2 Introduction to quantum mechanics

I ain't no physicist but I know what matters.
– Popeye the Sailor

Quantum mechanics: Real Black Magic Calculus
– Albert Einstein

Quantum mechanics is the most accurate and complete description of the world known. It is also the basis for an understanding of quantum computation and quantum information. This chapter provides all the necessary background knowledge of quantum mechanics needed for a thorough grasp of quantum computation and quantum information. No prior knowledge of quantum mechanics is assumed.

Quantum mechanics is easy to learn, despite its reputation as a difficult subject. The reputation comes from the difficulty of some *applications*, like understanding the structure of complicated molecules, which aren't fundamental to a grasp of the subject; we won't be discussing such applications. The only prerequisite for understanding is some familiarity with elementary linear algebra. Provided you have this background you can begin working out simple problems in a few hours, even with no prior knowledge of the subject.

Readers already familiar with quantum mechanics can quickly skim through this chapter, to become familiar with our (mostly standard) notational conventions, and to assure themselves of familiarity with all the material. Readers with little or no prior knowledge should work through the chapter in detail, pausing to attempt the exercises. If you have difficulty with an exercise, move on, and return later to make another attempt.

The chapter begins with a review of some material from linear algebra in Section 2.1. This section assumes familiarity with elementary linear algebra, but introduces the notation used by physicists to describe quantum mechanics, which is different to that used in most introductions to linear algebra. Section 2.2 describes the basic postulates of quantum mechanics. Upon completion of the section, you will have understood *all* of the fundamental principles of quantum mechanics. This section contains numerous simple exercises designed to help consolidate your grasp of this material. The remaining sections of the chapter, and of this book, elucidate upon this material, without introducing fundamentally new physical principles. Section 2.3 explains *superdense coding*, a surprising and illuminating example of quantum information processing which combines many of the postulates of quantum mechanics in a simple setting. Sections 2.4 and 2.5 develop powerful mathematical tools – the *density operator*, *purifications*, and the *Schmidt decomposition* – which are especially useful in the study of quantum computation and quantum information. Understanding these tools will also help you consolidate your understanding of elementary quantum mechanics. Finally, Section 2.6 examines the question of how quantum mechanics goes beyond the usual 'classical' understanding of the way the world works.

2.1 Linear algebra

This book is written as much to disturb and annoy as to instruct.
– The first line of *About Vectors*, by Banesh Hoffmann.

Life is complex – it has both real and imaginary parts.
– Anonymous

Linear algebra is the study of vector spaces and of linear operations on those vector spaces. A good understanding of quantum mechanics is based upon a solid grasp of elementary linear algebra. In this section we review some basic concepts from linear algebra, and describe the standard notations which are used for these concepts in the study of quantum mechanics. These notations are summarized in Figure 2.1 on page 62, with the quantum notation in the left column, and the linear-algebraic description in the right column. You may like to glance at the table, and see how many of the concepts in the right column you recognize.

In our opinion the chief obstacle to assimilation of the postulates of quantum mechanics is not the postulates themselves, but rather the large body of linear algebraic notions required to understand them. Coupled with the unusual Dirac notation adopted by physicists for quantum mechanics, it can appear (falsely) quite fearsome. For these reasons, we advise the reader not familiar with quantum mechanics to quickly read through the material which follows, pausing mainly to concentrate on understanding the absolute basics of the notation being used. Then proceed to a careful study of the main topic of the chapter – the postulates of quantum mechanics – returning to study the necessary linear algebraic notions and notations in more depth, as required.

The basic objects of linear algebra are *vector spaces*. The vector space of most interest to us is $\mathbf{C}^n$, the space of all n-tuples of complex numbers, $(z_1, \ldots, z_n)$. The elements of a vector space are called *vectors*, and we will sometimes use the column matrix notation

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \tag{2.1}$$

to indicate a vector. There is an *addition* operation defined which takes pairs of vectors to other vectors. In $\mathbf{C}^n$ the addition operation for vectors is defined by

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} + \begin{bmatrix} z'_1 \\ \vdots \\ z'_n \end{bmatrix} \equiv \begin{bmatrix} z_1 + z'_1 \\ \vdots \\ z_n + z'_n \end{bmatrix}, \tag{2.2}$$

where the addition operations on the right are just ordinary additions of complex numbers. Furthermore, in a vector space there is a *multiplication by a scalar* operation. In $\mathbf{C}^n$ this operation is defined by

$$z \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \equiv \begin{bmatrix} z z_1 \\ \vdots \\ z z_n \end{bmatrix}, \tag{2.3}$$

where z is a *scalar*, that is, a complex number, and the multiplications on the right are ordinary multiplication of complex numbers. Physicists sometimes refer to complex numbers as *c-numbers*.

Quantum mechanics is our main motivation for studying linear algebra, so we will use the standard notation of quantum mechanics for linear algebraic concepts. The standard quantum mechanical notation for a vector in a vector space is the following:

$$|\psi\rangle. \tag{2.4}$$

ψ is a label for the vector (any label is valid, although we prefer to use simple labels like ψ and φ). The $|\cdot\rangle$ notation is used to indicate that the object is a vector. The entire object $|\psi\rangle$ is sometimes called a *ket*, although we won't use that terminology often.

A vector space also contains a special *zero vector*, which we denote by 0. It satisfies the property that for any other vector $|v\rangle$, $|v\rangle + 0 = |v\rangle$. Note that we do not use the ket notation for the zero vector – it is the only exception we shall make. The reason for making the exception is because it is conventional to use the 'obvious' notation for the zero vector, $|0\rangle$, to mean something else entirely. The scalar multiplication operation is such that $z0 = 0$ for any complex number z. For convenience, we use the notation $(z_1, \ldots, z_n)$ to denote a column matrix with entries $z_1, \ldots, z_n$. In $\mathbf{C}^n$ the zero element is $(0, 0, \ldots, 0)$. A *vector subspace* of a vector space V is a subset W of V such that W is also a vector space, that is, W must be closed under scalar multiplication and addition.

Notation	Description
z^*	Complex conjugate of the complex number z. $(1 + i)^* = 1 - i$
$\|\psi\rangle$	Vector. Also known as a *ket*.
$\langle\psi\|$	Vector dual to $\|\psi\rangle$. Also known as a *bra*.
$\langle\varphi\|\psi\rangle$	Inner product between the vectors $\|\varphi\rangle$ and $\|\psi\rangle$.
$\|\varphi\rangle \otimes \|\psi\rangle$	Tensor product of $\|\varphi\rangle$ and $\|\psi\rangle$.
$\|\varphi\rangle\|\psi\rangle$	Abbreviated notation for tensor product of $\|\varphi\rangle$ and $\|\psi\rangle$.
A^*	Complex conjugate of the A matrix.
A^T	Transpose of the A matrix.
$A^\dagger$	Hermitian conjugate or adjoint of the A matrix, $A^\dagger = (A^T)^*$. $\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]^\dagger = \left[\begin{array}{cc} a^* & c^* \\ b^* & d^* \end{array}\right]$.
$\langle\varphi\|A\|\psi\rangle$	Inner product between $\|\varphi\rangle$ and $A\|\psi\rangle$. Equivalently, inner product between $A^\dagger\|\varphi\rangle$ and $\|\psi\rangle$.

Figure 2.1. Summary of some standard quantum mechanical notation for notions from linear algebra. This style of notation is known as the *Dirac* notation.

2.1.1 Bases and linear independence

A *spanning set* for a vector space is a set of vectors $|v_1\rangle, \ldots, |v_n\rangle$ such that any vector $|v\rangle$ in the vector space can be written as a linear combination $|v\rangle = \sum_i a_i |v_i\rangle$ of vectors

in that set. For example, a spanning set for the vector space $\mathbf{C}^2$ is the set

$$|v_1\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \quad |v_2\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{2.5}$$

since any vector

$$|v\rangle = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{2.6}$$

in $\mathbf{C}^2$ can be written as a linear combination $|v\rangle = a_1|v_1\rangle + a_2|v_2\rangle$ of the vectors $|v_1\rangle$ and $|v_2\rangle$. We say that the vectors $|v_1\rangle$ and $|v_2\rangle$ *span* the vector space $\mathbf{C}^2$.

Generally, a vector space may have many different spanning sets. A second spanning set for the vector space $\mathbf{C}^2$ is the set

$$|v_1\rangle \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad |v_2\rangle \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \tag{2.7}$$

since an arbitrary vector $|v\rangle = (a_1, a_2)$ can be written as a linear combination of $|v_1\rangle$ and $|v_2\rangle$,

$$|v\rangle = \frac{a_1 + a_2}{\sqrt{2}}|v_1\rangle + \frac{a_1 - a_2}{\sqrt{2}}|v_2\rangle. \tag{2.8}$$

A set of non-zero vectors $|v_1\rangle, \ldots, |v_n\rangle$ are *linearly dependent* if there exists a set of complex numbers $a_1, \ldots, a_n$ with $a_i \neq 0$ for at least one value of i, such that

$$a_1|v_1\rangle + a_2|v_2\rangle + \cdots + a_n|v_n\rangle = 0. \tag{2.9}$$

A set of vectors is *linearly independent* if it is not linearly dependent. It can be shown that any two sets of linearly independent vectors which span a vector space V contain the same number of elements. We call such a set a *basis* for V. Furthermore, such a basis set always exists. The number of elements in the basis is defined to be the *dimension* of V. In this book we will only be interested in *finite dimensional* vector spaces. There are many interesting and often difficult questions associated with infinite dimensional vector spaces. We won't need to worry about these questions.

Exercise 2.1: (Linear dependence: example) Show that $(1, -1), (1, 2)$ and $(2, 1)$ are linearly dependent.

2.1.2 Linear operators and matrices

A *linear operator* between vector spaces V and W is defined to be any function $A : V \rightarrow W$ which is linear in its inputs,

$$A\left(\sum_i a_i|v_i\rangle\right) = \sum_i a_i A\left(|v_i\rangle\right). \tag{2.10}$$

Usually we just write $A|v\rangle$ to denote $A(|v\rangle)$. When we say that a linear operator A is defined *on* a vector space, V, we mean that A is a linear operator from V to V. An important linear operator on any vector space V is the *identity operator*, I_V, defined by the equation $I_V|v\rangle \equiv |v\rangle$ for all vectors $|v\rangle$. Where no chance of confusion arises we drop the subscript V and just write I to denote the identity operator. Another important linear operator is the *zero operator*, which we denote 0. The zero operator maps all vectors to

the zero vector, $0|v\rangle \equiv 0$. It is clear from (2.10) that once the action of a linear operator A on a basis is specified, the action of A is completely determined on all inputs.

Suppose V, W, and X are vector spaces, and $A : V \rightarrow W$ and $B : W \rightarrow X$ are linear operators. Then we use the notation BA to denote the *composition* of B with A, defined by $(BA)(|v\rangle) \equiv B(A(|v\rangle))$. Once again, we write $BA|v\rangle$ as an abbreviation for $(BA)(|v\rangle)$.

The most convenient way to understand linear operators is in terms of their *matrix representations*. In fact, the linear operator and matrix viewpoints turn out to be completely equivalent. The matrix viewpoint may be more familiar to you, however. To see the connection, it helps to first understand that an m by n complex matrix A with entries A_{ij} is in fact a linear operator sending vectors in the vector space $\mathbf{C}^n$ to the vector space $\mathbf{C}^m$, under matrix multiplication of the matrix A by a vector in $\mathbf{C}^n$. More precisely, the claim that the matrix A is a linear operator just means that

$$A\left(\sum_i a_i|v_i\rangle\right) = \sum_i a_i A|v_i\rangle \tag{2.11}$$

is true as an equation where the operation is matrix multiplication of A by column vectors. Clearly, this is true!

We've seen that matrices can be regarded as linear operators. Can linear operators be given a matrix representation? In fact they can, as we now explain. This equivalence between the two viewpoints justifies our interchanging terms from matrix theory and operator theory throughout the book. Suppose $A : V \rightarrow W$ is a linear operator between vector spaces V and W. Suppose $|v_1\rangle, \ldots, |v_m\rangle$ is a basis for V and $|w_1\rangle, \ldots, |w_n\rangle$ is a basis for W. Then for each j in the range $1, \ldots, m$, there exist complex numbers A_{1j} through A_{nj} such that

$$A|v_j\rangle = \sum_i A_{ij}|w_i\rangle. \tag{2.12}$$

The matrix whose entries are the values A_{ij} is said to form a *matrix representation* of the operator A. This matrix representation of A is completely equivalent to the operator A, and we will use the matrix representation and abstract operator viewpoints interchangeably. Note that to make the connection between matrices and linear operators we must specify a set of input and output basis states for the input and output vector spaces of the linear operator.

Exercise 2.2: (Matrix representations: example) Suppose V is a vector space with basis vectors $|0\rangle$ and $|1\rangle$, and A is a linear operator from V to V such that $A|0\rangle = |1\rangle$ and $A|1\rangle = |0\rangle$. Give a matrix representation for A, with respect to the input basis $|0\rangle, |1\rangle$, and the output basis $|0\rangle, |1\rangle$. Find input and output bases which give rise to a different matrix representation of A.

Exercise 2.3: (Matrix representation for operator products) Suppose A is a linear operator from vector space V to vector space W, and B is a linear operator from vector space W to vector space X. Let $|v_i\rangle, |w_j\rangle$, and $|x_k\rangle$ be bases for the vector spaces V, W, and X, respectively. Show that the matrix representation for the linear transformation BA is the matrix product of the matrix representations for B and A, with respect to the appropriate bases.

