5

Linear vector spaces

Linear vector space is to quantum mechanics what calculus is to classical mechanics. In this chapter the essential ideas of linear vector spaces will be discussed. The reader is already familiar with vector calculus in three-dimensional Euclidean space E_3 (Chapter 1). We therefore present our discussion as a generalization of elementary vector calculus. The presentation will be, however, slightly abstract and more formal than the discussion of vectors in Chapter 1. Any reader who is not already familiar with this sort of discussion should be patient with the first few sections. You will then be amply repaid by finding the rest of this chapter relatively easy reading.

Euclidean n-space E_n

In the study of vector analysis in E_3, an ordered triple of numbers (a_1, a_2, a_3) has two different geometric interpretations. It represents a point in space, with a_1, a_2, a_3 being its coordinates; it also represents a vector, with a_1, a_2, and a_3 being its components along the three coordinate axes (Fig. 5.1). This idea of using triples of numbers to locate points in three-dimensional space was first introduced in the mid-seventeenth century. By the latter part of the nineteenth century physicists and mathematicians began to use the quadruples of numbers (a_1, a_2, a_3, a_4) as points in four-dimensional space, quintuples $(a_1, a_2, a_3, a_4, a_5)$ as points in five-dimensional space etc. We now extend this to n-dimensional space E_n, where n is a positive integer. Although our geometric visualization doesn't extend beyond three-dimensional space, we can extend many familiar ideas beyond three-dimensional space by working with analytic or numerical properties of points and vectors rather than their geometric properties.

For two- or three-dimensional space, we use the terms 'ordered pair' and 'ordered triple.' When $n > 3$, we use the term 'ordered-n-tuplet' for a sequence

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

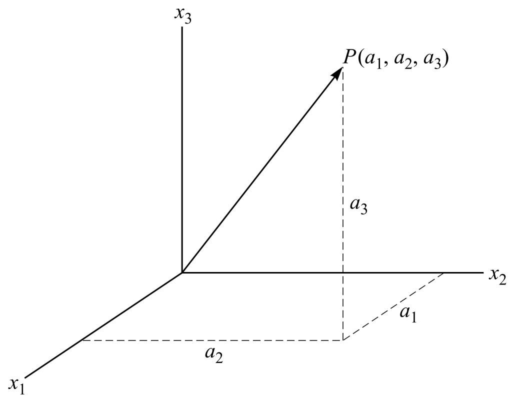

Figure 5.1. A space point P whose position vector is **A**.

of n numbers, real or complex, $(a_1, a_2, a_3, \ldots, a_n)$; they will be viewed either as a generalized point or a generalized vector in a n-dimensional space E_n.

Two vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in E_n are called equal if

$$u_i = v_i, \quad i = 1, 2, \ldots, n \tag{5.1}$$

The sum $\mathbf{u} + \mathbf{v}$ is defined by

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n) \tag{5.2}$$

and if k is any scalar, the scalar multiple $k\mathbf{u}$ is defined by

$$k\mathbf{u} = (ku_1, ku_2, \ldots, ku_n). \tag{5.3}$$

If $u = (u_1, u_2, \ldots, u_n)$ is any vector in E_n, its negative is given by

$$-\mathbf{u} = (-u_1, -u_2, \ldots, -u_n) \tag{5.4}$$

and the subtraction of vectors in E_n can be considered as addition: $\mathbf{v} - \mathbf{u} = \mathbf{v} + (-\mathbf{u})$. The null (zero) vector in E_n is defined to be the vector $0 = (0, 0, \ldots, 0)$.

The addition and scalar multiplication of vectors in E_n have the following arithmetic properties:

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}, \tag{5.5a}$$

$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}, \tag{5.5b}$$

$$\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}, \tag{5.5c}$$

$$a(b\mathbf{u}) = (ab)\mathbf{u}, \tag{5.5d}$$

$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}, \tag{5.5e}$$

$$(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}, \tag{5.5f}$$

where **u**, **v**, **w** are vectors in E_n and a and b are scalars.

