 0,\ \exists\,\delta > 0\ : \qquad \forall\,\varphi\ :\ ||\varphi - \psi|| < \delta\ \Rightarrow\ ||A\psi - A\varphi|| < \epsilon$$

can be translated into a statement about the origin 0, by writing it as $f = \varphi - \psi$ and $A(\psi) - A(\varphi) = A(\psi - \varphi) = A(f)$.

The fundamental commutation relation

$$[Q, P] = i\,\hbar \tag{7.19}$$

explains why unbounded operators occur as a rule in quantum mechanics. In fact, it follows from eqn (7.19) that

$$P\,Q^n - Q^n P = -i\,n\hbar\,Q^{n-1}\,, \tag{7.20}$$

and therefore

$$n\hbar\,||Q^{n-1}|| \leq 2||P\,Q^n|| \leq 2||P||\,||Q||\,||Q^{n-1}||,$$

where we used

$$||A\,B|| \leq ||A||\,||B||, \qquad ||A + B|| \leq ||A|| + ||B||,$$

valid for any pair of operators. Thus

$$2||P||\,||Q|| \geq n\,\hbar\,.$$

As the inequality is valid for any n, at least one of the operators P, Q must be unbounded.

Other common examples of unbounded operators are:

(i) The energy operator of a harmonic oscillator,

$$H = \frac{p^2}{2m} + \frac{1}{2} m\,\omega^2\,x^2,$$

is unbounded as there are states ψ_n such that $H\,\psi^{(n)} = E_n\,\psi^{(n)}$, $||\psi_n|| = 1$, with an arbitrarily large value of E_n.

(ii) The position operator x: for instance the states $\psi^{(n)} = \frac{e^{-x^2/2n^2}}{\pi^{1/4} n^{1/2}}$, are normalized but

$$||x\,\psi^{(n)}||^2 = \frac{n^2}{2}$$

can be arbitrarily large.

(iii) In one dimension $\psi(x) = \sqrt{\frac{a}{\pi}}\,\frac{1}{\sqrt{x^2+a^2}} \in \mathscr{H}$ but $x\,\psi \notin \mathscr{H}$.

As is clear from these examples, sometimes the action of an operator on a vector in $\mathscr{H}$ is not defined (i.e., does not generate a state in $\mathscr{H}$). It is therefore necessary to define the *domain* of an operator A, $\mathcal{D}(A)$, a subspace of $\mathscr{H}$ upon which A acts:

$$\psi \in \mathcal{D}(A) \subset \mathscr{H}, \qquad \text{if} \quad A\,\psi \in \mathscr{H}. \tag{7.21}$$

Definition (7.21) is flexible enough to allow for possible extension of the domain. One can define an *extension* of an operator A, B by

$$\mathcal{D}(A) \subset \mathcal{D}(B) \quad \text{e} \quad \forall\,|f\rangle \in \mathcal{D}(A)\ :\ A|f\rangle = B|f\rangle\,.$$

The necessity of dealing with unbounded operators renders it indispensable to require a more precise condition on the operators to be associated with dynamical variables.

First of all, recall the reasoning of the Subsection 2.3.2. An operator A is called *Hermitian* or *symmetric* if

$$\forall\, f, g \in \mathcal{D}(A) \quad \langle Af|g\rangle = \langle f|Ag\rangle. \tag{7.22}$$

The expectation value of a dynamical variable must be real in any state. The necessary and sufficient condition for this is that A be Hermitian.

In a finite-dimensional (N) space, this requirement is equivalent to the condition that the matrix representing the operator in a given basis is Hermitian. Let $\{|e_i\rangle\}$ be an orthonormal basis, and define the matrix as

$$A|e_i\rangle = \sum_{j=1}^{N} A_{ji} e_j. \tag{7.23}$$

From eqn (7.22) it follows immediately that $A_{ij} = A^*_{ji}$ e $A(\sum_i c_i|e_i\rangle) = (\sum_{ij}(A_{ij}c_j)|e_i\rangle$.

An important theorem in this context (known as the Hellinger–Toeplitz theorem) asserts that an operator defined everywhere in $\mathscr{H}$, with the property

$$\langle A\,\phi|\psi\rangle = \langle\phi|A\,\psi\rangle,$$

is necessarily bounded. It follows that for an unbounded operator the definition of "reality" requires more careful study.

7.3.1 Self-adjoint operators

If for any given vector $\psi \in \mathscr{H}$, another vector $\eta \in \mathscr{H}$ exists such that

$$\langle A\phi|\psi\rangle = \langle\phi|\eta\rangle\,, \quad \forall\phi \in \mathcal{D}(A)\,,$$

then we define

$$A^\dagger|\psi\rangle \equiv |\eta\rangle.$$

The existence of such a vector η defines $\mathcal{D}(A^\dagger)$. The operator $A^\dagger$ is called the *adjoint* of the operator A. This definition makes sense only if A is defined densely, that is, $\overline{\mathcal{D}(A)} = \mathscr{H}$. For otherwise let v be an element orthogonal to $\overline{\mathcal{D}(A)}$. We would have $\langle\phi|\eta\rangle = \langle\phi|\eta + v\rangle$, and thus the definition of η is not well defined.

Let us note that

$$\langle A\,\phi|\psi\rangle = \langle\phi|A^\dagger|\psi\rangle;$$

it follows from the definition of the scalar product that

$$\langle\phi|A^\dagger|\psi\rangle = (\langle\psi|A|\phi\rangle)^*.$$

In particular, if a matrix representation for A exists, one has $(A^\dagger)_{ij} = A^*_{ji}$, i.e., a Hermitian conjugate (adjoint) of A, known from linear algebra in finite-dimensional spaces. For a symmetric operator,

$$\langle A\varphi|\psi\rangle = \langle\varphi|A\psi\rangle \quad \text{for any vector} \quad \varphi, \psi \in \mathcal{D}(A)\,:$$

thus if $\varphi \in \mathcal{D}(A)$ holds, so does $\varphi \in \mathcal{D}(A^+)$. In other words, an operator is symmetric if

$$A^\dagger \psi = A \psi, \quad \forall \psi \in \mathcal{D}(A), \qquad \mathcal{D}(A) \subset \mathcal{D}(A^\dagger)\,. \tag{7.24}$$

If furthermore the condition

$$\mathcal{D}(A^\dagger) = \mathcal{D}(A) \tag{7.25}$$

is satisfied, the operator is said to be *self-adjoint*.

For any self-adjoint operator, the relation

$$\langle \psi | A | \psi \rangle = \langle \psi | A | \psi \rangle^*, \quad \forall \psi \in \mathcal{D}(A), \tag{7.26}$$

holds: i.e., its expectation value in any state (hence any of its eigenvalues) is real. By postulate,

Any dynamical variable is represented by a self-adjoint operator.

Such a requirement might appear at first sight somewhat arbitrary. Is the distinction between Hermitian operators and self-adjoint operators a physically irrelevant mathematical sophistication? Actually, apart from various mathematical requirements of consistency, this specification is physically based. Consider for instance an observable $\mathcal{A}$ represented by an operator A. A very reasonable physical requirement is that, whatever the correspondence between them, an observable which is a function $f(\mathcal{A})$ of $\mathcal{A}$ be associated with an *operator*, $f(A)$. It is precisely the self-adjoint operators among the symmetric operators which satisfy this requirement.

In most cases treated in this book we shall deal with a symmetric operator which is automatically also self-adjoint. Usually the question of the domains $\mathcal{D}(A) \subset \mathcal{D}(A^\dagger)$ arises in problems with nontrivial boundary conditions, and in these cases one must be careful how to extend the symmetric operator to a self-adjoint one. Sometimes the extension is not unique: we may find that there correspond more than one physically distinct quantum operators correspond to a given classical operator. See some examples discussed in the Supplement on Hilbert space.

The importance of having self-adjoint operators can be appreciated from the following two theorems, which we quote here without proof. The definition of unitary operators will be given below, in the Section 7.4.

Theorem 7.1 *Let A be a self-adjoint operator and*

$$U(t) \equiv e^{i\,t\,A} \tag{7.27}$$

with $t \in \mathbb{R}$ a continuous parameter. It follows then that

(a) *For t, s real, $U^\dagger(t)\,U(t) = \mathbf{1}$, $\quad U(t+s) = U(t)\,U(s)$.*

(b) *For any $\phi \in \mathscr{H}$ and for $t \to t_0$ $U(t)\,\phi \to U(t_0)\,\phi$ holds.*

(c) *For any $\psi \in \mathcal{D}(A)$ $\dfrac{U(t)\,\psi - \psi}{t} \overset{t\to 0}{\longrightarrow} i\,A\,\psi$ holds.*

(d) *If* $\lim_{t\to 0} \dfrac{U(t)\,\psi - \psi}{t}$ *exists, then* $\psi \in \mathcal{D}(A)$.