Exercise 2.4: (**Matrix representation for identity**) Show that the identity operator on a vector space V has a matrix representation which is one along the diagonal and zero everywhere else, if the matrix representation is taken with respect to the same input and output bases. This matrix is known as the *identity matrix*.

2.1.3 The Pauli matrices

Four extremely useful matrices which we shall often have occasion to use are the *Pauli matrices*. These are 2 by 2 matrices, which go by a variety of notations. The matrices, and their corresponding notations, are depicted in Figure 2.2. The Pauli matrices are so useful in the study of quantum computation and quantum information that we encourage you to memorize them by working through in detail the many examples and exercises based upon them in subsequent sections.

$$\sigma_0 \equiv I \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_1 \equiv \sigma_x \equiv X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\sigma_2 \equiv \sigma_y \equiv Y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_3 \equiv \sigma_z \equiv Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Figure 2.2. The Pauli matrices. Sometimes I is omitted from the list with just X, Y and Z known as the Pauli matrices.

2.1.4 Inner products

An *inner product* is a function which takes as input two vectors $|v\rangle$ and $|w\rangle$ from a vector space and produces a complex number as output. For the time being, it will be convenient to write the inner product of $|v\rangle$ and $|w\rangle$ as $(|v\rangle, |w\rangle)$. This is not the standard quantum mechanical notation; for pedagogical clarity the $(\cdot, \cdot)$ notation will be useful occasionally in this chapter. The standard quantum mechanical notation for the inner product $(|v\rangle, |w\rangle)$ is $\langle v|w\rangle$, where $|v\rangle$ and $|w\rangle$ are vectors in the inner product space, and the notation $\langle v|$ is used for the *dual vector* to the vector $|v\rangle$; the dual is a linear operator from the inner product space V to the complex numbers $\mathbf{C}$, defined by $\langle v|(|w\rangle) \equiv \langle v|w\rangle \equiv (|v\rangle, |w\rangle)$. We will see shortly that the matrix representation of dual vectors is just a row vector.

A function $(\cdot, \cdot)$ from $V \times V$ to $\mathbf{C}$ is an inner product if it satisfies the requirements that:

(1) $(\cdot, \cdot)$ is linear in the second argument,

$$\left(|v\rangle, \sum_i \lambda_i |w_i\rangle \right) = \sum_i \lambda_i \left(|v\rangle, |w_i\rangle \right). \tag{2.13}$$

(2) $(|v\rangle, |w\rangle) = (|w\rangle, |v\rangle)^*$.
(3) $(|v\rangle, |v\rangle) \geq 0$ with equality if and only if $|v\rangle = 0$.

For example, $\mathbf{C}^n$ has an inner product defined by

$$((y_1, \ldots, y_n), (z_1, \ldots, z_n)) \equiv \sum_i y_i^* z_i = \begin{bmatrix} y_1^* \ldots y_n^* \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}. \tag{2.14}$$

We call a vector space equipped with an inner product an *inner product space*.

Exercise 2.5: Verify that $(\cdot, \cdot)$ just defined is an inner product on $\mathbf{C}^n$.

Exercise 2.6: Show that any inner product $(\cdot, \cdot)$ is conjugate-linear in the first argument,

$$\left(\sum_i \lambda_i |w_i\rangle, |v\rangle \right) = \sum_i \lambda_i^* (|w_i\rangle, |v\rangle). \tag{2.15}$$

Discussions of quantum mechanics often refer to *Hilbert space*. In the finite dimensional complex vector spaces that come up in quantum computation and quantum information, a Hilbert space is *exactly the same thing* as an inner product space. From now on we use the two terms interchangeably, preferring the term Hilbert space. In infinite dimensions Hilbert spaces satisfy additional technical restrictions above and beyond inner product spaces, which we will not need to worry about.

Vectors $|w\rangle$ and $|v\rangle$ are *orthogonal* if their inner product is zero. For example, $|w\rangle \equiv (1, 0)$ and $|v\rangle \equiv (0, 1)$ are orthogonal with respect to the inner product defined by (2.14). We define the *norm* of a vector $|v\rangle$ by

$$\| |v\rangle \| \equiv \sqrt{\langle v|v\rangle}\,. \tag{2.16}$$

A *unit vector* is a vector $|v\rangle$ such that $\| |v\rangle \| = 1$. We also say that $|v\rangle$ is *normalized* if $\| |v\rangle \| = 1$. It is convenient to talk of *normalizing* a vector by dividing by its norm; thus $|v\rangle / \| |v\rangle \|$ is the *normalized* form of $|v\rangle$, for any non-zero vector $|v\rangle$. A set $|i\rangle$ of vectors with index i is *orthonormal* if each vector is a unit vector, and distinct vectors in the set are orthogonal, that is, $\langle i|j\rangle = \delta_{ij}$, where i and j are both chosen from the index set.

Exercise 2.7: Verify that $|w\rangle \equiv (1, 1)$ and $|v\rangle \equiv (1, -1)$ are orthogonal. What are the normalized forms of these vectors?

Suppose $|w_1\rangle, \ldots, |w_d\rangle$ is a basis set for some vector space V with an inner product. There is a useful method, the *Gram–Schmidt* procedure, which can be used to produce an orthonormal basis set $|v_1\rangle, \ldots, |v_d\rangle$ for the vector space V. Define $|v_1\rangle \equiv |w_1\rangle / \| |w_1\rangle \|$, and for $1 \leq k \leq d-1$ define $|v_{k+1}\rangle$ inductively by

$$|v_{k+1}\rangle \equiv \frac{|w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle}{\| |w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle \|}. \tag{2.17}$$

It is not difficult to verify that the vectors $|v_1\rangle, \ldots, |v_d\rangle$ form an orthonormal set which is also a basis for V. Thus, any finite dimensional vector space of dimension d has an orthonormal basis, $|v_1\rangle, \ldots, |v_d\rangle$.

Exercise 2.8: Prove that the Gram–Schmidt procedure produces an orthonormal basis for V.

From now on, when we speak of a matrix representation for a linear operator, we mean a matrix representation with respect to orthonormal input and output bases. We also use the convention that if the input and output spaces for a linear operator are the same, then the input and output bases are the same, unless noted otherwise.

With these conventions, the inner product on a Hilbert space can be given a convenient matrix representation. Let $|w\rangle = \sum_i w_i|i\rangle$ and $|v\rangle = \sum_j v_j|j\rangle$ be representations of vectors $|w\rangle$ and $|v\rangle$ with respect to some orthonormal basis $|i\rangle$. Then, since $\langle i|j\rangle = \delta_{ij}$,

$$\langle v|w\rangle = \left(\sum_i v_i|i\rangle, \sum_j w_j|j\rangle\right) = \sum_{ij} v_i^* w_j \delta_{ij} = \sum_i v_i^* w_i \tag{2.18}$$

$$= \begin{bmatrix} v_1^* \dots v_n^* \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}. \tag{2.19}$$

That is, the inner product of two vectors is equal to the vector inner product between two matrix representations of those vectors, provided the representations are written with respect to the same orthonormal basis. We also see that the dual vector $\langle v|$ has a nice interpretation as the row vector whose components are complex conjugates of the corresponding components of the column vector representation of $|v\rangle$.

There is a useful way of representing linear operators which makes use of the inner product, known as the *outer product* representation. Suppose $|v\rangle$ is a vector in an inner product space V, and $|w\rangle$ is a vector in an inner product space W. Define $|w\rangle\langle v|$ to be the linear operator from V to W whose action is defined by

$$\left(|w\rangle\langle v|\right)\left(|v'\rangle\right) \equiv |w\rangle\,\langle v|v'\rangle = \langle v|v'\rangle|w\rangle. \tag{2.20}$$

This equation fits beautifully into our notational conventions, according to which the expression $|w\rangle\langle v|v'\rangle$ could potentially have one of two meanings: we will use it to denote the result when the *operator* $|w\rangle\langle v|$ acts on $|v'\rangle$, and it has an existing interpretation as the result of multiplying $|w\rangle$ by the complex number $\langle v|v'\rangle$. Our definitions are chosen so that these two potential meanings coincide. Indeed, we *define* the former in terms of the latter!

We can take linear combinations of outer product operators $|w\rangle\langle v|$ in the obvious way. By definition $\sum_i a_i|w_i\rangle\langle v_i|$ is the linear operator which, when acting on $|v'\rangle$, produces $\sum_i a_i|w_i\rangle\langle v_i|v'\rangle$ as output.

The usefulness of the outer product notation can be discerned from an important result known as the *completeness relation* for orthonormal vectors. Let $|i\rangle$ be any orthonormal basis for the vector space V, so an arbitrary vector $|v\rangle$ can be written $|v\rangle = \sum_i v_i|i\rangle$ for some set of complex numbers v_i. Note that $\langle i|v\rangle = v_i$ and therefore

$$\left(\sum_i |i\rangle\langle i|\right)|v\rangle = \sum_i |i\rangle\langle i|v\rangle = \sum_i v_i|i\rangle = |v\rangle. \tag{2.21}$$

Since the last equation is true for all $|v\rangle$ it follows that

$$\sum_i |i\rangle\langle i| = I. \tag{2.22}$$

This equation is known as the *completeness relation.* One application of the completeness relation is to give a means for representing any operator in the outer product notation. Suppose $A : V \to W$ is a linear operator, $|v_i\rangle$ is an orthonormal basis for V, and $|w_j\rangle$ an orthonormal basis for W. Using the completeness relation twice we obtain

$$A = I_W A I_V \tag{2.23}$$

$$= \sum_{ij} |w_j\rangle\langle w_j|A|v_i\rangle\langle v_i| \tag{2.24}$$

$$= \sum_{ij} \langle w_j|A|v_i\rangle |w_j\rangle\langle v_i|, \tag{2.25}$$

which is the outer product representation for A. We also see from this equation that A has matrix element $\langle w_j|A|v_i\rangle$ in the ith column and jth row, with respect to the input basis $|v_i\rangle$ and output basis $|w_j\rangle$.

A second application illustrating the usefulness of the completeness relation is the *Cauchy–Schwarz inequality*. This important result is discussed in Box 2.1, on this page.

Exercise 2.9: (Pauli operators and the outer product) The Pauli matrices (Figure 2.2 on page 65) can be considered as operators with respect to an orthonormal basis $|0\rangle, |1\rangle$ for a two-dimensional Hilbert space. Express each of the Pauli operators in the outer product notation.

Exercise 2.10: Suppose $|v_i\rangle$ is an orthonormal basis for an inner product space V. What is the matrix representation for the operator $|v_j\rangle\langle v_k|$, with respect to the $|v_i\rangle$ basis?

Box 2.1: The Cauchy-Schwarz inequality

The *Cauchy–Schwarz inequality* is an important geometric fact about Hilbert spaces. It states that for any two vectors $|v\rangle$ and $|w\rangle$, $|\langle v|w\rangle|^2 \leq \langle v|v\rangle\langle w|w\rangle$. To see this, use the Gram–Schmidt procedure to construct an orthonormal basis $|i\rangle$ for the vector space such that the first member of the basis $|i\rangle$ is $|w\rangle/\sqrt{\langle w|w\rangle}$. Using the completeness relation $\sum_i |i\rangle\langle i| = I$, and dropping some non-negative terms gives

$$\langle v|v\rangle\langle w|w\rangle = \sum_i \langle v|i\rangle\langle i|v\rangle\langle w|w\rangle \tag{2.26}$$

$$\geq \frac{\langle v|w\rangle\langle w|v\rangle}{\langle w|w\rangle}\langle w|w\rangle \tag{2.27}$$

$$= \langle v|w\rangle\langle w|v\rangle = |\langle v|w\rangle|^2, \tag{2.28}$$

as required. A little thought shows that equality occurs if and only if $|v\rangle$ and $|w\rangle$ are linearly related, $|v\rangle = z|w\rangle$ or $|w\rangle = z|v\rangle$, for some scalar z.

2.1.5 Eigenvectors and eigenvalues

An *eigenvector* of a linear operator A on a vector space is a non-zero vector $|v\rangle$ such that $A|v\rangle = v|v\rangle$, where v is a complex number known as the *eigenvalue* of A corresponding to $|v\rangle$. It will often be convenient to use the notation v both as a label for the eigenvector, and to represent the eigenvalue. We assume that you are familiar with the elementary properties of eigenvalues and eigenvectors – in particular, how to find them, via the characteristic equation. The *characteristic function* is defined to be $c(\lambda) \equiv \det|A - \lambda I|$,

where det is the *determinant* function for matrices; it can be shown that the characteristic function depends only upon the operator A, and not on the specific matrix representation used for A. The solutions of the *characteristic equation* $c(\lambda) = 0$ are the eigenvalues of the operator A. By the fundamental theorem of algebra, every polynomial has at least one complex root, so every operator A has at least one eigenvalue, and a corresponding eigenvector. The *eigenspace* corresponding to an eigenvalue v is the set of vectors which have eigenvalue v. It is a vector subspace of the vector space on which A acts.

A *diagonal representation* for an operator A on a vector space V is a representation $A = \sum_i \lambda_i |i\rangle\langle i|$, where the vectors $|i\rangle$ form an orthonormal set of eigenvectors for A, with corresponding eigenvalues λ_i. An operator is said to be *diagonalizable* if it has a diagonal representation. In the next section we will find a simple set of necessary and sufficient conditions for an operator on a Hilbert space to be diagonalizable. As an example of a diagonal representation, note that the Pauli Z matrix may be written

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = |0\rangle\langle 0| - |1\rangle\langle 1|, \tag{2.29}$$

where the matrix representation is with respect to orthonormal vectors $|0\rangle$ and $|1\rangle$, respectively. Diagonal representations are sometimes also known as *orthonormal decompositions*.