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

We usually define the inner product of two vectors in E_3 in terms of lengths of the vectors and the angle between the vectors: $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta, \theta = \angle(\mathbf{A}, \mathbf{B})$. We do not define the inner product in E_n in the same manner. However, the inner product in E_3 has a second equivalent expression in terms of components: $\mathbf{A}\cdot\mathbf{B} = A_1B_1 + A_2B_2 + A_3B_3$. We choose to define a similar formula for the general case. We made this choice because of the further generalization that will be outlined in the next section. Thus, for any two vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in E_n, the inner (or dot) product $\mathbf{u}\cdot\mathbf{v}$ is defined by

$$\mathbf{u}\cdot\mathbf{v} = u_1^*v_1 + u_2^*v_2 + \cdots + u_n^*v_n \tag{5.6}$$

where the asterisk denotes complex conjugation. $\mathbf{u}$ is often called the prefactor and $\mathbf{v}$ the post-factor. The inner product is linear with respect to the post-factor, and anti-linear with respect to the prefactor:

$$\mathbf{u}\cdot(a\mathbf{v} + b\mathbf{w}) = a\mathbf{u}\cdot\mathbf{v} + b\mathbf{u}\cdot\mathbf{w}, \quad (a\mathbf{u} + b\mathbf{v})\cdot\mathbf{w} = a^*(\mathbf{u}\cdot\mathbf{v}) + b^*(\mathbf{u}\cdot\mathbf{w}).$$

We expect the inner product for the general case also to have the following three main features:

$$\mathbf{u}\cdot\mathbf{v} = (\mathbf{v}\cdot\mathbf{u})^* \tag{5.7a}$$

$$\mathbf{u}\cdot(a\mathbf{v} + b\mathbf{w}) = a\mathbf{u}\cdot\mathbf{v} + b\mathbf{u}\cdot\mathbf{w} \tag{5.7b}$$

$$\mathbf{u}\cdot\mathbf{u} \geq 0 \ (= 0, \text{ if and only if } \mathbf{u} = 0). \tag{5.7c}$$

Many of the familiar ideas from E_2 and E_3 have been carried over, so it is common to refer to E_n with the operations of addition, scalar multiplication, and with the inner product that we have defined here as Euclidean n-space.

General linear vector spaces

We now generalize the concept of vector space still further: a set of 'objects' (or elements) obeying a set of axioms, which will be chosen by abstracting the most important properties of vectors in E_n, forms a linear vector space V_n with the objects called vectors. Before introducing the requisite axioms, we first adapt a notation for our general vectors: general vectors are designated by the symbol $|\ \rangle$, which we call, following Dirac, ket vectors; the conjugates of ket vectors are denoted by the symbol $\langle\ |$, the bra vectors. However, for simplicity, we shall refer in the future to the ket vectors $|\ \rangle$ simply as vectors, and to the $\langle\ |$s as conjugate vectors. We now proceed to define two basic operations on these vectors: addition and multiplication by scalars.

By addition we mean a rule for forming the sum, denoted $|\psi_1\rangle + |\psi_2\rangle$, for any pair of vectors $|\psi_1\rangle$ and $|\psi_2\rangle$.

By scalar multiplication we mean a rule for associating with each scalar k and each vector $|\psi\rangle$ a new vector $k|\psi\rangle$.

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

We now proceed to generalize the concept of a vector space. An arbitrary set of n objects $|1\rangle, |2\rangle, |3\rangle, \ldots, |\phi\rangle, \ldots, |\varphi\rangle$ form a linear vector V_n if these objects, called vectors, meet the following axioms or properties:

A.1 If $|\phi\rangle$ and $|\varphi\rangle$ are objects in V_n and k is a scalar, then $|\phi\rangle + |\varphi\rangle$ and $k|\phi\rangle$ are in V_n, a feature called closure.
A.2 $|\phi\rangle + |\varphi\rangle = |\varphi\rangle + |\phi\rangle$; that is, addition is commutative.
A.3 $(|\phi\rangle + |\varphi\rangle) + |\psi\rangle = |\phi\rangle + (|\varphi\rangle + |\psi\rangle)$; that is, addition is associative.
A.4 $k(|\phi\rangle + |\varphi\rangle) = k|\phi\rangle + k|\varphi\rangle$; that is, scalar multiplication is distributive in the vectors.
A.5 $(k + \alpha)|\phi\rangle = k|\phi\rangle + \alpha|\phi\rangle$; that is, scalar multiplication is distributive in the scalars.
A.6 $k(\alpha|\phi\rangle) = k\alpha|\phi\rangle$; that is, scalar multiplication is associative.
A.7 There exists a null vector $|0\rangle$ in V_n such that $|\phi\rangle + |0\rangle = |\phi\rangle$ for all $|\phi\rangle$ in V_n.
A.8 For every vector $|\phi\rangle$ in V_n, there exists an inverse under addition, $|-\phi\rangle$ such that $|\phi\rangle + |-\phi\rangle = |0\rangle$.