Theorem 7.2. (Stone's theorem) *If* $U(t)$ *is a unitary operator in* $\mathscr{H}$ *and strongly continuous in* t *(i.e., satisfies the properties (a) and (b) above), then there exists a self-adjoint operator* A *in* $\mathscr{H}$ *such that*

$$U(t) = e^{itA}. \tag{7.28}$$

The operators $U(t)$ described above are unitary, and the family of operators obtained by varying t is a one-parameter group of unitary operators. Stone's theorem shows the close connection between one-parameter groups of unitary operators and self-adjoint operators. The operator A in Stone's theorem is usually called *infinitesimal generator* of the transformation. The reader will recognize in these statements a natural generalization of analogous properties of transformation groups in finite-dimensional spaces; it can be said that the self-adjoint operators are those operators for which many of the properties of Hermitian matrices continue to hold. A simple but important property of a self-adjoint operator is that if B is self-adjoint and if $A = B^\dagger B$, then A is semi-positive definite:

$$\langle\psi|\, B^\dagger\, B\,|\psi\rangle \geq 0, \tag{7.29}$$

with the equality holding if and only if $B\,|\psi\rangle = 0$.

Our starting point in the analysis of the Schrödinger equation was the study of its eigenvalues, which also determines the results of possible energy measurements. It is here that the importance of self-adjoint operators makes its appearance.

In finite-dimensional spaces it is well known that a Hermitian matrix can be diagonalized by a unitary transformation. In such a basis the matrix takes the form

$$\begin{pmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \vdots \\ 0 & 0 & \ddots & \vdots \\ \vdots & \dots & \dots & \lambda_k \end{pmatrix} \tag{7.30}$$

Each eigenvalue λ_i might appear with a certain multiplicity n_i. An important point is that the decomposition (7.30) can be expressed in a form independent of the basis chosen. In fact, within each block of dimension $n_i \times n_i$ the matrix is a multiple of identity, and this is nothing but the projection onto the subspace associated with the eigenvalue λ_i. If the eigenvectors corresponding to the the first eigenvalue λ_1 are $e_1^{(a)}, e_1^{(b)}, \dots$, then the first part of the above matrix is

$$\lambda_1 \left(|e_1^{(a)}\rangle\langle e_1^{(a)}| + |e_1^{(b)}\rangle\langle e_1^{(b)}| + \dots \right) = \lambda_1 \Pi_1,$$

where Π_1 is the projection operator onto the subspace, spun by $e_1^{(a)}, e_1^{(b)}, \dots$. In general, one has a decomposition

$$A = \sum_i \lambda_i \Pi_i \tag{7.31}$$

known as the *spectral decomposition* of the matrix A. Note that the projection operators are mutually orthogonal and the identity

$$\mathbb{I}_{N\times N} = \sum_i \Pi_i$$

holds ($\mathbb{I}_{N\times N}$ is the identity matrix), which expresses the *completeness* of the basis of eigenvectors.

The spectrum of self-adjoint operators

For self-adjoint operators it is possible to write an analogue of eqn (7.31). The *spectrum* of a self-adjoint operator A is the ensemble of its proper eigenvalues (discrete eigenvalues) and improper eigenvalues (continuous eigenvalues). The first, the discrete eigenvalues, correspond to the values of λ such that

$$(A - \lambda_m)|\psi_m\rangle = 0; \quad \|\psi_m\| = 1, \quad m = 0, 1, 2, \ldots \tag{7.32}$$

as is well known.

For operators (and the systems) possessing a continuous spectrum, however, the criterion for deciding whether a value belongs to its spectrum must be expressed in a more general manner. Let us define the **resolvent set** of an operator A as the ensemble of the points $x \in \rho(A) \subset \mathbb{R}$ for which $A - x\,\mathbf{1}$ has a bounded inverse,

$$R(x) = (A - x\,\mathbf{1})^{-1}, \qquad ||R|| < \infty$$

(known as the resolvent operator of A). It is physically clear that the spectrum of the operator A is the *complement* of $\rho(A)$, $\sigma(A)$, which by definition is the set of all $x \notin \rho(A)$. The resolvent set is obviously an open set; it follows that the spectrum $\sigma(A)$ forms a closed set.

These conditions can be expressed in a form similar to eqn (7.32), known as

Weyl's criterion:

A value λ belongs to the spectrum of a self-adjoint operator A if and only if there exists a sequence ψ_N, such that

$$\boxed{\lim_{N\to\infty} \|A\psi_N - \lambda\psi_N\| = 0, \qquad \|\psi_N\| = 1\,.} \tag{7.33}$$

In other words, the (continuous or improper) eigenvalues of an operator are those values for which there exist functions in $\mathcal{H}$, *arbitrarily close to the concept of the eigenfunction* (see eqn (7.32)), even if the limit of such a sequence might not be a normalizable function. Obviously, a discrete eigenvalue λ_m satisfies Weyl's criterion trivially, with $\psi_N = \psi_m, \quad \forall N$.

For example, in the case of the momentum operator, the existence of the sequence

$$\psi_N = \frac{1}{\pi^{1/4}N^{1/2}} e^{ipx/\hbar} e^{-x^2/2N^2}, \quad N = 1, 2, \ldots \tag{7.34}$$

shows that all real values belong to the spectrum of $p = -i\hbar(d/dx)$. Analogously, for the position operator one finds that

$$\lim_{N\to\infty} \|(x - x_0)\psi_N\| = 0, \tag{7.35}$$

for the sequence[3]

$$\psi_N = (\frac{2\,N}{\pi})^{1/4}\, e^{-N\,(x-x_0)^2}, \quad N = 1, 2, \ldots. \tag{7.36}$$

[3]**Exercise.** Show that eqn (7.33) is satisfied by the operator $\hat{p} = -i\hbar(d/dx)$ with sequence (7.34) and with $\lambda = p$. Verify eqn (7.35).

The foremost consequence of the requirement of the self-adjointness of an operator A is the existence of the set of real (proper and improper) eigenvalues $\{\lambda_n, \lambda\}$ and *a family of projection operators* (assuming a continuous spectrum $[\lambda_0, \infty)$):

$$\mathcal{P}_c(\lambda) = \int_{\lambda_0}^{\lambda} d\lambda' \sum_{\alpha} |\lambda', \alpha\rangle\langle\lambda', , \alpha|, \qquad \mathcal{P}_n = \sum_{\alpha} |n, \alpha\rangle\langle n, \alpha| \,. \tag{7.37}$$

The index α runs over the possible degenerate sets. For simplicity we use Dirac's notation with the normalization

$$\langle\lambda, \alpha|\lambda', \beta\rangle = \delta_{\alpha\beta}\, \delta(\lambda - \lambda') \,.$$

If the degeneracy were of continuous type, an obvious change of notation would be needed for the sum over α. The existence of the projection operators does not depend on the basis chosen. In the following the formulas will be given both in terms of the projection operators and in a base of improper eigenvectors.

It is convenient to unify the two terms in eqn (7.37), by using

$$\mathcal{P}(\lambda) = \int_{\lambda_0}^{\lambda} d\lambda' \left[\sum_{\alpha} |\lambda', \alpha\rangle\langle\lambda', , \alpha| + \sum_{n} \delta(\lambda' - \lambda_n) \sum_{\alpha} |n, \alpha\rangle\langle n, \alpha|\right] =$$
$$= \int_{\lambda_0}^{\lambda} d\mathcal{P}(\lambda') \,.$$

The function $\mathcal{P}(\lambda)$ is discontinuous at the points corresponding to a discrete spectrum: i.e., has a jump at such points, and is understood continuous on the right of these point:

$$\lim_{\lambda\to\lambda_n^+} \mathcal{P}(\lambda) = \mathcal{P}(\lambda_n) = \int_{\lambda_0}^{\lambda_n} d\lambda' \sum_{\alpha} |\lambda', \alpha\rangle\langle\lambda', \alpha| + \sum_{\lambda_k \le \lambda_n} \sum_{\alpha} |k, \alpha\rangle\langle k, \alpha; |$$

$$\lim_{\lambda\to\lambda_n^-} \mathcal{P}(\lambda) = \mathcal{P}(\lambda_n) = \int_{\lambda_0}^{\lambda_n} d\lambda' \sum_{\alpha} |\lambda', \alpha\rangle\langle\lambda', \alpha| + \sum_{\lambda_k < \lambda_n} \sum_{\alpha} |k, \alpha\rangle\langle k, \alpha| \,.$$