When an eigenspace is more than one dimensional we say that it is *degenerate*. For example, the matrix A defined by

$$A \equiv \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.30}$$

has a two-dimensional eigenspace corresponding to the eigenvalue 2. The eigenvectors $(1, 0, 0)$ and $(0, 1, 0)$ are said to be *degenerate* because they are linearly independent eigenvectors of A with the same eigenvalue.

Exercise 2.11: (Eigendecomposition of the Pauli matrices) Find the eigenvectors, eigenvalues, and diagonal representations of the Pauli matrices X, Y, and Z.

Exercise 2.12: Prove that the matrix

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{2.31}$$

is not diagonalizable.

2.1.6 Adjoints and Hermitian operators

Suppose A is any linear operator on a Hilbert space, V. It turns out that there exists a unique linear operator $A^\dagger$ on V such that for all vectors $|v\rangle, |w\rangle \in V$,

$$(|v\rangle, A|w\rangle) = (A^\dagger|v\rangle, |w\rangle). \tag{2.32}$$

This linear operator is known as the *adjoint* or *Hermitian conjugate* of the operator A. From the definition it is easy to see that $(AB)^\dagger = B^\dagger A^\dagger$. By convention, if $|v\rangle$ is a vector, then we define $|v\rangle^\dagger \equiv \langle v|$. With this definition it is not difficult to see that $(A|v\rangle)^\dagger = \langle v|A^\dagger$.

Exercise 2.13: If $|w\rangle$ and $|v\rangle$ are any two vectors, show that $(|w\rangle\langle v|)^\dagger = |v\rangle\langle w|$.

Exercise 2.14: (Anti-linearity of the adjoint) Show that the adjoint operation is anti-linear,

$$\left(\sum_i a_i A_i\right)^\dagger = \sum_i a_i^* A_i^\dagger. \tag{2.33}$$

Exercise 2.15: Show that $(A^\dagger)^\dagger = A$.

In a matrix representation of an operator A, the action of the Hermitian conjugation operation is to take the matrix of A to the conjugate-transpose matrix, $A^\dagger \equiv (A^*)^T$, where the $*$ indicates complex conjugation, and T indicates the transpose operation. For example, we have

$$\begin{bmatrix} 1+3i & 2i \\ 1+i & 1-4i \end{bmatrix}^\dagger = \begin{bmatrix} 1-3i & 1-i \\ -2i & 1+4i \end{bmatrix}. \tag{2.34}$$

An operator A whose adjoint is A is known as a *Hermitian* or *self-adjoint* operator. An important class of Hermitian operators is the *projectors*. Suppose W is a k-dimensional vector subspace of the d-dimensional vector space V. Using the Gram–Schmidt procedure it is possible to construct an orthonormal basis $|1\rangle, \ldots, |d\rangle$ for V such that $|1\rangle, \ldots, |k\rangle$ is an orthonormal basis for W. By definition,

$$P \equiv \sum_{i=1}^{k} |i\rangle\langle i| \tag{2.35}$$

is the *projector* onto the subspace W. It is easy to check that this definition is independent of the orthonormal basis $|1\rangle, \ldots, |k\rangle$ used for W. From the definition it can be shown that $|v\rangle\langle v|$ is Hermitian for any vector $|v\rangle$, so P is Hermitian, $P^\dagger = P$. We will often refer to the 'vector space' P, as shorthand for the vector space onto which P is a projector. The *orthogonal complement* of P is the operator $Q \equiv I - P$. It is easy to see that Q is a projector onto the vector space spanned by $|k+1\rangle, \ldots, |d\rangle$, which we also refer to as the *orthogonal complement* of P, and may denote by Q.

Exercise 2.16: Show that any projector P satisfies the equation $P^2 = P$.

An operator A is said to be *normal* if $AA^\dagger = A^\dagger A$. Clearly, an operator which is Hermitian is also normal. There is a remarkable representation theorem for normal operators known as the *spectral decomposition*, which states that an operator is a normal operator if and only if it is diagonalizable. This result is proved in Box 2.2 on page 72, which you should read closely.

Exercise 2.17: Show that a normal matrix is Hermitian if and only if it has real eigenvalues.

A matrix U is said to be *unitary* if $U^\dagger U = I$. Similarly an operator U is unitary if $U^\dagger U = I$. It is easily checked that an operator is unitary if and only if each of its matrix representations is unitary. A unitary operator also satisfies $UU^\dagger = I$, and therefore U is normal and has a spectral decomposition. Geometrically, unitary operators are important because they preserve inner products between vectors. To see this, let $|v\rangle$ and $|w\rangle$ be any

two vectors. Then the inner product of $U|v\rangle$ and $U|w\rangle$ is the same as the inner product of $|v\rangle$ and $|w\rangle$,

$$(U|v\rangle, U|w\rangle) = \langle v|U^\dagger U|w\rangle = \langle v|I|w\rangle = \langle v|w\rangle. \tag{2.36}$$

This result suggests the following elegant outer product representation of any unitary U. Let $|v_i\rangle$ be any orthonormal basis set. Define $|w_i\rangle \equiv U|v_i\rangle$, so $|w_i\rangle$ is also an orthonormal basis set, since unitary operators preserve inner products. Note that $U = \sum_i |w_i\rangle\langle v_i|$. Conversely, if $|v_i\rangle$ and $|w_i\rangle$ are any two orthonormal bases, then it is easily checked that the operator U defined by $U \equiv \sum_i |w_i\rangle\langle v_i|$ is a unitary operator.

Exercise 2.18: Show that all eigenvalues of a unitary matrix have modulus 1, that is, can be written in the form $e^{i\theta}$ for some real θ.

Exercise 2.19: (Pauli matrices: Hermitian and unitary) Show that the Pauli matrices are Hermitian and unitary.

Exercise 2.20: (Basis changes) Suppose A' and A'' are matrix representations of an operator A on a vector space V with respect to two different orthonormal bases, $|v_i\rangle$ and $|w_i\rangle$. Then the elements of A' and A'' are $A'_{ij} = \langle v_i|A|v_j\rangle$ and $A''_{ij} = \langle w_i|A|w_j\rangle$. Characterize the relationship between A' and A''.

A special subclass of Hermitian operators is extremely important. This is the *positive operators*. A positive operator A is defined to be an operator such that for any vector $|v\rangle$, $(|v\rangle, A|v\rangle)$ is a real, non-negative number. If $(|v\rangle, A|v\rangle)$ is *strictly* greater than zero for all $|v\rangle \neq 0$ then we say that A is *positive definite*. In Exercise 2.24 on this page you will show that any positive operator is automatically Hermitian, and therefore by the spectral decomposition has diagonal representation $\sum_i \lambda_i |i\rangle\langle i|$, with non-negative eigenvalues λ_i.

Exercise 2.21: Repeat the proof of the spectral decomposition in Box 2.2 for the case when M is Hermitian, simplifying the proof wherever possible.

Exercise 2.22: Prove that two eigenvectors of a Hermitian operator with different eigenvalues are necessarily orthogonal.

Exercise 2.23: Show that the eigenvalues of a projector P are all either 0 or 1.

Exercise 2.24: (Hermiticity of positive operators) Show that a positive operator is necessarily Hermitian. (*Hint*: Show that an arbitrary operator A can be written $A = B + iC$ where B and C are Hermitian.)

Exercise 2.25: Show that for any operator A, $A^\dagger A$ is positive.

2.1.7 Tensor products

The *tensor product* is a way of putting vector spaces together to form larger vector spaces. This construction is crucial to understanding the quantum mechanics of multiparticle systems. The following discussion is a little abstract, and may be difficult to follow if you're not already familiar with the tensor product, so feel free to skip ahead now and revisit later when you come to the discussion of tensor products in quantum mechanics.

Suppose V and W are vector spaces of dimension m and n respectively. For convenience we also suppose that V and W are Hilbert spaces. Then $V \otimes W$ (read 'V tensor

Box 2.2: The spectral decomposition – important!

The *spectral decomposition* is an extremely useful representation theorem for normal operators.

Theorem 2.1: (**Spectral decomposition**) Any normal operator M on a vector space V is diagonal with respect to some orthonormal basis for V. Conversely, any diagonalizable operator is normal.

Proof

The converse is a simple exercise, so we prove merely the forward implication, by induction on the dimension d of V. The case $d = 1$ is trivial. Let λ be an eigenvalue of M, P the projector onto the λ eigenspace, and Q the projector onto the orthogonal complement. Then $M = (P + Q)M(P + Q) = PMP + QMP + PMQ + QMQ$. Obviously $PMP = \lambda P$. Furthermore, $QMP = 0$, as M takes the subspace P into itself. We claim that $PMQ = 0$ also. To see this, let $|v\rangle$ be an element of the subspace P. Then $MM^\dagger|v\rangle = M^\dagger M|v\rangle = \lambda M^\dagger|v\rangle$. Thus, $M^\dagger|v\rangle$ has eigenvalue λ and therefore is an element of the subspace P. It follows that $QM^\dagger P = 0$. Taking the adjoint of this equation gives $PMQ = 0$. Thus $M = PMP + QMQ$. Next, we prove that QMQ is normal. To see this, note that $QM = QM(P + Q) = QMQ$, and $QM^\dagger = QM^\dagger(P + Q) = QM^\dagger Q$. Therefore, by the normality of M, and the observation that $Q^2 = Q$,

$$QMQ\,QM^\dagger Q = QMQM^\dagger Q \tag{2.37}$$
$$= QMM^\dagger Q \tag{2.38}$$
$$= QM^\dagger MQ \tag{2.39}$$
$$= QM^\dagger QMQ \tag{2.40}$$
$$= QM^\dagger Q\,QMQ\,, \tag{2.41}$$

so QMQ is normal. By induction, QMQ is diagonal with respect to some orthonormal basis for the subspace Q, and PMP is already diagonal with respect to some orthonormal basis for P. It follows that $M = PMP + QMQ$ is diagonal with respect to some orthonormal basis for the total vector space. □

In terms of the outer product representation, this means that M can be written as $M = \sum_i \lambda_i |i\rangle\langle i|$, where λ_i are the eigenvalues of M, $|i\rangle$ is an orthonormal basis for V, and each $|i\rangle$ an eigenvector of M with eigenvalue λ_i. In terms of projectors, $M = \sum_i \lambda_i P_i$, where λ_i are again the eigenvalues of M, and P_i is the projector onto the λ_i eigenspace of M. These projectors satisfy the completeness relation $\sum_i P_i = I$, and the orthonormality relation $P_i P_j = \delta_{ij} P_i$.

W') is an mn dimensional vector space. The elements of $V \otimes W$ are linear combinations of 'tensor products' $|v\rangle \otimes |w\rangle$ of elements $|v\rangle$ of V and $|w\rangle$ of W. In particular, if $|i\rangle$ and $|j\rangle$ are orthonormal bases for the spaces V and W then $|i\rangle \otimes |j\rangle$ is a basis for $V \otimes W$. We often use the abbreviated notations $|v\rangle|w\rangle$, $|v, w\rangle$ or even $|vw\rangle$ for the tensor product

$|v\rangle \otimes |w\rangle$. For example, if V is a two-dimensional vector space with basis vectors $|0\rangle$ and $|1\rangle$ then $|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle$ is an element of $V \otimes V$.

By definition the tensor product satisfies the following basic properties:

(1) For an arbitrary scalar z and elements $|v\rangle$ of V and $|w\rangle$ of W,

$$z\left(|v\rangle \otimes |w\rangle\right) = \left(z|v\rangle\right) \otimes |w\rangle = |v\rangle \otimes \left(z|w\rangle\right). \tag{2.42}$$

(2) For arbitrary $|v_1\rangle$ and $|v_2\rangle$ in V and $|w\rangle$ in W,

$$\left(|v_1\rangle + |v_2\rangle\right) \otimes |w\rangle = |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle. \tag{2.43}$$

(3) For arbitrary $|v\rangle$ in V and $|w_1\rangle$ and $|w_2\rangle$ in W,

$$|v\rangle \otimes \left(|w_1\rangle + |w_2\rangle\right) = |v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle. \tag{2.44}$$

What sorts of linear operators act on the space $V \otimes W$? Suppose $|v\rangle$ and $|w\rangle$ are vectors in V and W, and A and B are linear operators on V and W, respectively. Then we can define a linear operator $A \otimes B$ on $V \otimes W$ by the equation

$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A|v\rangle \otimes B|w\rangle. \tag{2.45}$$

The definition of $A \otimes B$ is then extended to all elements of $V \otimes W$ in the natural way to ensure linearity of $A \otimes B$, that is,

$$(A \otimes B)\left(\sum_i a_i |v_i\rangle \otimes |w_i\rangle\right) \equiv \sum_i a_i A|v_i\rangle \otimes B|w_i\rangle. \tag{2.46}$$

It can be shown that $A \otimes B$ defined in this way is a well-defined linear operator on $V \otimes W$. This notion of the tensor product of two operators extends in the obvious way to the case where $A : V \to V'$ and $B : W \to W'$ map between different vector spaces. Indeed, an arbitrary linear operator C mapping $V \otimes W$ to $V' \otimes W'$ can be represented as a linear combination of tensor products of operators mapping V to V' and W to W',

$$C = \sum_i c_i A_i \otimes B_i, \tag{2.47}$$

where by definition

$$\left(\sum_i c_i A_i \otimes B_i\right) |v\rangle \otimes |w\rangle \equiv \sum_i c_i A_i |v\rangle \otimes B_i |w\rangle. \tag{2.48}$$

The inner products on the spaces V and W can be used to define a natural inner product on $V \otimes W$. Define

$$\left(\sum_i a_i |v_i\rangle \otimes |w_i\rangle, \sum_j b_j |v'_j\rangle \otimes |w'_j\rangle\right) \equiv \sum_{ij} a_i^* b_j \langle v_i|v'_j\rangle \langle w_i|w'_j\rangle. \tag{2.49}$$

It can be shown that the function so defined is a well-defined inner product. From this inner product, the inner product space $V \otimes W$ inherits the other structure we are familiar with, such as notions of an adjoint, unitarity, normality, and Hermiticity.