The set of numbers $a, b, \ldots$ used in scalar multiplication of vectors is called the field over which the vector field is defined. If the field consists of real numbers, we have a real vector field; if they are complex, we have a complex field. *Note that the vectors themselves are neither real nor complex, the nature of the vectors is not specified. Vectors can be any kinds of objects; all that is required is that the vector space axioms be satisfied.* Thus we purposely do not use the symbol **V** to denote the vectors as the first step to turn the reader away from the limited concept of the vector as a directed line segment. Instead, we use Dirac's ket and bra symbols, $|\ \rangle$ and $\langle\ |$, to denote generic vectors.

The familiar three-dimensional space of position vectors E_3 is an example of a vector space over the field of real numbers. Let us now examine two simple examples.

Example 5.1
Let V be any plane through the origin in E_3. We wish to show that the points in the plane V form a vector space under the addition and scalar multiplication operations for vector in E_3.

Solution: Since E_3 itself is a vector space under the addition and scalar multiplication operations, thus Axioms A.2, A.3, A.4, A.5, and A.6 hold for all points in E_3 and consequently for all points in the plane V. We therefore need only show that Axioms A.1, A.7, and A.8 are satisfied.

Now the plane V, passing through the origin, has an equation of the form

$$ax_1 + bx_2 + cx_3 = 0.$$

Hence, if $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ are points in V, then we have

$$au_1 + bu_2 + cu_3 = 0 \quad \text{and} \quad av_1 + bv_2 + cv_3 = 0.$$

Addition gives

$$a(u_1 + v_1) + b(u_2 + v_2) + c(u_3 + v_3) = 0,$$

which shows that the point $\mathbf{u} + \mathbf{v}$ also lies in the plane V. This proves that Axiom A.1 is satisfied. Multiplying $au_1 + bu_2 + cu_3 = 0$ through by -1 gives

$$a(-u_1) + b(-u_2) + c(-u_3) = 0,$$

that is, the point $-\mathbf{u} = (-u_1, -u_2, -u_3)$ lies in V. This establishes Axiom A.8. The verification of Axiom A.7 is left as an exercise.

Example 5.2
Let V be the set of all $m \times n$ matrices with real elements. We know how to add matrices and multiply matrices by scalars. The corresponding rules obey closure, associativity and distributive requirements. The null matrix has all zeros in it, and the inverse under matrix addition is the matrix with all elements negated. Thus the set of all $m \times n$ matrices, together with the operations of matrix addition and scalar multiplication, is a vector space. We shall denote this vector space by the symbol M_{mn}.

Subspaces

Consider a vector space V. If W is a subset of V and forms a vector space under the addition and scalar multiplication, then W is called a subspace of V. For example, lines and planes passing through the origin form vector spaces and they are subspaces of E_3.

Example 5.3
We can show that the set of all 2×2 matrices having zero on the main diagonal is a subspace of the vector space M_{22} of all 2×2 matrices.

Solution: To prove this, let

$$\tilde{X} = \begin{pmatrix} 0 & x_{12} \\ x_{21} & 0 \end{pmatrix} \qquad \tilde{Y} = \begin{pmatrix} 0 & y_{12} \\ y_{21} & 0 \end{pmatrix}$$

be two matrices in W and k any scalar. Then

$$k\tilde{X} = \begin{pmatrix} 0 & x_{12} \\ kx_{21} & 0 \end{pmatrix} \quad \text{and} \quad \tilde{X} + \tilde{Y} = \begin{pmatrix} 0 & x_{12} + y_{12} \\ x_{21} + y_{21} & 0 \end{pmatrix}$$

and thus they lie in W. We leave the verification of other axioms as exercises.

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

Linear combination

A vector $|W\rangle$ is a linear combination of the vectors $|v_1\rangle$, $|v_2\rangle, \ldots, |v_r\rangle$ if it can be expressed in the form

$$|W\rangle = k_1|v_1\rangle + k_2|v_2\rangle + \cdots + k_r|v_r\rangle,$$

where $k_1, k_2, \ldots, k_r$ are scalars. For example, it is easy to show that the vector $|W\rangle = (9, 2, 7)$ in E_3 is a linear combination of $|v_1\rangle = (1, 2, -1)$ and $|v_2\rangle = (6, 4, 2)$. To see this, let us write