The operators $\mathcal{P}$ are the (sum of orthogonal projection operators) and satisfy

$$\lambda_1 < \lambda_2 \quad \Rightarrow \quad \mathcal{P}(\lambda_1)\mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_1) \,, \tag{7.38}$$

which is a natural generalization of the familiar property of projection operators in the case of a discrete spectrum:

$$\mathcal{P}_n^2 = \mathcal{P}_n, \qquad \mathcal{P}_n\, \mathcal{P}_m = 0, \quad m \neq n \,.$$

The operators $\mathcal{P}$ satisfy (*the spectral theorem*):

$$\mathbf{1} = \int d\mathcal{P}(\lambda) = \int d\mathcal{P}_c(\lambda) + \sum_n \mathcal{P}_n; \tag{7.39a}$$

$$|\psi\rangle = \int d\mathcal{P}(\lambda)\,|\psi\rangle = \int d\mathcal{P}_c(\lambda)\,|\psi\rangle + \sum_n \mathcal{P}_n\,|\psi\rangle, \quad \forall\psi \in \mathscr{H}; \tag{7.39b}$$

$$A = \int \lambda\, d\mathcal{P}(\lambda) = \int \lambda\, d\mathcal{P}_c(\lambda) + \sum_n \lambda_n\, \mathcal{P}_n. \tag{7.39c}$$

In particular, as $A|\psi\rangle$ must belong to the Hilbert space, one has

$$\langle A\psi|A\psi\rangle = \langle\psi|A^2\psi\rangle < \infty \tag{7.40}$$

and thus the dispersion of $\mathcal{A}$ is finite. More explicitly,

$$\begin{aligned}\langle\psi|A^2\psi\rangle &= \int \lambda^2\; d\,||\mathcal{P}(\lambda)\psi||^2 = \int \lambda^2\; d\,||\mathcal{P}_c(\lambda)\psi||^2 + \sum_n \lambda_n^2\,||\mathcal{P}_n\psi||^2 \\ &\equiv \int d\lambda\,\lambda^2 \sum_\alpha |\langle\lambda,\alpha|\psi\rangle|^2 + \sum_n \lambda_n^2 \sum_\alpha |\langle n,\alpha|\psi\rangle|^2\,.\end{aligned}$$

The spectral theorem allows us to define the operator $f(\mathcal{A})$, given the operator representing an observable A, where $f(x)$ is any function of x, as

$$f(A) = \int f(\lambda)\, d\mathcal{P}(\lambda) = \int f(\lambda)\, d\mathcal{P}_c(\lambda) + \sum_n f(\lambda_n)\,\mathcal{P}_n\,. \tag{7.41}$$

For instance its mean value is then given by

$$\begin{aligned}\langle\psi|f(A)|\psi\rangle &= \int f(\lambda)\; d\,||\mathcal{P}(\lambda)\psi||^2 = \int f(\lambda)\; d\,||\mathcal{P}_c(\lambda)\psi||^2 + \sum_n f(\lambda_n)||\mathcal{P}_n\psi||^2 \\ &\equiv \int d\lambda\, f(\lambda) \sum_\alpha |\langle\lambda,\alpha|\psi\rangle|^2 + \sum_n f(\lambda_n) \sum_\alpha |\langle\lambda_n,\alpha|\psi\rangle|^2\,.\end{aligned} \tag{7.42}$$

We recognize in eqn (7.39a,b) the completeness relation. These relations in fact guarantee the consistency of the rules of quantum mechanics, in particular, postulates P3, P4. If the mean value of the observable $\mathcal{A}$ is given by the expectation value of the operator A in a given state, one finds that

$$\begin{aligned}\langle\psi|A|\psi\rangle &= \int \lambda\; d\,||\mathcal{P}(\lambda)\psi||^2 = \int \lambda\; d\,||\mathcal{P}_c(\lambda)\psi||^2 + \sum_n \lambda_n||\mathcal{P}_n\psi||^2 \\ &\equiv \int d\lambda\,\lambda \sum_\alpha |\langle\lambda,\alpha|\psi\rangle|^2 + \sum_n \lambda_n \sum_\alpha |\langle\lambda_n,\alpha|\psi\rangle|^2\,.\end{aligned} \tag{7.43}$$

In writing eqn (7.43) we have used the fact that the operators $\mathcal{P}(\lambda)$ are projection operators, so that

$$\langle\psi|d\,\mathcal{P}(\lambda)|\psi\rangle = d\,\langle\psi|\mathcal{P}(\lambda)\psi\rangle = d\,\langle\mathcal{P}(\lambda)\psi|\mathcal{P}(\lambda)\psi\rangle = d\,||\mathcal{P}(\lambda)\psi||^2\,.$$

Equation (7.43) allows us to interpret (see also Section 2.3)

$$\rho(\lambda)\,d\lambda = \int \sum_{\alpha} |\langle \lambda, \alpha|\psi\rangle|^2\, d\lambda\,; \qquad P_n = \sum_{\alpha} |\langle n, \alpha|\psi\rangle|^2, \tag{7.44}$$

respectively, as the *probability* of observing a value contained in the range $[\lambda, \lambda + d\lambda]$ for the observable $\mathcal{A}$ (ρ is the probability density) or as the probability that $\mathcal{A}$ takes the value λ_n. Let us note, in particular, that the total probability is given by

$$\int \rho(\lambda) d\lambda + \sum_n P_n = \langle\psi|\,\{\int d\mathcal{P}_c(\lambda) + \sum_n \mathcal{P}_n\,\}|\psi\rangle = \langle\psi|\psi\rangle = 1. \tag{7.45}$$

[4]Consider two state vectors $f + g$ and $f + ig$, then

$$\langle f+g|A|f+g\rangle =$$
$$\langle f|A|f\rangle + \langle g|A|g\rangle + 2\,\mathrm{Re}[\langle g|A|f\rangle]\,;$$
$$\langle f+i\,g|A|f+i\,g\rangle =$$
$$\langle f|A|f\rangle + \langle g|A|g\rangle + 2\,\mathrm{Im}[\langle g|A|f\rangle].$$

$\mathrm{Re}[\langle g|A|f\rangle]$ and $\mathrm{Im}[\langle g|A|f\rangle]$ are thus determined by the expectation values of A.

Expectation values of a self-adjoint operator

We have already seen that the expectation value of a self-adjoint operator on any state is real. A remarkable property of a self-adjoint operator is that it is possible to reconstruct the operator itself from its expectation values;[4] for instance, two operators having the same expectation values in all states are equal.

Theorem 7.3 *If two (maximally extended) self-adjoint operators A, B are such that*

$$\forall\,\psi \qquad \langle\psi|A|\psi\rangle = \langle\psi|B|\psi\rangle,$$

then $A = B$.

Commuting operators

We have already noted that, in order for two observables $\mathcal{A}, \mathcal{B}$ to be measurable simultaneously, the corresponding operators A, B must commute (see page 33). A useful theorem in this connection is the following

Theorem 7.4. (von Neumann's theorem) *Given two (maximally extended) self-adjoint and commuting operators A, B there exists an operator R of which A and B are functions, i.e., $A = F(R)$, $B = G(R)$.*

Let us give an idea of the proof by considering operators having only discrete spectra. As we have already seen, if two operators A, B commute, it is possible to choose a base of common eigenvectors

$$A|n\rangle = \lambda_n|n\rangle\,, \qquad B|n\rangle = \mu_n|n\rangle\,.$$

Let us consider now some sequence x_n and construct a self-adjoint operator R defined by

$$R = \sum_n x_n|n\rangle\langle n|\,, \qquad R|n\rangle = x_n|n\rangle\,.$$

One can always choose two functions F, G such that $F(x_n) = \lambda_n$, $G(x_n) = \mu_n$ hold for each n. Then by using the spectral decomposition of R,

$$F(R) = \sum_n F(x_n)\,|n\rangle\langle n| = \sum_n \lambda_n\,|n\rangle\langle n| = A$$
$$G(R) = \sum_n G(x_n)\,|n\rangle\langle n| = \sum_n \mu_n\,|n\rangle\langle n| = B\,.$$

7.4 Unitary transformations

Physical quantities in quantum mechanics are associated with the matrix elements of various operators

$$\langle\phi|O|\psi\rangle\,, \tag{7.46}$$

or, $O|\psi\rangle$ being a vector itself, to various scalar products

$$\langle f|g\rangle\,. \tag{7.47}$$

It is natural to ask for which sort of change of basis, $f \to Uf$, $\forall f$, does the physics remain the same. In other words, for which kind of operators U does the relation

$$\langle Uf|Ug\rangle = \langle f|g\rangle \tag{7.48}$$

hold for all f, g? A "change of basis" of this sort will leave all physical predictions of the theory invariant, though each state, operator, etc., is described differently.[5] If a transformation of this sort leaves the *form* of the Hamiltonian invariant, then one is dealing actually with a *symmetry* (see Section 5.1); the notion of allowed base transformations we are considering here, is, however, a more general one.