All this discussion is rather abstract. It can be made much more concrete by moving

to a convenient matrix representation known as the *Kronecker product*. Suppose A is an m by n matrix, and B is a p by q matrix. Then we have the matrix representation:

$$A \otimes B \equiv \overbrace{\left[\begin{array}{cccc} A_{11}B & A_{12}B & \ldots & A_{1n}B \\ A_{21}B & A_{22}B & \ldots & A_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1}B & A_{m2}B & \ldots & A_{mn}B \end{array}\right]}^{nq} \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} mp \,. \tag{2.50}$$

In this representation terms like $A_{11}B$ denote p by q submatrices whose entries are proportional to B, with overall proportionality constant A_{11}. For example, the tensor product of the vectors $(1, 2)$ and $(2, 3)$ is the vector

$$\left[\begin{array}{c} 1 \\ 2 \end{array}\right] \otimes \left[\begin{array}{c} 2 \\ 3 \end{array}\right] = \left[\begin{array}{c} 1 \times 2 \\ 1 \times 3 \\ 2 \times 2 \\ 2 \times 3 \end{array}\right] = \left[\begin{array}{c} 2 \\ 3 \\ 4 \\ 6 \end{array}\right] . \tag{2.51}$$

The tensor product of the Pauli matrices X and Y is

$$X \otimes Y = \left[\begin{array}{cc} 0 \cdot Y & 1 \cdot Y \\ 1 \cdot Y & 0 \cdot Y \end{array}\right] = \left[\begin{array}{cccc} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{array}\right] . \tag{2.52}$$

Finally, we mention the useful notation $|\psi\rangle^{\otimes k}$, which means $|\psi\rangle$ tensored with itself k times. For example $|\psi\rangle^{\otimes 2} = |\psi\rangle \otimes |\psi\rangle$. An analogous notation is also used for operators on tensor product spaces.

Exercise 2.26: Let $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Write out $|\psi\rangle^{\otimes 2}$ and $|\psi\rangle^{\otimes 3}$ explicitly, both in terms of tensor products like $|0\rangle|1\rangle$, and using the Kronecker product.

Exercise 2.27: Calculate the matrix representation of the tensor products of the Pauli operators (a) X and Z; (b) I and X; (c) X and I. Is the tensor product commutative?

Exercise 2.28: Show that the transpose, complex conjugation, and adjoint operations distribute over the tensor product,

$$(A \otimes B)^* = A^* \otimes B^*; \; (A \otimes B)^T = A^T \otimes B^T; \; (A \otimes B)^\dagger = A^\dagger \otimes B^\dagger . \tag{2.53}$$

Exercise 2.29: Show that the tensor product of two unitary operators is unitary.

Exercise 2.30: Show that the tensor product of two Hermitian operators is Hermitian.

Exercise 2.31: Show that the tensor product of two positive operators is positive.

Exercise 2.32: Show that the tensor product of two projectors is a projector.

Exercise 2.33: The Hadamard operator on one qubit may be written as

$$H = \frac{1}{\sqrt{2}} \Big[(|0\rangle + |1\rangle)\langle 0| + (|0\rangle - |1\rangle)\langle 1| \Big] . \tag{2.54}$$

Show explicitly that the Hadamard transform on n qubits, $H^{\otimes n}$, may be written as

$$H^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x,y} (-1)^{x \cdot y} |x\rangle\langle y|. \tag{2.55}$$

Write out an explicit matrix representation for $H^{\otimes 2}$.

2.1.8 Operator functions

There are many important functions which can be defined for operators and matrices. Generally speaking, given a function f from the complex numbers to the complex numbers, it is possible to define a corresponding matrix function on normal matrices (or some subclass, such as the Hermitian matrices) by the following construction. Let $A = \sum_a a|a\rangle\langle a|$ be a spectral decomposition for a normal operator A. Define $f(A) \equiv \sum_a f(a)|a\rangle\langle a|$. A little thought shows that $f(A)$ is uniquely defined. This procedure can be used, for example, to define the square root of a positive operator, the logarithm of a positive-definite operator, or the exponential of a normal operator. As an example,

$$\exp(\theta Z) = \begin{bmatrix} e^{\theta} & 0 \\ 0 & e^{-\theta} \end{bmatrix}, \tag{2.56}$$

since Z has eigenvectors $|0\rangle$ and $|1\rangle$.

Exercise 2.34: Find the square root and logarithm of the matrix

$$\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}. \tag{2.57}$$

Exercise 2.35: (Exponential of the Pauli matrices) Let $\vec{v}$ be any real, three-dimensional unit vector and θ a real number. Prove that

$$\exp(i\theta \vec{v} \cdot \vec{\sigma}) = \cos(\theta) I + i \sin(\theta) \vec{v} \cdot \vec{\sigma}, \tag{2.58}$$

where $\vec{v} \cdot \vec{\sigma} \equiv \sum_{i=1}^{3} v_i \sigma_i$. This exercise is generalized in Problem 2.1 on page 117.

Another important matrix function is the *trace* of a matrix. The trace of A is defined to be the sum of its diagonal elements,

$$\mathrm{tr}(A) \equiv \sum_i A_{ii}. \tag{2.59}$$

The trace is easily seen to be *cyclic*, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, and *linear*, $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$, $\mathrm{tr}(zA) = z\,\mathrm{tr}(A)$, where A and B are arbitrary matrices, and z is a complex number. Furthermore, from the cyclic property it follows that the trace of a matrix is invariant under the unitary *similarity transformation* $A \rightarrow UAU^{\dagger}$, as $\mathrm{tr}(UAU^{\dagger}) = \mathrm{tr}(U^{\dagger}UA) = \mathrm{tr}(A)$. In light of this result, it makes sense to define the trace of an *operator* A to be the trace of any matrix representation of A. The invariance of the trace under unitary similarity transformations ensures that the trace of an operator is well defined.

As an example of the trace, suppose $|\psi\rangle$ is a unit vector and A is an arbitrary operator. To evaluate $\mathrm{tr}(A|\psi\rangle\langle\psi|)$ use the Gram–Schmidt procedure to extend $|\psi\rangle$ to an

orthonormal basis $|i\rangle$ which includes $|\psi\rangle$ as the first element. Then we have

$$\text{tr}(A|\psi\rangle\langle\psi|) = \sum_i \langle i|A|\psi\rangle\langle\psi|i\rangle \tag{2.60}$$

$$= \langle\psi|A|\psi\rangle. \tag{2.61}$$

This result, that $\text{tr}(A|\psi\rangle\langle\psi|) = \langle\psi|A|\psi\rangle$ is extremely useful in evaluating the trace of an operator.

Exercise 2.36: Show that the Pauli matrices except for I have trace zero.

Exercise 2.37: (Cyclic property of the trace) If A and B are two linear operators show that

$$\text{tr}(AB) = \text{tr}(BA). \tag{2.62}$$

Exercise 2.38: (Linearity of the trace) If A and B are two linear operators, show that

$$\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B) \tag{2.63}$$

and if z is an arbitrary complex number show that

$$\text{tr}(zA) = z\text{tr}(A). \tag{2.64}$$

Exercise 2.39: (The Hilbert–Schmidt inner product on operators) The set L_V of linear operators on a Hilbert space V is obviously a vector space – the sum of two linear operators is a linear operator, zA is a linear operator if A is a linear operator and z is a complex number, and there is a zero element 0. An important additional result is that the vector space L_V can be given a natural inner product structure, turning it into a Hilbert space.

(1) Show that the function $(\cdot,\cdot)$ on $L_V \times L_V$ defined by

$$(A, B) \equiv \text{tr}(A^\dagger B) \tag{2.65}$$

is an inner product function. This inner product is known as the *Hilbert–Schmidt* or *trace* inner product.

(2) If V has d dimensions show that L_V has dimension d^2.

(3) Find an orthonormal basis of Hermitian matrices for the Hilbert space L_V.

2.1.9 The commutator and anti-commutator

The *commutator* between two operators A and B is defined to be

$$[A, B] \equiv AB - BA. \tag{2.66}$$

If $[A, B] = 0$, that is, $AB = BA$, then we say A *commutes* with B. Similarly, the *anti-commutator* of two operators A and B is defined by

$$\{A, B\} \equiv AB + BA; \tag{2.67}$$

we say A *anti-commutes* with B if $\{A, B\} = 0$. It turns out that many important properties of pairs of operators can be deduced from their commutator and anti-commutator. Perhaps the most useful relation is the following connection between the commutator and the property of being able to *simultaneously diagonalize* Hermitian operators A and B,

that is, write $A = \sum_i a_i |i\rangle\langle i|$, $B = \sum_i b_i |i\rangle\langle i|$, where $|i\rangle$ is some common orthonormal set of eigenvectors for A and B.

Theorem 2.2: **(Simultaneous diagonalization theorem)** Suppose A and B are Hermitian operators. Then $[A, B] = 0$ if and only if there exists an orthonormal basis such that both A and B are diagonal with respect to that basis. We say that A and B are *simultaneously diagonalizable* in this case.

This result connects the commutator of two operators, which is often easy to compute, to the property of being simultaneously diagonalizable, which is *a priori* rather difficult to determine. As an example, consider that

$$[X, Y] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} - \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.68}$$

$$= 2i \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.69}$$

$$= 2iZ\,, \tag{2.70}$$

so X and Y do not commute. You have already shown, in Exercise 2.11, that X and Y do not have common eigenvectors, as we expect from the simultaneous diagonalization theorem.

Proof
You can (and should!) easily verify that if A and B are diagonal in the same orthonormal basis then $[A, B] = 0$. To show the converse, let $|a, j\rangle$ be an orthonormal basis for the eigenspace V_a of A with eigenvalue a; the index j is used to label possible degeneracies. Note that

$$AB|a, j\rangle = BA|a, j\rangle = aB|a, j\rangle, \tag{2.71}$$

and therefore $B|a, j\rangle$ is an element of the eigenspace V_a. Let P_a denote the projector onto the space V_a and define $B_a \equiv P_a B P_a$. It is easy to see that the restriction of B_a to the space V_a is Hermitian on V_a, and therefore has a spectral decomposition in terms of an orthonormal set of eigenvectors which span the space V_a. Let's call these eigenvectors $|a, b, k\rangle$, where the indices a and b label the eigenvalues of A and B_a, and k is an extra index to allow for the possibility of a degenerate B_a. Note that $B|a, b, k\rangle$ is an element of V_a, so $B|a, b, k\rangle = P_a B|a, b, k\rangle$. Moreover we have $P_a|a, b, k\rangle = |a, b, k\rangle$, so

$$B|a, b, k\rangle = P_a B P_a |a, b, k\rangle = b|a, b, k\rangle. \tag{2.72}$$

It follows that $|a, b, k\rangle$ is an eigenvector of B with eigenvalue b, and therefore $|a, b, k\rangle$ is an orthonormal set of eigenvectors of both A and B, spanning the entire vector space on which A and B are defined. That is, A and B are simultaneously diagonalizable. □

Exercise 2.40: (Commutation relations for the Pauli matrices) Verify the commutation relations

$$[X, Y] = 2iZ; \quad [Y, Z] = 2iX; \quad [Z, X] = 2iY. \tag{2.73}$$

There is an elegant way of writing this using ϵ_{jkl}, the antisymmetric tensor on

three indices, for which $\epsilon_{jkl} = 0$ except for $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, and $\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1$:

$$[\sigma_j, \sigma_k] = 2i \sum_{l=1}^{3} \epsilon_{jkl} \sigma_l. \tag{2.74}$$

Exercise 2.41: (**Anti-commutation relations for the Pauli matrices**) Verify the anti-commutation relations

$$\{\sigma_i, \sigma_j\} = 0 \tag{2.75}$$

where $i \neq j$ are both chosen from the set $1, 2, 3$. Also verify that ($i = 0, 1, 2, 3$)

$$\sigma_i^2 = I. \tag{2.76}$$

Exercise 2.42: Verify that

$$AB = \frac{[A, B] + \{A, B\}}{2}. \tag{2.77}$$

Exercise 2.43: Show that for $j, k = 1, 2, 3$,

$$\sigma_j \sigma_k = \delta_{jk} I + i \sum_{l=1}^{3} \epsilon_{jkl} \sigma_l. \tag{2.78}$$

Exercise 2.44: Suppose $[A, B] = 0$, $\{A, B\} = 0$, and A is invertible. Show that B must be 0.

Exercise 2.45: Show that $[A, B]^\dagger = [B^\dagger, A^\dagger]$.