$$(9, 2, 7) = k_1(1, 2, -1) + k_2(6, 4, 2)$$

or

$$(9, 2, 7) = (k_1 + 6k_2, 2k_1 + 4k_2, -k_1 + 2k_2).$$

Equating corresponding components gives

$$k_1 + 6k_2 = 9, \quad 2k_1 + 4k_2 = 2, \quad -k_1 + 2k_2 = 7.$$

Solving this system yields $k_1 = -3$ and $k_2 = 2$ so that

$$|W\rangle = -3|v_1\rangle + 2|v_2\rangle.$$

Linear independence, bases, and dimensionality

Consider a set of vectors $|1\rangle, |2\rangle, \ldots, |r\rangle, \ldots |n\rangle$ in a linear vector space V. If every vector in V is expressible as a linear combination of $|1\rangle, |2\rangle, \ldots, |r\rangle, \ldots, |n\rangle$, then we say that these vectors *span* the vector space V, and they are called the *base vectors* or *basis* of the vector space V. For example, the three unit vectors $e_1 = (1, 0, 0), e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$ span E_3 because every vector in E_3 is expressible as a linear combination of e_1, e_2, and e_3. But the following three vectors in E_3 do not span E_3: $|1\rangle = (1, 1, 2), |2\rangle = (1, 0, 1)$, and $|3\rangle = (2, 1, 3)$.

Base vectors are very useful in a variety of problems since it is often possible to study a vector space by first studying the vectors in a base set, then extending the results to the rest of the vector space. Therefore it is desirable to keep the spanning set as small as possible. Finding the spanning sets for a vector space depends upon the notion of linear independence.

We say that a finite set of n vectors $|1\rangle, |2\rangle, \ldots, |r\rangle, \ldots, |n\rangle$, none of which is a null vector, is linearly independent if no set of non-zero numbers a_k exists such that

$$\sum_{k=1}^{n} a_k|k\rangle = |0\rangle. \tag{5.8}$$

In other words, the set of vectors is linearly independent if it is impossible to construct the null vector from a linear combination of the vectors except when all the coefficients vanish. For example, non-zero vectors $|1\rangle$ and $|2\rangle$ of E_2 that lie along the same coordinate axis, say x_1, are not linearly independent, since we can write one as a multiple of the other: $|1\rangle = a|2\rangle$, where a is a scalar which may be positive or negative. That is, $|1\rangle$ and $|2\rangle$ depend on each other and so they are not linearly independent. Now let us move the term $a|2\rangle$ to the left hand side and the result is the null vector: $|1\rangle - a|2\rangle = |0\rangle$. Thus, for these two vectors $|1\rangle$ and $|2\rangle$ in E_2, we can find two non-zero numbers $(1, -a)$ such that Eq. (5.8) is satisfied, and so they are not linearly independent.

On the other hand, the n vectors $|1\rangle, |2\rangle, \ldots, |r\rangle, \ldots, |n\rangle$ are linearly dependent if it is possible to find scalars $a_1, a_2, \ldots, a_n$, at least two of which are non-zero, such that Eq. (5.8) is satisfied. Let us say $a_9 \neq 0$. Then we could express $|9\rangle$ in terms of the other vectors

$$|9\rangle = \sum_{i=1,\neq 9}^{n} \frac{-a_i}{a_9}|i\rangle.$$

That is, the n vectors in the set are linearly dependent if any one of them can be expressed as a linear combination of the remaining $n - 1$ vectors.

Example 5.4

The set of three vectors $|1\rangle = (2, -1, 0, 3)$, $|2\rangle = (1, 2, 5, -1)$, $|3\rangle = (7, -1, 5, 8)$ is linearly dependent, since $3|1\rangle + |2\rangle - |3\rangle = |0\rangle$.

Example 5.5

The set of three unit vectors $|e_1\rangle = (1, 0, 0)$, $|e_2\rangle = (0, 1, 0)$, and $|e_3\rangle = (0, 0, 1)$ in E_3 is linearly independent. To see this, let us start with Eq. (5.8) which now takes the form

$$a_1|e_1\rangle + a_2|e_2\rangle + a_3|e_3\rangle = |0\rangle$$

or

$$a_1(1, 0, 0) + a_2(0, 1, 0) + a_3(0, 0, 1) = (0, 0, 0)$$

from which we obtain

$$(a_1, a_2, a_3) = (0, 0, 0);$$

the set of three unit vectors $|e_1\rangle, |e_2\rangle$, and $|e_3\rangle$ is therefore linearly independent.

Example 5.6

The set S of the following four matrices

$$|1\rangle = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad |3\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad |4\rangle = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

is a basis for the vector space M_{22} of 2×2 matrices. To see that S spans M_{22}, note that a typical 2×2 vector (matrix) can be written as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$
$$= a|1\rangle + b|2\rangle + c|3\rangle + d|4\rangle.$$

To see that S is linearly independent, assume that

$$a|1\rangle + b|2\rangle + c|3\rangle + d|4\rangle = |0\rangle,$$

that is,

$$a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

from which we find $a = b = c = d = 0$ so that S is linearly independent.