[5] To be precise, a transformation in which a relation $\langle Vf|Vg\rangle = (\langle f|g\rangle)^*$ holds for all states—known as anti-unitary transformations—is also allowed, as has been already noted in connection with the time reversal symmetry in Section 5.1.

As we wish to perform the same transformation on every vector of the system, it must be that $\mathcal{D}(U) = \mathscr{H}$; furthermore, as one is simply changing representation, the role of the original and the new one must be interchangeable: the image of U must also be $\mathscr{H}$. We thus arrive at the following

Definition An operator U with domain $\mathscr{H}$ and image $\mathscr{H}$ is said to be *unitary* if

$$\boxed{\forall\, x, y \in \mathscr{H} \;:\quad (Ux, Uy) = (x, y)\,.} \tag{7.49}$$

Note that no requirement of linearity has been made; we have only required the definability on the entire Hilbert space and that its action on $\mathscr{H}$ is surjective. One can show immediately that U has an inverse, that is, if $Uf = Ug$ then $f = g$:

$$\begin{aligned} 0 &= (Uf - Ug, Uf - Ug) = (Uf, Uf) - (Uf, Ug) - (Ug, Uf) + (Ug, Ug) \\ &= (f, f) - (f, g) - (g, f) + (g, g) = (f - g, f - g) \quad \Rightarrow \quad f = g \end{aligned}$$

Clearly U^{-1} is unitary also. Let us write $f = \alpha_1 x + \alpha_2 y$ and use the existence of the inverse,

$$\begin{aligned} &(g, Uf) = (U^{-1}g, f) = (U^{-1}g, \alpha_1 x + \alpha_2 y) = \alpha_1 (U^{-1}g, x) + \alpha_2 (U^{-1}g, y) \\ &= \alpha_1 (g, Ux) + \alpha_2 (g, Uy);\ g \,\text{arbitrary} \Rightarrow U(\alpha_1 x + \alpha_2 y) = \alpha_1 Ux + \alpha_2 Uy\,. \end{aligned}$$

Therefore a unitary operator is also *linear*. From the preceding relations it follows that $U^\dagger = U^{-1}$. From the relation $(Uf, Uf) = |Uf|^2 = |f|^2$ it follows that $||U|| = 1$ hence U is a linear bounded operator, and hence continuous.

Recapitulating,

$$\boxed{U^\dagger U = U U^\dagger = \mathbf{1}; \qquad U^\dagger = U^{-1}\,.} \tag{7.50}$$

The transformation law of the *operators* under a unitary transformation follows from eqn (7.50): let us rewrite eqn (7.46) twice inserting the identity operator $\mathbf{1} = U^\dagger U$:

$$\langle\phi|O|\psi\rangle = \langle\phi|U^\dagger U O U^\dagger U|\psi\rangle = \langle\tilde{\phi}|\tilde{O}|\tilde{\psi}\rangle, \tag{7.51}$$

where

$$\boxed{|\tilde{\psi}\rangle \equiv U|\psi\rangle; \quad |\tilde{\phi}\rangle \equiv U|\phi\rangle; \qquad \tilde{O} \equiv U\,O\,U^\dagger\,.} \tag{7.52}$$

Note that the norm of the states remains, as it should, invariant:

$$\langle\tilde{\psi}|\tilde{\psi}\rangle = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle.$$

The transformation of the state vectors and the operators given by eqns (7.51, 7.52) is known as a *unitary transformation*. As all the quantities dealt with in quantum mechanics reduce to some combinations of the matrix elements of type (7.47), it follows that the theory is invariant under unitary transformations. But this means that

the states and operators in quantum mechanics are defined up to unitary transformations.

For the eigenvalues of the unitary operator U we have the following obvious properties:

- The eigenvalues of a unitary operator have a norm 1, $|\lambda| = 1$.
- The eigenvectors relative to two distinct eigenvalues are orthogonal.

In fact, $\langle\psi|\psi\rangle = \langle U\psi|U\psi\rangle = |\lambda|^2\langle\psi|\psi\rangle$. For two eigenvectors belonging to different eigenvalues,

$$\langle\psi_1|\psi_2\rangle = \langle U\psi_1|U\psi_2\rangle = \lambda_1^*\lambda_2\langle\psi_1|\psi_2\rangle.$$

As $\lambda_1^*\lambda_2 \neq 1$ (as $\lambda_1 \neq \lambda_2$ and because they have the absolute value 1), it follows that $\langle\psi_1|\psi_2\rangle = 0$.

Also for the unitary operators it is possible to write a spectral representation, in a form analogous to eqn (7.39):

$$\mathbf{1} = \int_0^{2\pi} d\mathcal{P}(\lambda) = \int_0^{2\pi} d\mathcal{P}_c(\lambda) + \sum_n \mathcal{P}_n; \tag{7.53a}$$

$$|\psi\rangle = \int_0^{2\pi} d\mathcal{P}(\lambda)\,|\psi\rangle = \int_0^{2\pi} d\mathcal{P}_c(\lambda)\,|\psi\rangle + \sum_n \mathcal{P}_n\,|\psi\rangle, \quad \forall\psi \in \mathscr{H}; \tag{7.53b}$$

$$U = \int_0^{2\pi} e^{i\lambda}\,d\mathcal{P}(\lambda) = \int_0^{2\pi} e^{i\lambda}\,d\mathcal{P}_c(\lambda) + \sum_n e^{i\lambda_n}\,\mathcal{P}_n. \tag{7.53c}$$

Remark. It is possible to generalize eqn (7.48)—and it is sometimes useful to do so—to the cases in which U is an operator transforming the states of a Hilbert space to those in another Hilbert space, $U : \mathscr{H} \to \mathscr{H}'$; in these cases, the transformations are often termed *isometries*.

7.5 The Heisenberg picture

A significant result in classical mechanics (see Supplement 20.1) is that the time evolution $q(t), p(t) \to q(t+dt), p(t+dt)$ is a succession of infinitesimal canonical transformations, with the Hamiltonian playing the role of their generator. Analogously, in quantum mechanics, the time evolution of the system is described as a unitary transformation,

$$|\psi(t)\rangle = e^{-iHt/\hbar}|\psi(0)\rangle, \tag{7.54}$$

with the Hamiltonian operator playing the role of infinitesimal generator of time evolution. Indeed, eqn (7.54) is the formal solution of the Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle, \quad |\psi(t)\rangle|_{t=0} = |\psi(0)\rangle\,,$$

and as H is self-adjoint (which is always assumed to be the case) the operator $\exp(-iHt/\hbar)$ is unitary (see Theorem 7.2 on page 162).

The discussion of the previous section allows us to study the time evolution in quantum mechanics from a new point of view. Let us consider a particular, time-dependent unitary transformation

$$U(t) = e^{iHt/\hbar}\,.$$

The states and generic operators of the system transform according to eqn (7.52):

$$|\psi_H\rangle = U(t)|\psi(t)\rangle = e^{iHt/\hbar}|\psi(t)\rangle = |\psi(0)\rangle; \tag{7.55}$$

$$O_H(t) = U(t)\,O\,U(t)^\dagger = e^{iHt/\hbar}\,O\,e^{-iHt/\hbar}. \tag{7.56}$$

All the matrix elements are invariant under such transformations,

$$\langle\psi(t)|O|\psi(t)\rangle = \langle\psi_H|O_H(t)|\psi_H\rangle,$$

but now the time evolution of the systems is no longer described by the Schrödinger equation; it resides in a nontrivial time dependence of the operators!

The equation of motion of a generic operator O follows from eqn (7.56); it is given by[6]

$$i\hbar\frac{d\,O_H}{d\,t} = i\hbar\frac{\partial\,O_H}{\partial\,t} + [O_H, H], \tag{7.57}$$

where the first term is present if the operator explicitly depends on time. Equation (7.57) is known as the *Heisenberg equation*.