Exercise 2.46: Show that $[A, B] = -[B, A]$.

Exercise 2.47: Suppose A and B are Hermitian. Show that $i[A, B]$ is Hermitian.

2.1.10 The polar and singular value decompositions

The *polar* and *singular value* decompositions are useful ways of breaking linear operators up into simpler parts. In particular, these decompositions allow us to break general linear operators up into products of unitary operators and positive operators. While we don't understand the structure of general linear operators terribly well, we do understand unitary operators and positive operators in quite some detail. The polar and singular value decompositions allow us to apply this understanding to better understand general linear operators.

Theorem 2.3: (**Polar decomposition**) Let A be a linear operator on a vector space V. Then there exists unitary U and positive operators J and K such that

$$A = UJ = KU, \tag{2.79}$$

where the unique positive operators J and K satisfying these equations are defined by $J \equiv \sqrt{A^\dagger A}$ and $K \equiv \sqrt{AA^\dagger}$. Moreover, if A is invertible then U is unique.

We call the expression $A = UJ$ the *left polar decomposition* of A, and $A = KU$ the *right polar decomposition* of A. Most often, we'll omit the 'right' or 'left' nomenclature, and use the term 'polar decomposition' for both expressions, with context indicating which is meant.

Proof
$J \equiv \sqrt{A^\dagger A}$ is a positive operator, so it can be given a spectral decomposition, $J = \sum_i \lambda_i |i\rangle\langle i|$ ($\lambda_i \geq 0$). Define $|\psi_i\rangle \equiv A|i\rangle$. From the definition, we see that $\langle\psi_i|\psi_i\rangle = \lambda_i^2$. Consider for now only those i for which $\lambda_i \neq 0$. For those i define $|e_i\rangle \equiv |\psi_i\rangle/\lambda_i$, so the $|e_i\rangle$ are normalized. Moreover, they are orthogonal, since if $i \neq j$ then $\langle e_i|e_j\rangle = \langle i|A^\dagger A|j\rangle/\lambda_i\lambda_j = \langle i|J^2|j\rangle/\lambda_i\lambda_j = 0$.

We have been considering i such that $\lambda_i \neq 0$. Now use the Gram–Schmidt procedure to extend the orthonormal set $|e_i\rangle$ so it forms an orthonormal basis, which we also label $|e_i\rangle$. Define a unitary operator $U \equiv \sum_i |e_i\rangle\langle i|$. When $\lambda_i \neq 0$ we have $UJ|i\rangle = \lambda_i|e_i\rangle = |\psi_i\rangle = A|i\rangle$. When $\lambda_i = 0$ we have $UJ|i\rangle = 0 = |\psi_i\rangle$. We have proved that the action of A and UJ agree on the basis $|i\rangle$, and thus that $A = UJ$.

J is unique, since multiplying $A = UJ$ on the left by the adjoint equation $A^\dagger = JU^\dagger$ gives $J^2 = A^\dagger A$, from which we see that $J = \sqrt{A^\dagger A}$, uniquely. A little thought shows that if A is invertible, then so is J, so U is uniquely determined by the equation $U = AJ^{-1}$. The proof of the right polar decomposition follows, since $A = UJ = UJU^\dagger U = KU$, where $K \equiv UJU^\dagger$ is a positive operator. Since $AA^\dagger = KUU^\dagger K = K^2$ we must have $K = \sqrt{AA^\dagger}$, as claimed. □

The singular value decomposition combines the polar decomposition and the spectral theorem.

Corollary 2.4: (**Singular value decomposition**) Let A be a square matrix. Then there exist unitary matrices U and V, and a diagonal matrix D with non-negative entries such that

$$A = UDV\,. \tag{2.80}$$

The diagonal elements of D are called the *singular values* of A.

Proof
By the polar decomposition, $A = SJ$, for unitary S, and positive J. By the spectral theorem, $J = TDT^\dagger$, for unitary T and diagonal D with non-negative entries. Setting $U \equiv ST$ and $V \equiv T^\dagger$ completes the proof. □

Exercise 2.48: What is the polar decomposition of a positive matrix P? Of a unitary matrix U? Of a Hermitian matrix, H?

Exercise 2.49: Express the polar decomposition of a normal matrix in the outer product representation.

Exercise 2.50: Find the left and right polar decompositions of the matrix

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{2.81}$$

2.2 The postulates of quantum mechanics

All understanding begins with our not accepting the world as it appears.
– Alan Kay

The most incomprehensible thing about the world is that it is comprehensible.
– Albert Einstein

Quantum mechanics is a mathematical framework for the development of physical theories. On its own quantum mechanics doesn't tell you what laws a physical system must obey, but it does provide a mathematical and conceptual framework for the development of such laws. In the next few sections we give a complete description of the basic postulates of quantum mechanics. These postulates provide a connection between the physical world and the mathematical formalism of quantum mechanics.

The postulates of quantum mechanics were derived after a long process of trial and (mostly) error, which involved a considerable amount of guessing and fumbling by the originators of the theory. Don't be surprised if the motivation for the postulates is not always clear; even to experts the basic postulates of quantum mechanics appear surprising. What you should expect to gain in the next few sections is a good working grasp of the postulates – how to apply them, and when.

2.2.1 State space

The first postulate of quantum mechanics sets up the arena in which quantum mechanics takes place. The arena is our familiar friend from linear algebra, Hilbert space.

> **Postulate 1**: Associated to any isolated physical system is a complex vector space with inner product (that is, a Hilbert space) known as the *state space* of the system. The system is completely described by its *state vector*, which is a unit vector in the system's state space.

Quantum mechanics does *not* tell us, for a given physical system, what the state space of that system is, nor does it tell us what the state vector of the system is. Figuring that out for a *specific* system is a difficult problem for which physicists have developed many intricate and beautiful rules. For example, there is the wonderful theory of quantum electrodynamics (often known as QED), which describes how atoms and light interact. One aspect of QED is that it tells us what state spaces to use to give quantum descriptions of atoms and light. We won't be much concerned with the intricacies of theories like QED (except in so far as they apply to physical realizations, in Chapter 7), as we are mostly interested in the general framework provided by quantum mechanics. For our purposes it will be sufficient to make some very simple (and reasonable) assumptions about the state spaces of the systems we are interested in, and stick with those assumptions.

The simplest quantum mechanical system, and the system which we will be most concerned with, is the *qubit*. A qubit has a two-dimensional state space. Suppose $|0\rangle$ and $|1\rangle$ form an orthonormal basis for that state space. Then an arbitrary state vector in the state space can be written

$$|\psi\rangle = a|0\rangle + b|1\rangle, \tag{2.82}$$

where a and b are complex numbers. The condition that $|\psi\rangle$ be a unit vector, $\langle\psi|\psi\rangle = 1$, is therefore equivalent to $|a|^2 + |b|^2 = 1$. The condition $\langle\psi|\psi\rangle = 1$ is often known as the *normalization condition* for state vectors.

We will take the qubit as our fundamental quantum mechanical system. Later, in Chapter 7, we will see that there are real physical systems which may be described in terms of qubits. For now, though, it is sufficient to think of qubits in abstract terms, without reference to a specific realization. Our discussions of qubits will always be referred to some orthonormal set of basis vectors, $|0\rangle$ and $|1\rangle$, which should be thought of as being fixed in advance. Intuitively, the states $|0\rangle$ and $|1\rangle$ are analogous to the two values 0 and 1 which a bit may take. The way a qubit differs from a bit is that *superpositions* of these two states, of the form $a|0\rangle + b|1\rangle$, can also exist, in which it is not possible to say that the qubit is definitely in the state $|0\rangle$, or definitely in the state $|1\rangle$.

We conclude with some useful terminology which is often used in connection with the description of quantum states. We say that any linear combination $\sum_i \alpha_i|\psi_i\rangle$ is a superposition of the states $|\psi_i\rangle$ with *amplitude* α_i for the state $|\psi_i\rangle$. So, for example, the state

$$\frac{|0\rangle - |1\rangle}{\sqrt{2}} \tag{2.83}$$

is a superposition of the states $|0\rangle$ and $|1\rangle$ with amplitude $1/\sqrt{2}$ for the state $|0\rangle$, and amplitude $-1/\sqrt{2}$ for the state $|1\rangle$.

2.2.2 Evolution

How does the state, $|\psi\rangle$, of a quantum mechanical system change with time? The following postulate gives a prescription for the description of such state changes.

> **Postulate 2**: The evolution of a *closed* quantum system is described by a *unitary transformation*. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2,
>
> $$|\psi'\rangle = U|\psi\rangle\,. \tag{2.84}$$

Just as quantum mechanics does not tell us the state space or quantum state of a *particular* quantum system, it does not tell us which unitary operators U describe real-world quantum dynamics. Quantum mechanics merely assures us that the evolution of any closed quantum system may be described in such a way. An obvious question to ask is: what unitary operators are natural to consider? In the case of single qubits, it turns out that *any* unitary operator at all can be realized in realistic systems.

Let's look at a few examples of unitary operators on a single qubit which are important in quantum computation and quantum information. We have already seen several examples of such unitary operators – the Pauli matrices, defined in Section 2.1.3, and the quantum gates described in Chapter 1. As remarked in Section 1.3.1, the X matrix is often known as the quantum NOT gate, by analogy to the classical NOT gate. The X and Z Pauli matrices are also sometimes referred to as the *bit flip* and *phase flip* matrices: the X matrix takes $|0\rangle$ to $|1\rangle$, and $|1\rangle$ to $|0\rangle$, thus earning the name bit flip; and the Z matrix leaves $|0\rangle$ invariant, and takes $|1\rangle$ to $-|1\rangle$, with the extra factor of -1 added known as a *phase factor*, thus justifying the term phase flip. We will not use the term phase flip for

Z very often, since it is easily confused with the phase gate to be defined in Chapter 4. (Section 2.2.7 contains more discussion of the many uses of the term 'phase'.)

Another interesting unitary operator is the *Hadamard gate*, which we denote H. This has the action $H|0\rangle \equiv (|0\rangle + |1\rangle)/\sqrt{2}$, $H|1\rangle \equiv (|0\rangle - |1\rangle)/\sqrt{2}$, and corresponding matrix representation

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{2.85}$$

Exercise 2.51: Verify that the Hadamard gate H is unitary.

Exercise 2.52: Verify that $H^2 = I$.

Exercise 2.53: What are the eigenvalues and eigenvectors of H?

Postulate 2 requires that the system being described be closed. That is, it is not interacting in any way with other systems. In reality, of course, all systems (except the Universe as a whole) interact at least somewhat with other systems. Nevertheless, there are interesting systems which can be described to a good approximation as being closed, and which are described by unitary evolution to some good approximation. Furthermore, at least in principle every open system can be described as part of a larger closed system (the Universe) which is undergoing unitary evolution. Later, we'll introduce more tools which allow us to describe systems which are not closed, but for now we'll continue with the description of the evolution of closed systems.

Postulate 2 describes how the quantum states of a closed quantum system at two different times are related. A more refined version of this postulate can be given which describes the evolution of a quantum system in *continuous time*. From this more refined postulate we will recover Postulate 2. Before we state the revised postulate, it is worth pointing out two things. First, a notational remark. The operator H appearing in the following discussion is not the same as the Hadamard operator, which we just introduced. Second, the following postulate makes use of the apparatus of differential equations. Readers with little background in the study of differential equations should be reassured that they will not be necessary for much of the book, with the exception of parts of Chapter 7, on real physical implementations of quantum information processing.

> **Postulate 2′**: The time evolution of the state of a closed quantum system is described by the *Schrödinger equation*,
>
> $$i\hbar \frac{d|\psi\rangle}{dt} = H|\psi\rangle. \tag{2.86}$$
>
> In this equation, $\hbar$ is a physical constant known as *Planck's constant* whose value must be experimentally determined. The exact value is not important to us. In practice, it is common to absorb the factor $\hbar$ into H, effectively setting $\hbar = 1$. H is a fixed Hermitian operator known as the *Hamiltonian* of the closed system.

If we know the Hamiltonian of a system, then (together with a knowledge of $\hbar$) we understand its dynamics completely, at least in principle. In general figuring out the Hamiltonian needed to describe a particular physical system is a very difficult problem – much of twentieth century physics has been concerned with this problem – which requires substantial input from experiment in order to be answered. From our point of

view this is a problem of *detail* to be addressed by physical theories built within the framework of quantum mechanics – what Hamiltonian do we need to describe atoms in such-and-such a configuration – and is not a question that needs to be addressed by the theory of quantum mechanics itself. Most of the time in our discussion of quantum computation and quantum information we won't need to discuss Hamiltonians, and when we do, we will usually just posit that some matrix is the Hamiltonian as a starting point, and proceed from there, without attempting to justify the use of that Hamiltonian.

Because the Hamiltonian is a Hermitian operator it has a spectral decomposition

$$H = \sum_E E|E\rangle\langle E|, \tag{2.87}$$

with eigenvalues E and corresponding normalized eigenvectors $|E\rangle$. The states $|E\rangle$ are conventionally referred to as *energy eigenstates*, or sometimes as *stationary states*, and E is the *energy* of the state $|E\rangle$. The lowest energy is known as the *ground state energy* for the system, and the corresponding energy eigenstate (or eigenspace) is known as the *ground state*. The reason the states $|E\rangle$ are sometimes known as stationary states is because their only change in time is to acquire an overall numerical factor,

$$|E\rangle \rightarrow \exp(-iEt/\hbar)|E\rangle. \tag{2.88}$$

As an example, suppose a single qubit has Hamiltonian

$$H = \hbar\omega X. \tag{2.89}$$

In this equation ω is a parameter that, in practice, needs to be experimentally determined. We won't worry about the parameter overly much here – the point is to give you a feel for the sort of Hamiltonians that are sometimes written down in the study of quantum computation and quantum information. The energy eigenstates of this Hamiltonian are obviously the same as the eigenstates of X, namely $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}$, with corresponding energies $\hbar\omega$ and $-\hbar\omega$. The ground state is therefore $(|0\rangle - |1\rangle)/\sqrt{2}$, and the ground state energy is $-\hbar\omega$.

