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, $A^\dagger = (A^T)^*$. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^\dagger = \begin{bmatrix} a^* & c^* \\ b^* & d^* \end{bmatrix}.$
$\langle\varphi\|A\|\psi\rangle$	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 $A : V \to W$ 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 $\langle i|j\rangle = \delta_{ij}$,

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

$$= \left[v_1^* \dots v_n^*\right] \left[\begin{array}{c} w_1 \\ \vdots \\ w_n \end{array}\right]. \tag{2.19}$$

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 $|w\rangle$,

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

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 $Q^2 = Q$,

$$\begin{aligned} QMQ\,QM^\dagger Q &= QMQM^\dagger Q && (2.37)\\ &= QMM^\dagger Q && (2.38)\\ &= QM^\dagger MQ && (2.39)\\ &= QM^\dagger QMQ && (2.40)\\ &= QM^\dagger Q\,QMQ\,, && (2.41)\end{aligned}$$

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:

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

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

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \otimes \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \times 2 \\ 1 \times 3 \\ 2 \times 2 \\ 2 \times 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 6 \end{bmatrix}. \tag{2.51}$$

The tensor product of the Pauli matrices X and Y is

$$X \otimes Y = \begin{bmatrix} 0 \cdot Y & 1 \cdot Y \\ 1 \cdot Y & 0 \cdot Y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix}. \tag{2.52}$$

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* $A \to UAU^\dagger$, 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

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

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

This result, that $\mathrm{tr}(A|\psi\rangle\langle\psi|) = \langle\psi|A|\psi\rangle$ 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

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

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

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

and if z is an arbitrary complex number show that

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

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

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

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 $B_a \equiv P_aBP_a$. 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 $B|a, b, k\rangle = P_aB|a, b, k\rangle$. Moreover we have $P_a|a, b, k\rangle = |a, b, k\rangle$, so

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

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

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