[6] Note the resemblance of this formula to the classical equation of motion expressed in terms of Poisson's brackets:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \{H, f\}\,.$$

The description of the time evolution in quantum mechanics in terms of eqns (7.55), (7.56), (7.57) is known as the **Heisenberg picture** or Heisenberg representation, as opposed to the **Schrödinger picture** or Schrödinger representation, in which the state vectors evolve with time. At an instant (e.g., at $t = 0$) the two pictures coincide:

$$O_H(0) = O; \quad |\psi_H\rangle = |\psi(0)\rangle.$$

It is of crucial importance that the fundamental commutators *at equal time* have the same form at any instant and independently of the Hamiltonian,

$$[q_{iH}(t), p_{jH}(t)] = i\,\hbar\,\delta_{ij}, \quad [q_{iH}(t), q_{jH}(t)] = [p_{iH}(t), p_{jH}(t)] = 0. \tag{7.58}$$

For instance,

$$\begin{aligned}[q_{iH}(t), p_{jH}(t)] &= [e^{i\,Ht/\hbar}\,q_i\,e^{-i\,Ht/\hbar}, e^{i\,Ht/\hbar}\,p_j\,e^{-i\,Ht/\hbar}] \\ &= e^{i\,Ht/\hbar}\,[q_i, p_j]\,e^{-i\,Ht/\hbar} = i\,\hbar\,\delta_{ij}.\end{aligned}$$

The usual commutators in the Schrödinger picture can, in fact, be seen as a particular case (for $t = 0$) of the more general result (7.58). The fact that the fundamental commutators have the same form at any time and hold independently of the particular Hamiltonian (details of dynamics) is essential for the consistency of the entire structure of quantum mechanics, as there is no privileged instant of time.[7]

Vice versa, the commutators at two different times

$$[q_{iH}(t), p_{jH}(t')], \quad [q_{iH}(t), q_{jH}(t')], \quad [p_{iH}(t), p_{jH}(t')]$$

contain dynamical information and depend on the system considered.

Let us note that each operator $f(q, p)$ in the Schrödinger picture becomes $f(q_H, p_H)$ in the Heisenberg picture. It is easy to verify this by expanding in powers of q and p, and inserting $U^{-1}U$ for each monomial, for instance:

$$\begin{aligned}U q^k p^n U^{-1} &= U q \ldots q\,p \ldots p\,U^{-1} = U q\,U^{-1} U q\,U^{-1} U \ldots U p\,U^{-1} \ldots \\ &= (U q\,U^{-1})^k (U p\,U^{-1})^n = q_H^k\,p_H^n.\end{aligned}$$

The problem of time evolution is thus reduced to solving the Heisenberg equations for q_H, p_H:[8]

$$\frac{d}{dt} q_H = -\frac{i}{\hbar}\,[q_H, H]\,; \qquad \frac{d}{dt} p_H = -\frac{i}{\hbar}\,[p_H, H]\,. \tag{7.59}$$

The time-dependent operators which are solutions of the above must satisfy constraint (7.58). The invariance of the latter during the time evolution guarantees that if the constraint is satisfied at the initial time, it will be satisfied at later times as well.

[7] We are here assuming the standard properties of the physical world, where time is homogeneous and the total energy conservation holds to a very high precision. Near the initial Big Bang singularity of the expansion of the Universe, when quantum gravity effects become important and spacetime highly curved, the whole concept of quantum mechanics itself might require some drastic revisions.

[8] **Exercise.** Solve the Heisenberg equation for a *free* particle of mass m moving in one dimension. Evaluate the commutator $[q_H(t), q_H(0)]$ at $t \neq 0$. (Answer: $[q_H(t), q_H(0)] = -i\,\hbar t/m$.)

7.5.1 The harmonic oscillator in the Heisenberg picture

Consider the linear oscillator

$$H = \frac{p^2}{2m} + \frac{m\,\omega^2}{2}\,x^2.$$

The Hamiltonian in the Heisenberg picture is

$$H_H = U\,H(\,x,\,p\,)\,U^\dagger = H(\,U\,x\,U^\dagger, U\,p\,U^\dagger) = \frac{p_H^2}{2\,m} + \frac{m\,\omega^2}{2}x_H^2.$$

The Heisenberg equations are (7.59)

$$m\,\dot{x}_H = p_H; \qquad \dot{p}_H = -m\,\omega^2\,x_H\,.$$

The solution has a form identical to that in classical mechanics,

$$x_H(t) = x\,\cos\omega\,t + \frac{1}{m\,\omega}\,p\,\sin\omega\,t\,; \tag{7.60}$$

$$p_H(t) = p\,\cos\omega\,t - m\,\omega\,x\,\sin\omega\,t\,. \tag{7.61}$$

In terms of the creation and annihilation operators (3.34), (3.35),

$$a = \sqrt{\frac{m\omega}{2\hbar}}x + i\sqrt{\frac{1}{2m\omega\hbar}}\,p = \sqrt{\frac{m\omega}{2\hbar}}\left(x + i\,\frac{p}{m\omega}\right),$$

$$a^\dagger = \sqrt{\frac{m\omega}{2\hbar}}x - i\sqrt{\frac{1}{2m\omega\hbar}}\,p = \sqrt{\frac{m\omega}{2\hbar}}\left(x - i\,\frac{p}{m\omega}\right),$$

one finds from eqn (7.61) that

$$a_H(t) = \sqrt{\frac{m\omega}{2\hbar}}\left[x\,\cos\omega t + \frac{p}{m\omega}\sin\omega t + i\frac{p}{m\omega}\cos\omega t - ix\sin\omega t\right] = a_H(0)\,e^{-i\omega t}$$

and, analogously,

$$a_H^\dagger(t) = a_H^\dagger(0)\,e^{i\omega t}\,.$$

Thus

$$a_H(t) = a\,e^{-i\omega t}; \quad a_H^\dagger(t) = a^\dagger\,e^{i\omega t} \Rightarrow \quad \dot{a}_H = -i\,\omega\,a_H; \quad \dot{a}_H^\dagger = +i\,\omega\,a_H\,.$$

These solutions can naturally be found directly by using the Heisenberg form of the Hamiltonian in terms of the creation and annihilation operators

$$H = \frac{\omega\hbar}{2}(aa^\dagger + a^\dagger a) = \omega\hbar\left(a^\dagger a + \frac{1}{2}\right)$$

and using the commutation relation

$$[a, a^\dagger] = 1\,.$$

As an example of the use of the Heisenberg picture, let us consider a linear oscillator, described by a *real* wave packet $\psi_0(x)$ at $t = 0$. Suppose

that the expectation values $\langle\psi_0|p^2|\psi_0\rangle \equiv p_0^2$ and $\langle\psi_0|x^2|\psi_0\rangle \equiv x_0^2$ are known, and we want to determine $\langle\psi(t)|p^2|\psi(t)\rangle$. In the Schrödinger picture one has to solve the Schrödinger equation to find $|\psi(t)\rangle$, and then compute the expectation value of p^2. In the Heisenberg picture the problem is easily solved by

$$\langle\psi(t)|p^2|\psi(t)\rangle = \langle\psi_0|\,U(t)\,p^2\,U^{-1}(t)\,|\psi_0\rangle = \langle\psi_0|p_H(t)^2|\psi_0\rangle.$$

But

$$p_H(t)^2 = p^2\,\cos^2\omega\,t + m^2\,\omega^2\,x^2\,\sin^2\omega\,t - m\,\omega\,(x\,p + p\,x)\,\cos\omega\,t\,\sin\omega\,t\,,$$

where use was made of eqn (7.61); furthermore note that

$$\langle\psi_0|\,x\,p + p\,x\,|\psi_0\rangle = 0$$

(the left-hand side must be real, being the expectation value of a Hermitian operator, but it is also purely imaginary for any real wave function). It follows then that

$$\langle\psi(t)|p^2|\psi(t)\rangle = p_0^2\,\cos^2\omega\,t + m^2\,\omega^2\,x_0^2\,\sin^2\omega\,t.$$

Analogously one finds that

$$\langle\psi(t)|x^2|\psi(t)\rangle = x_0^2\,\cos^2\omega\,t + \frac{1}{m^2\,\omega^2}\,p_0^2\,\sin^2\omega\,t.$$

For non-real initial wave functions (moving wave packets) the expectation value of the operator $x\,p + p\,x$ at $t = 0$ must also be taken into account.

7.6 The uncertainty principle

The connection between the mathematical structure of the abstract Hilbert space and physics is made upon the association, in the Schrödinger picture, between the dynamical variables q, p and the operators Q, P:

$$q \to Q, \quad Q\,\psi(x) = x\psi(x)\,; \qquad p \to P, \quad P\,\psi(x) = \frac{\hbar}{i}\frac{\partial}{\partial x}\,\psi(x)$$

it is the mathematical realization of de Broglie's hypothesis. This leads to

$$\boxed{[Q, P] = i\,\hbar\,,} \tag{7.62}$$

which has been stated as one of the postulate, P5.

It is invariant under unitary transformations: indeed if $\tilde{Q} = UQU^{-1}$, $\tilde{P} = UPU^{-1}$ one has

$$[\tilde{Q}, \tilde{P}] = UQU^{-1}UPU^{-1} - UPU^{-1}UQU^{-1} = U[Q, P]U^{-1} = i\,\hbar\,:$$

it is valid in the abstract Hilbert space.