What is the connection between the Hamiltonian picture of dynamics, Postulate 2′, and the unitary operator picture, Postulate 2? The answer is provided by writing down the solution to Schrödinger's equation, which is easily verified to be:

$$|\psi(t_2)\rangle = \exp\left[\frac{-iH(t_2 - t_1)}{\hbar}\right]|\psi(t_1)\rangle = U(t_1, t_2)|\psi(t_1)\rangle\,, \tag{2.90}$$

where we define

$$U(t_1, t_2) \equiv \exp\left[\frac{-iH(t_2 - t_1)}{\hbar}\right]\,. \tag{2.91}$$

You will show in the exercises that this operator is unitary, and furthermore, that any unitary operator U can be realized in the form $U = \exp(iK)$ for some Hermitian operator K. There is therefore a one-to-one correspondence between the discrete-time description of dynamics using unitary operators, and the continuous time description using Hamiltonians. For most of the book we use the unitary formulation of quantum dynamics.

Exercise 2.54: Suppose A and B are commuting Hermitian operators. Prove that $\exp(A)\exp(B) = \exp(A + B)$. (*Hint*: Use the results of Section 2.1.9.)

Exercise 2.55: Prove that $U(t_1, t_2)$ defined in Equation (2.91) is unitary.

Exercise 2.56: Use the spectral decomposition to show that $K \equiv -i \log(U)$ is Hermitian for any unitary U, and thus $U = \exp(iK)$ for some Hermitian K.

In quantum computation and quantum information we often speak of *applying* a unitary operator to a particular quantum system. For example, in the context of quantum circuits we may speak of applying the unitary gate X to a single qubit. Doesn't this contradict what we said earlier, about unitary operators describing the evolution of a *closed* quantum system? After all, if we are 'applying' a unitary operator, then that implies that there is an external 'we' who is interacting with the quantum system, and the system is not closed.

An example of this occurs when a laser is focused on an atom. After a lot of thought and hard work it is possible to write down a Hamiltonian describing the total atom–laser system. The interesting thing is that when we write down the Hamiltonian for the atom–laser system and consider the effects on the atom alone, the behavior of the state vector of the atom turns out to be almost but not quite perfectly described by another Hamiltonian, the *atomic Hamiltonian*. The atomic Hamiltonian contains terms related to laser intensity, and other parameters of the laser, which we can vary at will. It is *as if* the evolution of the atom were being described by a Hamiltonian which we can vary at will, despite the atom not being a closed system.

More generally, for many systems like this it turns out to be possible to write down a *time-varying* Hamiltonian for a quantum system, in which the Hamiltonian for the system is not a constant, but varies according to some parameters which are under an experimentalist's control, and which may be changed during the course of an experiment. The system is not, therefore, closed, but it does evolve according to Schrödinger's equation with a time-varying Hamiltonian, to some good approximation.

The upshot is that to begin we will often describe the evolution of quantum systems – even systems which aren't closed – using unitary operators. The main exception to this, quantum measurement, will be described in the next section. Later on we will investigate in more detail possible deviations from unitary evolution due to the interaction with other systems, and understand more precisely the dynamics of realistic quantum systems.

2.2.3 Quantum measurement

We postulated that closed quantum systems evolve according to unitary evolution. The evolution of systems which don't interact with the rest of the world is all very well, but there must also be times when the experimentalist and their experimental equipment – an external physical system in other words – observes the system to find out what is going on inside the system, an interaction which makes the system no longer closed, and thus not necessarily subject to unitary evolution. To explain what happens when this is done, we introduce Postulate 3, which provides a means for describing the effects of measurements on quantum systems.

> **Postulate 3**: Quantum measurements are described by a collection $\{M_m\}$ of *measurement operators*. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ immediately before the measurement then the probability that result m occurs is

given by

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle\,, \tag{2.92}$$

and the state of the system after the measurement is

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}\,. \tag{2.93}$$

The measurement operators satisfy the *completeness equation*,

$$\sum_m M_m^\dagger M_m = I\,. \tag{2.94}$$

The completeness equation expresses the fact that probabilities sum to one:

$$1 = \sum_m p(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle\,. \tag{2.95}$$

This equation being satisfied for all $|\psi\rangle$ is equivalent to the completeness equation. However, the completeness equation is much easier to check directly, so that's why it appears in the statement of the postulate.

A simple but important example of a measurement is the *measurement of a qubit in the computational basis*. This is a measurement on a single qubit with two outcomes defined by the two measurement operators $M_0 = |0\rangle\langle 0|$, $M_1 = |1\rangle\langle 1|$. Observe that each measurement operator is Hermitian, and that $M_0^2 = M_0, M_1^2 = M_1$. Thus the completeness relation is obeyed, $I = M_0^\dagger M_0 + M_1^\dagger M_1 = M_0 + M_1$. Suppose the state being measured is $|\psi\rangle = a|0\rangle + b|1\rangle$. Then the probability of obtaining measurement outcome 0 is

$$p(0) = \langle\psi|M_0^\dagger M_0|\psi\rangle = \langle\psi|M_0|\psi\rangle = |a|^2. \tag{2.96}$$

Similarly, the probability of obtaining the measurement outcome 1 is $p(1) = |b|^2$. The state after measurement in the two cases is therefore

$$\frac{M_0|\psi\rangle}{|a|} = \frac{a}{|a|}|0\rangle \tag{2.97}$$

$$\frac{M_1|\psi\rangle}{|b|} = \frac{b}{|b|}|1\rangle. \tag{2.98}$$

We will see in Section 2.2.7 that multipliers like $a/|a|$, which have modulus one, can effectively be ignored, so the two post-measurement states are effectively $|0\rangle$ and $|1\rangle$, just as described in Chapter 1.

The status of Postulate 3 as a fundamental postulate intrigues many people. Measuring devices are quantum mechanical systems, so the quantum system being measured and the measuring device together are part of a larger, isolated, quantum mechanical system. (It may be necessary to include quantum systems other than the system being measured and the measuring device to obtain a completely isolated system, but the point is that this can be done.) According to Postulate 2, the evolution of this larger isolated system can be described by a unitary evolution. Might it be possible to *derive* Postulate 3 as a consequence of this picture? Despite considerable investigation along these lines there is still disagreement between physicists about whether or not this is possible. We, however, are going to take the very pragmatic approach that in practice it is clear when to apply

Postulate 2 and when to apply Postulate 3, and not worry about deriving one postulate from the other.

Over the next few sections we apply Postulate 3 to several elementary but important measurement scenarios. Section 2.2.4 examines the problem of *distinguishing* a set of quantum states. Section 2.2.5 explains a special case of Postulate 3, the *projective* or *von Neumann* measurements. Section 2.2.6 explains another special case of Postulate 3, known as *POVM* measurements. Many introductions to quantum mechanics only discuss projective measurements, omitting a full discussion of Postulate 3 or of POVM elements. For this reason we have included Box 2.5 on page 91 which comments on the relationship between the different classes of measurement we describe.

Exercise 2.57: (Cascaded measurements are single measurements) Suppose $\{L_l\}$ and $\{M_m\}$ are two sets of measurement operators. Show that a measurement defined by the measurement operators $\{L_l\}$ followed by a measurement defined by the measurement operators $\{M_m\}$ is physically equivalent to a single measurement defined by measurement operators $\{N_{lm}\}$ with the representation $N_{lm} \equiv M_m L_l$.

2.2.4 Distinguishing quantum states

An important application of Postulate 3 is to the problem of *distinguishing quantum states*. In the classical world, distinct states of an object are usually distinguishable, at least in principle. For example, we can always identify whether a coin has landed heads or tails, at least in the ideal limit. Quantum mechanically, the situation is more complicated. In Section 1.6 we gave a plausible argument that non-orthogonal quantum states cannot be distinguished. With Postulate 3 as a firm foundation we can now give a much more convincing demonstration of this fact.

Distinguishability, like many ideas in quantum computation and quantum information, is most easily understood using the metaphor of a game involving two parties, Alice and Bob. Alice chooses a state $|\psi_i\rangle$ ($1 \leq i \leq n$) from some fixed set of states known to both parties. She gives the state $|\psi_i\rangle$ to Bob, whose task it is to identify the index i of the state Alice has given him.

Suppose the states $|\psi_i\rangle$ are orthonormal. Then Bob can do a quantum measurement to *distinguish* these states, using the following procedure. Define measurement operators $M_i \equiv |\psi_i\rangle\langle\psi_i|$, one for each possible index i, and an additional measurement operator M_0 defined as the positive square root of the positive operator $I - \sum_{i \neq 0} |\psi_i\rangle\langle\psi_i|$. These operators satisfy the completeness relation, and if the state $|\psi_i\rangle$ is prepared then $p(i) = \langle\psi_i|M_i|\psi_i\rangle = 1$, so the result i occurs with certainty. Thus, it is possible to reliably distinguish the orthonormal states $|\psi_i\rangle$.

By contrast, if the states $|\psi_i\rangle$ are not orthonormal then we can prove that there is *no quantum measurement capable of distinguishing the states*. The idea is that Bob will do a measurement described by measurement operators M_j, with outcome j. Depending on the outcome of the measurement Bob tries to guess what the index i was using some rule, $i = f(j)$, where $f(\cdot)$ represents the rule he uses to make the guess. The key to why Bob can't distinguish non-orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ is the observation that $|\psi_2\rangle$ can be decomposed into a (non-zero) component parallel to $|\psi_1\rangle$, and a component orthogonal to $|\psi_1\rangle$. Suppose j is a measurement outcome such that $f(j) = 1$, that is, Bob guesses that the state was $|\psi_1\rangle$ when he observes j. But because of the component of $|\psi_2\rangle$ parallel

to $|\psi_1\rangle$, there is a non-zero probability of getting outcome j when $|\psi_2\rangle$ is prepared, so sometimes Bob will make an error identifying which state was prepared. A more rigorous argument that non-orthogonal states can't be distinguished is given in Box 2.3, but this captures the essential idea.

Box 2.3: Proof that non-orthogonal states can't be reliably distinguished

A proof by contradiction shows that no measurement distinguishing the non-orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ is possible. Suppose such a measurement is possible. If the state $|\psi_1\rangle$ ($|\psi_2\rangle$) is prepared then the probability of measuring j such that $f(j) = 1$ ($f(j) = 2$) must be 1. Defining $E_i \equiv \sum_{j:f(j)=i} M_j^\dagger M_j$, these observations may be written as:

$$\langle\psi_1|E_1|\psi_1\rangle = 1; \quad \langle\psi_2|E_2|\psi_2\rangle = 1. \tag{2.99}$$

Since $\sum_i E_i = I$ it follows that $\sum_i \langle\psi_1|E_i|\psi_1\rangle = 1$, and since $\langle\psi_1|E_1|\psi_1\rangle = 1$ we must have $\langle\psi_1|E_2|\psi_1\rangle = 0$, and thus $\sqrt{E_2}|\psi_1\rangle = 0$. Suppose we decompose $|\psi_2\rangle = \alpha|\psi_1\rangle + \beta|\varphi\rangle$, where $|\varphi\rangle$ is orthonormal to $|\psi_1\rangle$, $|\alpha|^2 + |\beta|^2 = 1$, and $|\beta| < 1$ since $|\psi_1\rangle$ and $|\psi_2\rangle$ are not orthogonal. Then $\sqrt{E_2}|\psi_2\rangle = \beta\sqrt{E_2}|\varphi\rangle$, which implies a contradiction with (2.99), as

$$\langle\psi_2|E_2|\psi_2\rangle = |\beta|^2\langle\varphi|E_2|\varphi\rangle \leq |\beta|^2 < 1, \tag{2.100}$$

where the second last inequality follows from the observation that

$$\langle\varphi|E_2|\varphi\rangle \leq \sum_i \langle\varphi|E_i|\varphi\rangle = \langle\varphi|\varphi\rangle = 1. \tag{2.101}$$

2.2.5 Projective measurements

In this section we explain an important special case of the general measurement postulate, Postulate 3. This special class of measurements is known as *projective measurements*. For many applications of quantum computation and quantum information we will be concerned primarily with projective measurements. Indeed, projective measurements actually turn out to be *equivalent* to the general measurement postulate, when they are augmented with the ability to perform unitary transformations, as described in Postulate 2. We will explain this equivalence in detail in Section 2.2.8, as the statement of the measurement postulate for projective measurements is superficially rather different from the general postulate, Postulate 3.

Projective measurements: A projective measurement is described by an *observable*, M, a Hermitian operator on the state space of the system being observed. The observable has a spectral decomposition,

$$M = \sum_m m P_m, \tag{2.102}$$

where P_m is the projector onto the eigenspace of M with eigenvalue m. The possible outcomes of the measurement correspond to the eigenvalues, m, of the observable. Upon measuring the state $|\psi\rangle$, the probability of getting result m is

given by

$$p(m) = \langle \psi | P_m | \psi \rangle \, . \tag{2.103}$$

Given that outcome m occurred, the state of the quantum system immediately after the measurement is

$$\frac{P_m |\psi\rangle}{\sqrt{p(m)}} \, . \tag{2.104}$$

Projective measurements can be understood as a special case of Postulate 3. Suppose the measurement operators in Postulate 3, in addition to satisfying the completeness relation $\sum_m M_m^\dagger M_m = I$, also satisfy the conditions that M_m are orthogonal projectors, that is, the M_m are Hermitian, and $M_m M_{m'} = \delta_{m,m'} M_m$. With these additional restrictions, Postulate 3 reduces to a projective measurement as just defined.

Projective measurements have many nice properties. In particular, it is very easy to calculate average values for projective measurements. By definition, the average (see Appendix 1 for elementary definitions and results in probability theory) value of the measurement is

$$\begin{aligned} \mathbf{E}(M) &= \sum_m m \, p(m) && (2.110) \\ &= \sum_m m \langle \psi | P_m | \psi \rangle && (2.111) \\ &= \langle \psi | \left(\sum_m m P_m \right) | \psi \rangle && (2.112) \\ &= \langle \psi | M | \psi \rangle . && (2.113) \end{aligned}$$

This is a useful formula, which simplifies many calculations. The average value of the observable M is often written $\langle M \rangle \equiv \langle \psi | M | \psi \rangle$. From this formula for the average follows a formula for the standard deviation associated to observations of M,

$$\begin{aligned} [\Delta(M)]^2 &= \langle (M - \langle M \rangle)^2 \rangle && (2.114) \\ &= \langle M^2 \rangle - \langle M \rangle^2 . && (2.115) \end{aligned}$$

The standard deviation is a measure of the typical spread of the observed values upon measurement of M. In particular, if we perform a large number of experiments in which the state $|\psi\rangle$ is prepared and the observable M is measured, then the standard deviation $\Delta(M)$ of the observed values is determined by the formula $\Delta(M) = \sqrt{\langle M^2 \rangle - \langle M \rangle^2}$. This formulation of measurement and standard deviations in terms of observables gives rise in an elegant way to results such as the *Heisenberg uncertainty principle* (see Box 2.4).

Exercise 2.58: Suppose we prepare a quantum system in an eigenstate $|\psi\rangle$ of some observable M, with corresponding eigenvalue m. What is the average observed value of M, and the standard deviation?

Two widely used nomenclatures for measurements deserve emphasis. Rather than giving an observable to describe a projective measurement, often people simply list a complete set of orthogonal projectors P_m satisfying the relations $\sum_m P_m = I$ and $P_m P_{m'} =$

Box 2.4: The Heisenberg uncertainty principle

Perhaps the best known result of quantum mechanics is the *Heisenberg uncertainty principle*. Suppose A and B are two Hermitian operators, and $|\psi\rangle$ is a quantum state. Suppose $\langle\psi|AB|\psi\rangle = x + iy$, where x and y are real. Note that $\langle\psi|[A,B]|\psi\rangle = 2iy$ and $\langle\psi|\{A,B\}|\psi\rangle = 2x$. This implies that

$$|\langle\psi|[A,B]|\psi\rangle|^2 + |\langle\psi|\{A,B\}|\psi\rangle|^2 = 4\,|\langle\psi|AB|\psi\rangle|^2\,. \tag{2.105}$$

By the Cauchy–Schwarz inequality

$$|\langle\psi|AB|\psi\rangle|^2 \leq \langle\psi|A^2|\psi\rangle\langle\psi|B^2|\psi\rangle, \tag{2.106}$$

which combined with Equation (2.105) and dropping a non-negative term gives

$$|\langle\psi|[A,B]|\psi\rangle|^2 \leq 4\langle\psi|A^2|\psi\rangle\langle\psi|B^2|\psi\rangle. \tag{2.107}$$

Suppose C and D are two observables. Substituting $A = C - \langle C\rangle$ and $B = D - \langle D\rangle$ into the last equation, we obtain Heisenberg's uncertainty principle as it is usually stated:

$$\Delta(C)\Delta(D) \geq \frac{|\langle\psi|[C,D]|\psi\rangle|}{2}\,. \tag{2.108}$$

You should be wary of a common misconception about the uncertainty principle, that measuring an observable C to some 'accuracy' $\Delta(C)$ causes the value of D to be 'disturbed' by an amount $\Delta(D)$ in such a way that some sort of inequality similar to (2.108) is satisfied. While it is true that measurements in quantum mechanics cause disturbance to the system being measured, this is most emphatically *not* the content of the uncertainty principle.

The correct interpretation of the uncertainty principle is that if we prepare a large number of quantum systems in identical states, $|\psi\rangle$, and then perform measurements of C on some of those systems, and of D in others, then the standard deviation $\Delta(C)$ of the C results times the standard deviation $\Delta(D)$ of the results for D will satisfy the inequality (2.108).

As an example of the uncertainty principle, consider the observables X and Y when measured for the quantum state $|0\rangle$. In Equation (2.70) we showed that $[X,Y] = 2iZ$, so the uncertainty principle tells us that

$$\Delta(X)\Delta(Y) \geq \langle 0|Z|0\rangle = 1\,. \tag{2.109}$$

One elementary consequence of this is that $\Delta(X)$ and $\Delta(Y)$ must both be strictly greater than 0, as can be verified by direct calculation.

$\delta_{mm'}P_m$. The corresponding observable implicit in this usage is $M = \sum_m mP_m$. Another widely used phrase, to 'measure in a basis $|m\rangle$', where $|m\rangle$ form an orthonormal basis, simply means to perform the projective measurement with projectors $P_m = |m\rangle\langle m|$.

Let's look at an example of projective measurements on single qubits. First is the measurement of the observable Z. This has eigenvalues +1 and -1 with corresponding eigenvectors $|0\rangle$ and $|1\rangle$. Thus, for example, measurement of Z on the state $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ gives the result +1 with probability $\langle\psi|0\rangle\langle 0|\psi\rangle = 1/2$, and similarly the

result -1 with probability $1/2$. More generally, suppose $\vec{v}$ is any real three-dimensional unit vector. Then we can define an observable:

$$\vec{v} \cdot \vec{\sigma} \equiv v_1 \sigma_1 + v_2 \sigma_2 + v_3 \sigma_3. \tag{2.116}$$

Measurement of this observable is sometimes referred to as a 'measurement of spin along the $\vec{v}$ axis', for historical reasons. The following two exercises encourage you to work out some elementary but important properties of such a measurement.

Exercise 2.59: Suppose we have qubit in the state $|0\rangle$, and we measure the observable X. What is the average value of X? What is the standard deviation of X?

Exercise 2.60: Show that $\vec{v} \cdot \vec{\sigma}$ has eigenvalues ± 1, and that the projectors onto the corresponding eigenspaces are given by $P_{\pm} = (I \pm \vec{v} \cdot \vec{\sigma})/2$.

Exercise 2.61: Calculate the probability of obtaining the result $+1$ for a measurement of $\vec{v} \cdot \vec{\sigma}$, given that the state prior to measurement is $|0\rangle$. What is the state of the system after the measurement if $+1$ is obtained?

2.2.6 POVM measurements

The quantum measurement postulate, Postulate 3, involves two elements. First, it gives a rule describing the measurement statistics, that is, the respective probabilities of the different possible measurement outcomes. Second, it gives a rule describing the post-measurement state of the system. However, for some applications the post-measurement state of the system is of little interest, with the main item of interest being the probabilities of the respective measurement outcomes. This is the case, for example, in an experiment where the system is measured only once, upon conclusion of the experiment. In such instances there is a mathematical tool known as the *POVM formalism* which is especially well adapted to the analysis of the measurements. (The acronym POVM stands for 'Positive Operator-Valued Measure', a technical term whose historical origins we won't worry about.) This formalism is a simple consequence of the general description of measurements introduced in Postulate 3, but the theory of POVMs is so elegant and widely used that it merits a separate discussion here.

Suppose a measurement described by measurement operators M_m is performed upon a quantum system in the state $|\psi\rangle$. Then the probability of outcome m is given by $p(m) = \langle \psi | M_m^\dagger M_m | \psi \rangle$. Suppose we define

$$E_m \equiv M_m^\dagger M_m. \tag{2.117}$$

Then from Postulate 3 and elementary linear algebra, E_m is a positive operator such that $\sum_m E_m = I$ and $p(m) = \langle \psi | E_m | \psi \rangle$. Thus the set of operators E_m are sufficient to determine the probabilities of the different measurement outcomes. The operators E_m are known as the *POVM elements* associated with the measurement. The complete set $\{E_m\}$ is known as a *POVM*.

As an example of a POVM, consider a projective measurement described by measurement operators P_m, where the P_m are projectors such that $P_m P_{m'} = \delta_{mm'} P_m$ and $\sum_m P_m = I$. In this instance (and only this instance) all the POVM elements are the same as the measurement operators themselves, since $E_m \equiv P_m^\dagger P_m = P_m$.

Box 2.5: General measurements, projective measurements, and POVMs

Most introductions to quantum mechanics describe only projective measurements, and consequently the general description of measurements given in Postulate 3 may be unfamiliar to many physicists, as may the POVM formalism described in Section 2.2.6. The reason most physicists don't learn the general measurement formalism is because most physical systems can only be measured in a very coarse manner. In quantum computation and quantum information we aim for an exquisite level of control over the measurements that may be done, and consequently it helps to use a more comprehensive formalism for the description of measurements.

Of course, when the other axioms of quantum mechanics are taken into account, projective measurements augmented by unitary operations turn out to be completely *equivalent* to general measurements, as shown in Section 2.2.8. So a physicist trained in the use of projective measurements might ask to what end we start with the general formalism, Postulate 3? There are several reasons for doing so. First, mathematically general measurements are in some sense simpler than projective measurements, since they involve fewer restrictions on the measurement operators; there is, for example, no requirement for general measurements analogous to the condition $P_i P_j = \delta_{ij} P_i$ for projective measurements. This simpler structure also gives rise to many useful properties for general measurements that are not possessed by projective measurements. Second, it turns out that there are important problems in quantum computation and quantum information – such as the optimal way to distinguish a set of quantum states – the answer to which involves a general measurement, rather than a projective measurement.

A third reason for preferring Postulate 3 as a starting point is related to a property of projective measurements known as *repeatability*. Projective measurements are repeatable in the sense that if we perform a projective measurement once, and obtain the outcome m, repeating the measurement gives the outcome m again and does not change the state. To see this, suppose $|\psi\rangle$ was the initial state. After the first measurement the state is $|\psi_m\rangle = \left(P_m|\psi\rangle\right) / \sqrt{\langle\psi|P_m|\psi\rangle}$. Applying P_m to $|\psi_m\rangle$ does not change it, so we have $\langle\psi_m|P_m|\psi_m\rangle = 1$, and therefore repeated measurement gives the result m each time, without changing the state.

This repeatability of projective measurements tips us off to the fact that many important measurements in quantum mechanics are not projective measurements. For instance, if we use a silvered screen to measure the position of a photon we destroy the photon in the process. This certainly makes it impossible to repeat the measurement of the photon's position! Many other quantum measurements are also not repeatable in the same sense as a projective measurement. For such measurements, the general measurement postulate, Postulate 3, must be employed.

Where do POVMs fit in this picture? POVMs are best viewed as a special case of the general measurement formalism, providing the simplest means by which one can study general measurement statistics, without the necessity for knowing the post-measurement state. They are a mathematical convenience that sometimes gives extra insight into quantum measurements.

Exercise 2.62: Show that any measurement where the measurement operators and the POVM elements coincide is a projective measurement.

Above we noticed that the POVM operators are positive and satisfy $\sum_m E_m = I$. Suppose now that $\{E_m\}$ is some arbitrary set of positive operators such that $\sum_m E_m = I$. We will show that there exists a set of measurement operators M_m defining a measurement described by the POVM $\{E_m\}$. Defining $M_m \equiv \sqrt{E_m}$ we see that $\sum_m M_m^\dagger M_m = \sum_m E_m = I$, and therefore the set $\{M_m\}$ describes a measurement with POVM $\{E_m\}$. For this reason it is convenient to *define* a POVM to be any set of operators $\{E_m\}$ such that: (a) each operator E_m is positive; and (b) the *completeness relation* $\sum_m E_m = I$ is obeyed, expressing the fact that probabilities sum to one. To complete the description of POVMs, we note again that given a POVM $\{E_m\}$, the probability of outcome m is given by $p(m) = \langle\psi|E_m|\psi\rangle$.

We've looked at projective measurements as an example of the use of POVMs, but it wasn't very exciting since we didn't learn much that was new. The following more sophisticated example illustrates the use of the POVM formalism as a guide for our intuition in quantum computation and quantum information. Suppose Alice gives Bob a qubit prepared in one of two states, $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. As explained in Section 2.2.4 it is impossible for Bob to determine whether he has been given $|\psi_1\rangle$ or $|\psi_2\rangle$ with perfect reliability. However, it is possible for him to perform a measurement which distinguishes the states some of the time, but *never* makes an error of mis-identification. Consider a POVM containing three elements,

$$E_1 \equiv \frac{\sqrt{2}}{1+\sqrt{2}}|1\rangle\langle 1|, \tag{2.118}$$

$$E_2 \equiv \frac{\sqrt{2}}{1+\sqrt{2}}\frac{\left(|0\rangle - |1\rangle\right)\left(\langle 0| - \langle 1|\right)}{2}, \tag{2.119}$$

$$E_3 \equiv I - E_1 - E_2. \tag{2.120}$$

It is straightforward to verify that these are positive operators which satisfy the completeness relation $\sum_m E_m = I$, and therefore form a legitimate POVM.