A useful generalization of eqn (7.20) is

$$Pf(Q) - f(Q)P = -i\,\hbar f'(Q)\,.$$

In fact,

$$\begin{aligned} Pf(Q)\psi(x) &= -i\hbar\,\frac{\partial}{\partial x}\,(f(x)\,\psi(x)) = -i\hbar f'(x)\psi(x) + f(x)\,\psi'(x) = \\ &= -i\,\hbar f'(Q)\psi(x) + f(Q)(P\psi(x))\,. \end{aligned}$$

Let us now come back to the uncertainty relation. Let $\mathcal{A}$ be an observable, associated with the operator A. It will have an expectation value $\overline{A}_\psi = \langle\psi|A|\psi\rangle$ in a state ψ. Define the dispersion by

$$(\Delta A)^2 = \langle\psi|A^2|\psi\rangle - (\langle\psi|A|\psi\rangle)^2 \equiv \langle\psi|(A - \overline{A}_\psi)^2|\psi\rangle\,.$$

Consider now, two non-commuting self-adjoint operators A, B:

$$[A, B] = AB - BA = i\,\hbar\,C, \tag{7.63}$$

C is Hermitian. In each state in which the expectation value of the both sides of eqn (7.63) can be defined one finds the uncertainty relation

$$\Delta A \cdot \Delta B \geq \frac{\hbar}{2}\,|\langle\psi|C|\psi\rangle|\,. \tag{7.64}$$

The proof is the repetition of the one given in Section 2.3.4. Define operators

$$\tilde{A} = A - \overline{A}_\psi\,; \quad \tilde{B} = B - \overline{B}_\psi\,; \quad [\tilde{A}, \tilde{B}] = i\,\hbar\,C\,; \qquad O = \tilde{A} + i\,\alpha\tilde{B}.$$

It follows from the positivity $\langle O\psi|O\psi\rangle \geq 0$ that

$$\begin{aligned} 0 \leq \langle O\psi|O\psi\rangle &= \langle\tilde{A}\psi|\tilde{A}\psi\rangle + \alpha^2\langle\tilde{B}\psi|\tilde{B}\psi\rangle + i\,\alpha\,\langle\psi|(\tilde{A}\tilde{B} - \tilde{B}\tilde{A})|\psi\rangle = \\ &= \langle\tilde{A}\psi|\tilde{A}\psi\rangle + \alpha^2\langle\tilde{B}\psi|\tilde{B}\psi\rangle - \alpha\,\hbar\psi C|\psi\rangle\,. \end{aligned}$$

Imposing the condition that the discriminant of the quadratic form in α be negative, one finds eqn (7.64).

In particular, by applying the above to the canonical commutation relation we recover the Heisenberg's relation

$$\Delta P \cdot \Delta Q \geq \frac{\hbar}{2}$$

as discussed earlier.

7.7 Mixed states and the density matrix

The description of a physical state of a given system in terms of a state vector (wave function) in a Hilbert space, is the most detailed description in quantum mechanics.[9] There are many situations (strictly speaking, practically always), however, in which such a "complete" description is

[9] The idea that the probabilistic interpretation of quantum mechanics is to be attributed to an unknown distribution of certain inaccessible but classical variables goes under the generic name of "hidden-variable theories". Einstein famously argued for the incompleteness of quantum mechanics, in favor of these alternative theories. That no such theory can actually reproduce fully the content of quantum mechanics, hence can be experimentally distinguished from the latter (and indeed discarded), was shown in a milestone paper by J.S. Bell in 1960. These questions will be discussed in Chapters 18 and 19.

either impossible or unnecessary. It suffices to consider the case in which we deal with a subsystem of a larger, closed system: having access only to variables belonging to the subsystem, there is no hope of describing it in terms of a wave function. Another important class of cases are systems of many degrees of freedom (macroscopic systems, solid, gas, etc.). In these cases it is obviously impossible to have a complete knowledge of the wave function of, e.g., 10^{23} molecules: we must work with average quantities, defined in various ways. Still other cases concern either unpolarized or partially polarized beams of particles in a scattering problem. Again an incomplete knowledge of the spin state of the particles in the beam makes the description in terms of a wave function unavailable.

In all these cases we are dealing with **mixed states**; in contrast, states described by wave functions are **pure states**. In the case of a mixed state the role of the wave function is taken by a **density matrix**. A mixed state is characterized by various degrees of incomplete (or lack of) information.

Consider for concreteness the case of the first type: a subsystem S of a larger, closed system Σ. We have access, by assumption, to the variables ($\{x\}$) in S only. Even though the total system Σ can have a wave function $\psi(q,x)$, it is not in general factorized:

$$\Psi(q,x) \neq \psi_S(x)\psi_{\Sigma/S}(q);$$

the subsystem in itself does not have a wave function. How can one calculate the expectation value of an operator $\hat{f}_x$ which acts only on the variables $\{x\}$ of the subsystem?

By denoting an arbitrarily chosen orthonormal basis for each system S and Σ/S as $|j\rangle$ and $|\alpha\rangle$, respectively, a generic state is described by

$$|\Psi\rangle = \sum_{j,\alpha} c_{j,\alpha}\,|j\rangle \otimes |\alpha\rangle\ . \tag{7.65}$$

The expectation value of $\hat{f}$ in this state is

$$\langle\Psi|\hat{f}|\Psi\rangle = \sum_{j,k}\sum_{\alpha} c_{k,\alpha}\,c^*_{j,\alpha}\,\langle j|\hat{f}|k\rangle = \mathrm{Tr}\,(\mathbf{f}\rho)\ ,$$

where

$$\rho_{jk} \equiv \sum_{\alpha} c_{j,\alpha}\,c^*_{k,\alpha}\ , \qquad \mathbf{f}_{jk} = \langle j|\hat{f}|k\rangle\ .$$

The *density matrix* ρ has the following general properties:

$$\mathrm{Tr}\,\rho = 1\ ; \tag{7.66a}$$

$$\rho^\dagger = \rho\ ; \quad \text{(Hermiticity)} \tag{7.66b}$$

$$0 \leq \rho_{jj} \leq 1\ ; \tag{7.66c}$$

$$|\rho_{jk}|^2 \leq \rho_{jj}\,\rho_{kk}\ . \tag{7.66d}$$

Equations (7.66a)–(7.66c) are obvious. The last one can be shown di-

rectly:

$$\begin{aligned}\rho_{jj}\,\rho_{kk} &- \rho_{jk}\,\rho_{kj} \\ &= \sum_{\alpha,\beta}[\,c_{j,\alpha}\,c^*_{j,\alpha}\,c_{k,\beta}\,c^*_{k,\beta} - c_{j,\alpha}\,c^*_{k,\alpha}\,c_{k,\beta}\,c^*_{j,\beta}\,] \\ &= \frac{1}{2}\sum_{\alpha,\beta}[\,c_{j,\alpha}\,c_{k,\beta} - c_{k,\alpha}\,c_{j,\beta}\,][\,c_{j,\alpha}\,c_{k,\beta} - c_{k,\alpha}\,c_{j,\beta}\,]^* \;\geq\quad 0\,.\end{aligned}$$

In the case of a pure state, with the wave function (i.e., no sum over α)

$$|\psi\rangle = \sum_j c_j\,|j\rangle,$$

the density matrix is simply

$$\rho_{jk} = c_j c^*_k.$$

The density matrix of a pure state thus satisfies the characteristic property

$$\rho^2 = \rho\,,$$

as can be proven easily by using $\sum_\ell c^*_\ell c_\ell = 1$.

Clearly, the concept of a mixed state is a more general one than that of a pure state. A pure state can always be regarded as a special type of mixed state, but not all mixed states are pure.

An important class of applications of the density matrix formalism concerns statistical physics. There the enormous number of the degrees of freedom forces us to appeal to a statistical treatment (Boltzmann). The density matrix $\rho_{jk} = w_{jk}$ is known as a *statistical matrix* in these cases. Let W_n be the probability that one of the microscopic systems (for example, an atom) is in the n-th energy eigenstate

$$|\psi^{(n)}(t)\rangle = \sum_j a^n_j(t)|\psi_j\rangle,$$

where $\{\psi_j\}$ is a generic orthonormal basis (e.g., eigenstates of some operator). For instance, in a canonical ensemble at temperature T the energy distribution is given by Boltzmann's formula

$$W_n = e^{-E_n/kT}/\mathcal{N}, \quad \sum_n W_n = 1,$$

where $\mathcal{N}$ is the partition function $\mathcal{N} = \sum_n e^{-E_n/kT}$. The discussion below is, however, valid in any type of statistical ensemble.

The expectation value of an operator f is thus given by

$$\langle f\rangle = \sum_n W_n\,\langle\psi^{(n)}|f|\psi^{(n)}\rangle = \sum_n\sum_{j,k} W_n\,a^{(n)*}_j a^{(n)}_k f_{jk} = \mathrm{Tr}\,(\rho\,\mathbf{f}),$$

where we have introduced the density (or statistical) matrix

$$\rho_{jk} = \sum_n W_n\,a^{(n)*}_k\,a^{(n)}_j. \tag{7.67}$$

Note that, thanks to the positivity of the classical probability $W_n \geq 0$, the density matrix defined here satisfies the same defining properties (7.66a)–(7.66d) introduced before. In both cases the density matrix reflects our ignorance about the system.