Suppose Bob is given the state $|\psi_1\rangle = |0\rangle$. He performs the measurement described by the POVM $\{E_1, E_2, E_3\}$. There is zero probability that he will observe the result E_1, since E_1 has been cleverly chosen to ensure that $\langle\psi_1|E_1|\psi_1\rangle = 0$. Therefore, if the result of his measurement is E_1 then Bob can safely conclude that the state he received must have been $|\psi_2\rangle$. A similar line of reasoning shows that if the measurement outcome E_2 occurs then it must have been the state $|\psi_1\rangle$ that Bob received. Some of the time, however, Bob will obtain the measurement outcome E_3, and he can infer nothing about the identity of the state he was given. The key point, however, is that Bob *never* makes a mistake identifying the state he has been given. This infallibility comes at the price that sometimes Bob obtains no information about the identity of the state.

This simple example demonstrates the utility of the POVM formalism as a simple and intuitive way of gaining insight into quantum measurements in instances where only the measurement statistics matter. In many instances later in the book we will only be concerned with measurement statistics, and will therefore use the POVM formalism rather than the more general formalism for measurements described in Postulate 3.

Exercise 2.63: Suppose a measurement is described by measurement operators M_m.

Show that there exist unitary operators U_m such that $M_m = U_m\sqrt{E_m}$, where E_m is the POVM associated to the measurement.

Exercise 2.64: Suppose Bob is given a quantum state chosen from a set $|\psi_1\rangle, \ldots, |\psi_m\rangle$ of linearly independent states. Construct a POVM $\{E_1, E_2, \ldots, E_{m+1}\}$ such that if outcome E_i occurs, $1 \leq i \leq m$, then Bob knows with certainty that he was given the state $|\psi_i\rangle$. (The POVM must be such that $\langle\psi_i|E_i|\psi_i\rangle > 0$ for each i.)

2.2.7 Phase

'Phase' is a commonly used term in quantum mechanics, with several different meanings dependent upon context. At this point it is convenient to review a couple of these meanings. Consider, for example, the state $e^{i\theta}|\psi\rangle$, where $|\psi\rangle$ is a state vector, and θ is a real number. We say that the state $e^{i\theta}|\psi\rangle$ is equal to $|\psi\rangle$, up to the *global phase factor* $e^{i\theta}$. It is interesting to note that the *statistics of measurement* predicted for these two states are the same. To see this, suppose M_m is a measurement operator associated to some quantum measurement, and note that the respective probabilities for outcome m occurring are $\langle\psi|M_m^\dagger M_m|\psi\rangle$ and $\langle\psi|e^{-i\theta}M_m^\dagger M_m e^{i\theta}|\psi\rangle = \langle\psi|M_m^\dagger M_m|\psi\rangle$. Therefore, from an observational point of view these two states are identical. For this reason we may ignore global phase factors as being irrelevant to the observed properties of the physical system.

There is another kind of phase known as the *relative phase*, which has quite a different meaning. Consider the states

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}} \qquad \text{and} \qquad \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \tag{2.121}$$

In the first state the amplitude of $|1\rangle$ is $1/\sqrt{2}$. For the second state the amplitude is $-1/\sqrt{2}$. In each case the *magnitude* of the amplitudes is the same, but they differ in sign. More generally, we say that two amplitudes, a and b, *differ by a relative phase* if there is a real θ such that $a = \exp(i\theta)b$. More generally still, two states are said to *differ by a relative phase* in some basis if each of the amplitudes in that basis is related by such a phase factor. For example, the two states displayed above are the same up to a relative phase shift because the $|0\rangle$ amplitudes are identical (a relative phase factor of 1), and the $|1\rangle$ amplitudes differ only by a relative phase factor of -1. The difference between relative phase factors and global phase factors is that for relative phase the phase factors may vary from amplitude to amplitude. This makes the relative phase a basis-dependent concept unlike global phase. As a result, states which differ only by relative phases in some basis give rise to physically observable differences in measurement statistics, and it is not possible to regard these states as physically equivalent, as we do with states differing by a global phase factor

Exercise 2.65: Express the states $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}$ in a basis in which they are *not* the same up to a relative phase shift.

2.2.8 Composite systems

Suppose we are interested in a composite quantum system made up of two (or more) distinct physical systems. How should we describe states of the composite system? The following postulate describes how the state space of a composite system is built up from the state spaces of the component systems.

Postulate 4: The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numbered 1 through n, and system number i is prepared in the state $|\psi_i\rangle$, then the joint state of the total system is $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$.

Why is the tensor product the mathematical structure used to describe the state space of a composite physical system? At one level, we can simply accept it as a basic postulate, not reducible to something more elementary, and move on. After all, we certainly expect that there be *some canonical way* of describing composite systems in quantum mechanics. Is there some other way we can arrive at this postulate? Here is one heuristic that is sometimes used. Physicists sometimes like to speak of the *superposition principle of quantum mechanics*, which states that if $|x\rangle$ and $|y\rangle$ are two states of a quantum system, then any superposition $\alpha|x\rangle + \beta|y\rangle$ should also be an allowed state of a quantum system, where $|\alpha|^2 + |\beta|^2 = 1$. For composite systems, it seems natural that if $|A\rangle$ is a state of system A, and $|B\rangle$ is a state of system B, then there should be some corresponding state, which we might denote $|A\rangle|B\rangle$, of the joint system AB. Applying the superposition principle to product states of this form, we arrive at the tensor product postulate given above. This is not a derivation, since we are not taking the superposition principle as a fundamental part of our description of quantum mechanics, but it gives you the flavor of the various ways in which these ideas are sometimes reformulated.

A variety of different notations for composite systems appear in the literature. Part of the reason for this proliferation is that different notations are better adapted for different applications, and we will also find it convenient to introduce some specialized notations on occasion. At this point it suffices to mention a useful subscript notation to denote states and operators on different systems, when it is not clear from context. For example, in a system containing three qubits, X_2 is the Pauli σ_x operator acting on the second qubit.

Exercise 2.66: Show that the average value of the observable $X_1 Z_2$ for a two qubit system measured in the state $(|00\rangle + |11\rangle)/\sqrt{2}$ is zero.

In Section 2.2.5 we claimed that projective measurements together with unitary dynamics are sufficient to implement a general measurement. The proof of this statement makes use of composite quantum systems, and is a nice illustration of Postulate 4 in action. Suppose we have a quantum system with state space Q, and we want to perform a measurement described by measurement operators M_m on the system Q. To do this, we introduce an *ancilla system*, with state space M, having an orthonormal basis $|m\rangle$ in one-to-one correspondence with the possible outcomes of the measurement we wish to implement. This ancilla system can be regarded as merely a mathematical device appearing in the construction, or it can be interpreted physically as an extra quantum system introduced into the problem, which we assume has a state space with the required properties.

Letting $|0\rangle$ be any fixed state of M, define an operator U on products $|\psi\rangle|0\rangle$ of states $|\psi\rangle$ from Q with the state $|0\rangle$ by

$$U|\psi\rangle|0\rangle \equiv \sum_m M_m|\psi\rangle|m\rangle. \tag{2.122}$$

Using the orthonormality of the states $|m\rangle$ and the completeness relation $\sum_m M_m^\dagger M_m =$

I, we can see that U preserves inner products between states of the form $|\psi\rangle|0\rangle$,

$$\langle\varphi|\langle 0|U^\dagger U|\psi\rangle|0\rangle = \sum_{m,m'} \langle\varphi|M_m^\dagger M_{m'}|\psi\rangle \langle m|m'\rangle \tag{2.123}$$

$$= \sum_m \langle\varphi|M_m^\dagger M_m|\psi\rangle \tag{2.124}$$

$$= \langle\varphi|\psi\rangle. \tag{2.125}$$

By the results of Exercise 2.67 it follows that U can be extended to a unitary operator on the space $Q \otimes M$, which we also denote by U.

Exercise 2.67: Suppose V is a Hilbert space with a subspace W. Suppose $U : W \rightarrow V$ is a linear operator which preserves inner products, that is, for any $|w_1\rangle$ and $|w_2\rangle$ in W,

$$\langle w_1|U^\dagger U|w_2\rangle = \langle w_1|w_2\rangle. \tag{2.126}$$

Prove that there exists a unitary operator $U' : V \rightarrow V$ which *extends* U. That is, $U'|w\rangle = U|w\rangle$ for all $|w\rangle$ in W, but U' is defined on the entire space V. Usually we omit the prime symbol $'$ and just write U to denote the extension.

Next, suppose we perform a projective measurement on the two systems described by projectors $P_m \equiv I_Q \otimes |m\rangle\langle m|$. Outcome m occurs with probability

$$p(m) = \langle\psi|\langle 0|U^\dagger P_m U|\psi\rangle|0\rangle \tag{2.127}$$

$$= \sum_{m',m''} \langle\psi|M_{m'}^\dagger\langle m'|(I_Q \otimes |m\rangle\langle m|)M_{m''}|\psi\rangle|m''\rangle \tag{2.128}$$

$$= \langle\psi|M_m^\dagger M_m|\psi\rangle, \tag{2.129}$$

just as given in Postulate 3. The joint state of the system QM after measurement, conditional on result m occurring, is given by

$$\frac{P_m U|\psi\rangle|0\rangle}{\sqrt{\langle\psi|U^\dagger P_m U|\psi\rangle}} = \frac{M_m|\psi\rangle|m\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{2.130}$$

It follows that the state of system M after the measurement is $|m\rangle$, and the state of system Q is

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}, \tag{2.131}$$

just as prescribed by Postulate 3. Thus unitary dynamics, projective measurements, and the ability to introduce ancillary systems, together allow any measurement of the form described in Postulate 3 to be realized.

Postulate 4 also enables us to define one of the most interesting and puzzling ideas associated with composite quantum systems – *entanglement*. Consider the two qubit state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \tag{2.132}$$

This state has the remarkable property that there are no single qubit states $|a\rangle$ and $|b\rangle$ such that $|\psi\rangle = |a\rangle|b\rangle$, a fact which you should now convince yourself of:

Exercise 2.68: Prove that $|\psi\rangle \neq |a\rangle|b\rangle$ for all single qubit states $|a\rangle$ and $|b\rangle$.

We say that a state of a composite system having this property (that it can't be written as a product of states of its component systems) is an *entangled* state. For reasons which nobody fully understands, entangled states play a crucial role in quantum computation and quantum information, and arise repeatedly through the remainder of this book. We have already seen entanglement play a crucial role in quantum teleportation, as described in Section 1.3.7. In this chapter we give two examples of the strange effects enabled by entangled quantum states, superdense coding (Section 2.3), and the violation of Bell's inequality (Section 2.6).

2.2.9 Quantum mechanics: a global view

We have now explained *all* the fundamental postulates of quantum mechanics. Most of the rest of the book is taken up with deriving consequences of these postulates. Let's quickly review the postulates and try to place them in some kind of global perspective.

Postulate 1 sets the arena for quantum mechanics, by specifying how the state of an isolated quantum system is to be described. Postulate 2 tells us that the dynamics of *closed* quantum systems are described by the Schrödinger equation, and thus by unitary evolution. Postulate 3 tells us how to extract information from our quantum systems by giving a prescription for the description of measurement. Postulate 4 tells us how the state spaces of different quantum systems may be combined to give a description of the composite system.

What's odd about quantum mechanics, at least by our classical lights, is that we can't directly observe the state vector. It's a little bit like a game of chess where you can never find out exactly where each piece is, but only know the rank of the board they are on. Classical physics – and our intuition – tells us that the fundamental properties of an object, like energy, position, and velocity, are directly accessible to observation. In quantum mechanics these quantities no longer appear as fundamental, being replaced by the state vector, which can't be directly observed. It is as though there is a *hidden world* in quantum mechanics, which we can only indirectly and imperfectly access. Moreover, merely observing a classical system does not necessarily change the state of the system. Imagine how difficult it would be to play tennis if each time you looked at the ball its position changed! But according to Postulate 3, observation in quantum mechanics is an invasive procedure that typically changes the state of the system.

What conclusions should we draw from these strange features of quantum mechanics? Might it be possible to reformulate quantum mechanics in a mathematically equivalent way so that it had a structure more like classical physics? In Section 2.6 we'll prove *Bell's inequality*, a surprising result that shows any attempt at such a reformulation is doomed to failure. We're stuck with the counter-intuitive nature of quantum mechanics. Of course, the proper reaction to this is glee, not sorrow! It gives us an opportunity to develop tools of thought that make quantum mechanics intuitive. Moreover, we can exploit the hidden nature of the state vector to do information processing tasks beyond what is possible in the classical world. Without this counter-intuitive behavior, quantum computation and quantum information would be a lot less interesting.

We can also turn this discussion about, and ask ourselves: 'If quantum mechanics is so different from classical physics, then how come the everyday world looks so classical?' Why do we see no evidence of a hidden state vector in our everyday lives? It turns out

that the classical world we see can be *derived* from quantum mechanics as an approximate description of the world that will be valid on the sort of time, length and mass scales we commonly encounter in our everyday lives. Explaining the details of how quantum mechanics gives rise to classical physics is beyond the scope of this book, but the interested reader should check out the discussion of this topic in 'History and further reading'at the end of Chapter 8.