The time evolution of the density matrix follows from the fact that $|\psi^{(i)}(t)\rangle$ obeys the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\psi^{(i)}(t)\rangle = H|\psi^{(i)}(t)\rangle.$$

As

$$a_n^{(i)}(t) = \langle\psi_n|\psi^{(i)}(t)\rangle,$$

one has

$$i\hbar\,\dot{a}_n^{(i)}(t) = \langle\psi_n|H|\psi^{(i)}(t)\rangle = \sum_k a_k^{(i)} H_{nk}.$$

Analogously

$$-i\hbar\,\dot{a}_m^{(i)*}(t) = \langle\psi^{(i)}(t)|H|\psi_m\rangle = \sum_k a_k^{(i)*} H_{km}.$$

One thus finds that

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}\rho_{nm} &= \sum_i W_i \sum_k (a_m^{(i)*} H_{nk} a_k^{(i)} - a_k^{(i)*} H_{km} a_n^{(i)}) \\ &= \sum_k (H_{nk}\,\rho_{km} - \rho_{nk}\,H_{km}) = [\mathbf{H}, \rho]_{nm}. \end{aligned} \tag{7.68}$$

This equation substitutes, for a mixed state, the Schrödinger equation or Heisenberg equation for a pure state.[10]

[10] Equation (7.68) formally resembles the Heisenberg equation; note, however, the curious (but well-known) sign difference in the two equations.

7.7.1 Photon polarization

Let us illustrate the use of the density matrix, by taking the example of the polarization of a photon. As discussed in Section 2.1.2, the state of a photon, all other attributes such as the energy, momentum (the wavelength and the direction of propagation) being neglected, is described as a typical two-level system (or a *q-bit*), with the base states $|1\rangle$ and $|2\rangle$. $|1\rangle$ and $|2\rangle$ can be taken as the states of linear (and orthogonal) polarizations, as the two independent circular polarization states, and so on. A pure state is described by a wave function

$$|\psi\rangle = c_1|1\rangle + c_2|2\rangle \equiv \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \tag{7.69}$$

where

$$|1\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad |2\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \qquad \langle 1| = (1,\,0); \quad \langle 2| = (0,\,1), \tag{7.70}$$

c_1, c_2 are arbitrary complex numbers such that

$$|c_1|^2 + |c_2|^2 = 1.$$

The two base states are taken to be orthonormal:

$$\langle 1|1\rangle = \langle 2|2\rangle = 1; \quad \langle 1|2\rangle = \langle 2|1\rangle = 0.$$

The operators which correspond to the measurement of the photon polarizations 1 and 2 act as

$$P_1|1\rangle = |1\rangle; \quad P_1|2\rangle = 0; \qquad P_2|2\rangle = |2\rangle; \quad P_2|1\rangle = 0;$$

that is

$$P_1 = |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \quad P_2 = |2\rangle\langle 2| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} : \tag{7.71}$$

they are the projection operators on the states $|1\rangle$ and $|2\rangle$, respectively. The density matrix for the pure state (7.69) is simply

$$\rho = \begin{pmatrix} |c_1|^2 & c_1 c_2^* \\ c_1^* c_2 & |c_2|^2 \end{pmatrix}.$$

A beam of partially polarized or unpolarized light, is described as a mixed state. A photon in an unpolarized beam is described by the density matrix

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \tag{7.72}$$

in fact, the mean value of the polarization in directions 1 or 2 is

$$< P_1 >= \mathrm{Tr}\,(P_1\,\rho) = \frac{1}{2}; \quad < P_2 >= \mathrm{Tr}\,(P_2\,\rho) = \frac{1}{2}, \tag{7.73}$$

respectively. It is not difficult to check that the mean value of the polarization in *any* polarization, in state (7.72), is given by $\frac{1}{2}$.

The state of partial polarization has a general representation of the form

$$\rho = \frac{1}{2}\begin{pmatrix} 1+\xi_3 & \xi_1 - i\xi_2 \\ \xi_1 + i\xi_2 & 1-\xi_3 \end{pmatrix} = \frac{1}{2}(\,\mathbf{1} + \sigma_i\,\xi_i\,),$$

where

$$\xi_1^2 + \xi_2^2 + \xi_3^2 \le 1,$$

and

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli matrices. The three numbers ξ_1, ξ_2, ξ_3 (real) are known as the Stokes parameters. It can be easily seen that

$$\rho^2 = \rho,$$

if and only if

$$\boldsymbol{\xi}^2 \equiv \xi_1^2 + \xi_2^2 + \xi_3^2 = 1 :$$

in this case the system is pure. $1 - \boldsymbol{\xi}^2$ thus can be seen as a measure of our ignorance of the state of polarization. ξ_3 describes the degree of polarization in directions 1 or 2, for instance

$$\langle P_1\rangle = \mathrm{Tr}\,P_1\,\rho = \frac{1+\xi_3}{2} = \begin{cases} 1 & \text{if } \xi_3 = 1, \\ 0 & \text{if } \xi_3 = -1\,. \end{cases}$$

Analogously ξ_1 is the measure of the linear polarization in the direction wichh makes n angle $\pm\frac{\pi}{4}$ with 1 and 2, as can be verified by constructing the projection operator

$$P_1' = |1'\rangle\langle 1'|, \qquad P_2' = |2'\rangle\langle 2'|, \qquad |1'\rangle = \frac{|1\rangle + |2\rangle}{\sqrt{2}}, \quad |2'\rangle = \frac{|1\rangle - |2\rangle}{\sqrt{2}},$$

and evaluating $< P_1' >$, etc. Finally, ξ_2 corresponds to the measure of circular polarizations,

$$|+\rangle = \frac{1}{\sqrt{2}}(|1\rangle + i\,|2\rangle); \quad |-\rangle = \frac{1}{\sqrt{2}}(|1\rangle - i\,|2\rangle).$$

7.8 Quantization in general coordinates

The rule for the quantum Hamiltonian introduced in eqns (2.29)–(2.30)

$$\hat{H}(\{\hat{q}_i\}, \{\hat{p}_i\}; t) = H_{clas}(\{q_i\}, \{p_i\}; t)|_{q_i \to \hat{q}_i = q_i;\, p_i \to \hat{p}_i = -i\hbar\frac{\partial}{\partial q_i}} \qquad (7.74)$$

looks straightforward, but a little thought shows that it is far from being trivial. Let us discuss some of the questions involved.

In order not to mix up physically distinct problems, let us separate the following two questions:

(i) quantization of a "simple" class of systems, i.e., in a flat space, with the standard quadratic kinetic term, which has the form

$$\sum_i \sum_{\alpha=x,y,z} \frac{p_{i,\alpha}^2}{2m_i}$$

in cartesian coordinates, with the potential V depending only on $\{q_{i,\alpha}\}$, and

(ii) quantization of physical systems of more general type, such as particles moving in a curved space or in a topologically nontrivial space, for instance, with nontrivial periodicity (defects), or particles moving in a momentum-dependent potential (e.g., a charged particle moving in an external electromagnetic field $\mathbf{A}_i$ with minimal interactions).

Consider the first class of systems, to start with. Is the quantization procedure (7.74) valid in generic curvilinear coordinates? One could consider any generalized coordinates, construct the canonically conjugate momenta $p_i \equiv \frac{\partial L}{\partial \dot{q}_i}$, and then make the replacement

$$p_i \to -i\hbar\frac{\partial}{\partial q_i} \qquad (7.75)$$

in the Hamiltonian, expressed in terms of the coordinates (q_i, p_i). It turns out that this procedure in general leads to a quantum Hamiltonian operator different from what is obtained in cartesian coordinates. For

instance, in spherical coordinates $q_i = (r, \theta, \phi)$, the kinetic term is given by

$$L = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\right) ,$$

and the canonical momenta are

$$p_r = m\,\dot{r}; \qquad p_\theta = m\,r^2\,\dot{\theta}; \qquad p_\phi = m\,r^2\sin^2\theta\,\dot{\phi} .$$

The kinetic term of the Hamiltonian takes the form

$$H_{kin} = \frac{p_r^2}{2\,m} + \frac{p_\theta^2}{2\,m\,r^2} + \frac{p_\phi^2}{2\,m\,r^2\,\sin^2\theta} .$$

The replacement rule (7.75) yields an operator

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) .$$

It can be seen immediately that such a procedure misses the terms

$$-\frac{\hbar^2}{2m}\left(\frac{2}{r}\frac{\partial}{\partial r} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial}{\partial\theta}\right)$$

present in the standard quantum Hamiltonian, *written* in spherical coordinates, see eqns (4.1), (4.2).

Analogously, a free particle moving in a plane

$$H = \frac{1}{2\,m}(p_x^2 + p_y^2),$$

has the form

$$H = \frac{1}{2\,m}\left(p_r^2 + \frac{1}{r^2}p_\phi^2\right).$$

in polar coordinates. The substitution

$$p_r \to -i\hbar\frac{\partial}{\partial r}, \qquad p_\phi \to -i\hbar\frac{\partial}{\partial\phi}$$

in the second Hamiltonian would yield an operator which differs [Messiah (2000)] from the standard operator obtained in cartesian coordinates, by

$$-\frac{\hbar^2}{2\,m\,r}\frac{\partial}{\partial r} .$$

The problem can actually look even more serious if we note that in general curvilinear coordinates procedure (7.74) is not even well defined, due to the well-known "operator ordering problem". Indeed, if g_{ij} is the metric, the kinetic term in the Lagrangian has the form $L_{(kin)} = (m/2)g_{ij}(q)\,\dot{q}^i\,\dot{q}^j$. Furthermore, define g^{ij} as the inverse of the metric g_{ij} and $g \equiv \det g = \det g_{ij}$. In the passage from classical to quantum mechanics, however, there is a fatal arbitrariness in writing the kinetic term as

$$\frac{1}{2m}g^{ij}(q)\,p_i\,p_j, \quad \frac{1}{2m}\,p_i\,g^{ij}(q)\,p_j, \quad \frac{1}{2m}\,\sqrt{g}\,p_j(\sqrt{g}\,g^{jk}(q)\,p_k),$$

or something else. This would appear to irreparably compromise the uniqueness of the quantum Hamiltonian. Note that the requirement of Hermiticity is not sufficient to eliminate this ambiguity.

In systems of the first type (i) there is actually no ambiguity whatsoever. A unique quantum operator corresponds to the classical Hamiltonian. It is an empirical law that the correct quantum Hamiltonian is given by rule (7.74), when one *works in cartesian coordinates*.[11] The replacement rule (7.75) must be applied on the classical Hamiltonian, written in cartesian coordinates and similarly for systems with more than one particle. As long as cartesian coordinates are used, the result does not depend on the particular choice of coordinates.

[11] It would thus appear that the basic quantum law privileges particular sets of coordinates. Its deep meaning, if any, is not known.

This does not preclude of course that, once the quantum Hamiltonian is correctly identified, *any* other coordinate system can be used by a *simple change of variables*. The general rule of quantization at the end of this section ultimately reduces to this prescription. (See also the books [Brillouin (1938)], [Kemble (1937)].)

The situation concerning systems of more general types (ii) is subtler. In the cases of minimal coupling, describing the interactions of a charged particle with an electromagnetic potential (for instance), the procedure

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = m\,\dot{q}_i + \frac{e}{c}A_i(q), \qquad p_i \to \hat{p}_i = -i\,\hbar\,\frac{\partial}{\partial q_i}$$

(in cartesian coordinates) gives the correct result: eqns (14.8) and (14.12).

In general, the requirement that the quantum Hamiltonian be self-adjoint (though necessary) is not sufficient to determine uniquely the quantum Hamiltonian. In the case of a particle moving on a circle (or equivalently, a particle moving on a periodic lattice), it is possible to define an infinite number of distinct quantum systems.

An important point, which sometimes risks being overlooked behind the mathematical armor of these considerations, is the fact that in certain systems physical parameters appear that characterize each quantum system. Even though the physical significance of these parameters is often quite clear, the *way* they characterize the quantum systems is quite remarkable, and unfamiliar from the classical point of view. For instance the θ parameter which characterizes the particle moving on a circle (Subsection 3.2.2) might be related to the magnetic flux flowing through the area encircled (in the physical example of a toroidal superconductor), or to the lattice momentum in the case of a particle moving through a periodic potential (see Subsection 3.6.1).

In the case of a particle moving in a general curved space, the classical Hamiltonian

$$H = \frac{1}{2m}g^{ij}(q)\,p_i\,p_j$$

corresponds to the quantum Hamiltonian [DeWitt (1957)]

$$\hat{H} = \frac{1}{2m}g^{-1/4}\hat{p}_i\,g^{1/2}\,g^{ij}\,\hat{p}_j\,g^{-1/4} + C\,\hbar^2\,R,$$

$$g^{1/4}\,\hat{p}_j\,g^{-1/4} = -i\hbar\frac{\partial}{\partial q^j},$$

where the constant C in front of the term of order $\hbar^2$ and proportional to the scalar curvature R of the space—and only this term—remains undetermined. On the other hand, it is quite understandable that, given the classical system (for which $\hbar = 0$ hence C has no meaning), there can be no *a priori* reason to privilege one particular value or another for C. This arbitrariness is known as DeWitt's ambiguity. In spite of the name, it must be regarded as signaling the presence of a physical parameter which characterizes each quantum system of this kind (which must be determined from experiments), and *not* as a sort of inconsistency or as a genuine ambiguity.

Let us come back now to particles moving in flat space ($R = 0$), a system of the first kind. From what has been just said it follows that the quantization procedure in a general curvilinear coordinate system is given by

$$\boxed{H \to \hat{H} = \frac{1}{2m}\, g^{-1/4} \hat{p}_i\, g^{1/2}\, g^{ij}\, \hat{p}_j\, g^{-1/4} + V(q)\,,} \tag{7.76}$$

$$\boxed{\hat{p}_i = g^{-1/4} \cdot \left(-i\hbar \frac{\partial}{\partial q^i}\right) \cdot g^{1/4}\,.} \tag{7.77}$$

Rule (7.77) can be rewritten as a sort of covariant derivative. The correctness of this procedure is guaranteed by the fact that the differential operator which follows from eqns (7.76), (7.77) is precisely the Laplace–Beltrami operator

$$\frac{1}{\sqrt{g}} \frac{\partial}{\partial q^i} \left(\sqrt{g}\, g^{ij}\, \frac{\partial}{\partial q^j}\right),$$

which is *invariant* under general coordinate transformations, and by the fact that in cartesian coordinate it reduces to the empirically correct prescription, (7.74). Thus in general coordinates neither the substitution rule $p_i \to \hat{p}_i = -i\hbar \frac{\partial}{\partial q_i}$ nor the canonical commutation rule $[q_i, p_j] = i\hbar\, \delta_{ij}$ are valid. These must be replaced by (7.77) and by the commutation relations which follow from it.

Further reading

The reader interested in the mathematical aspects of the theory can read the book [von Neumann (1932)]. A proof of Stone's theorem and von Neumann's theorems can be found in Vol. 1 of the book [Reed and Simon (1980a)], Chapter 8. A pedagogical introduction to self-adjoint extensions is in [Bonneau, Faraut, and Valent (2001)].

Guide to the Supplements

One of the main postulates of quantum mechanics is that each physical state corresponds to a ray in the Hilbert space (P1) (see Section 7.1). Does the inverse statement, "each vector in the Hilbert space describes a physical state", hold true as well? The question puts the superposition principle under scrutiny: given two physical states $|\alpha\rangle$, $|\beta\rangle$, is the state $|\alpha\rangle + |\beta\rangle$ also a physical state, for *whatever* choice of $|\alpha\rangle$ and $|\beta\rangle$? Supplement 20.14 is dedicated to the discussion of this subtle issue. The conclusion will be that the superposition principle actually admits exceptions. A well-known example is related to the exactly conserved electric charge. Only superpositions among the states with the same electric charge are allowed. An analogous restriction holds for the fermion number. Superposition of states with different fermion numbers is unphysical. These restrictions are known as *superselection rules*.

In a second Supplement some details of the von Neumann theorem are given. The theorem basically ensures the uniqueness of the Schrödinger representation for the Heisenberg commutation relations. Related questions on the relevance of canonical transformations in quantum mechanics and the problem of self-adjoint extensions of the operators are briefly addressed.

Problems

(7.1) Solve the Heisenberg equations of motion for a free particle and for a particle in an external uniform field.

(7.2) Solve the Heisenberg equations of motion for a harmonic oscillator in a uniform constant external field. Generalize the solution for a uniform time dependent force $F(t)$.

(7.3) Suppose that the system described by the wave function $\psi_S(x)$ at the instant $t = 0$ is an eigenstate of the operator f, with eigenvalue, f_0. Show that the wave function at time t is an eigenstate of the Heisenberg operator $\hat{f}_H(-t)$, with the same eigenvalue. (This technique is used in the book [Kogan and Galitsky (1963)] to compute the Green functions of several simple models).