We now come to the dimensionality of a vector space. We think of space around us as three-dimensional. How do we extend the notion of dimension to a linear vector space? Recall that the three-dimensional Euclidean space E_3 is spanned by the three base vectors: $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. Similarly, the dimension n of a vector space V is defined to be the number n of linearly independent base vectors that span the vector space V. The vector space will be denoted by $V_n(R)$ if the field is real and by $V_n(C)$ if the field is complex. For example, as shown in Example 5.6, 2×2 matrices form a four-dimensional vector space whose base vectors are

$$|1\rangle = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad |3\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad |4\rangle = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix};$$

since any arbitrary 2×2 matrix can be written in terms of these:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a|1\rangle + b|2\rangle + c|3\rangle + d|4\rangle.$$

If the scalars a, b, c, d are real, we have a real four-dimensional space, if they are complex we have a complex four-dimensional space.

Inner product spaces (unitary spaces)

In this section the structure of the vector space will be greatly enriched by the addition of a numerical function, the inner product (or scalar product). Linear vector spaces in which an inner product is defined are called inner-product spaces (or unitary spaces). The study of inner-product spaces enables us to make a real juncture with physics.

In our earlier discussion, the inner product of two vectors in E_n was defined by Eq. (5.6), a generalization of the inner product of two vectors in E_3. In a general linear vector space, an inner product is defined axiomatically analogously with the inner product on E_n. Thus given two vectors $|U\rangle$ and $|W\rangle$

$$|U\rangle = \sum_{i=1}^{n} u_i |i\rangle, \quad |W\rangle = \sum_{i=1}^{n} w_i |i\rangle, \tag{5.9}$$

where $|U\rangle$ and $|W\rangle$ are expressed in terms of the n base vectors $|i\rangle$, the inner product, denoted by the symbol $\langle U|W\rangle$, is defined to be

$$\langle U|W\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} u_i^* w_j \langle i|j\rangle. \tag{5.10}$$

$\langle U|$ is often called the pre-factor and $|W\rangle$ the post-factor. The inner product obeys the following rules (or axioms):

B.1 $\langle U|W\rangle = \langle W|U\rangle^*$ (skew-symmetry);
B.2 $\langle U|U\rangle \geq 0, \quad = 0$ if and only if $|U\rangle = |0\rangle$ (positive semidefiniteness);
B.3 $\langle U|X\rangle + |W\rangle) = \langle U|X\rangle + \langle U|W\rangle$ (additivity);
B.4 $\langle aU|W\rangle = a^*\langle U|W\rangle, \quad \langle U|bW\rangle = b\langle U|W\rangle$ (homogeneity);

where a and b are scalars and the asterisk (*) denotes complex conjugation. Note that Axiom B.1 is different from the one for the inner product on E_3: the inner product on a general linear vector space depends on the order of the two factors for a complex vector space. In a real vector space E_3, the complex conjugation in Axioms B.1 and B.4 adds nothing and may be ignored. In either case, real or complex, Axiom B.1 implies that $\langle U|U\rangle$ is real, so the inequality in Axiom B.2 makes sense.

The inner product is linear with respect to the post-factor:

$$\langle U|aW + bX\rangle = a\langle U|W\rangle + b\langle U|X\rangle,$$

and anti-linear with respect to the prefactor,

$$\langle aU + bX|W\rangle = a^*\langle U|W\rangle + b^*\langle X|W\rangle.$$

Two vectors are said to be orthogonal if their inner product vanishes. And we will refer to the quantity $\langle U|U\rangle^{1/2} = \| U \|$ as the norm or length of the vector. A normalized vector, having unit norm, is a unit vector. Any given non-zero vector may be normalized by dividing it by its length. An orthonormal basis is a set of basis vectors that are all of unit norm and pair-wise orthogonal. It is very handy to have an orthonormal set of vectors as a basis for a vector space, so for $\langle i|j\rangle$ in Eq. (5.10) we shall assume

$$\langle i|j\rangle = \delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases},$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

then Eq. (5.10) reduces to

$$\langle U|W\rangle = \sum_i \sum_j u_i^* w_j \delta_{ij} = \sum_i u_i^* \left(\sum_j w_j \delta_{ij} \right) = \sum_i u_i^* w_i. \tag{5.11}$$

Note that Axiom B.2 implies that if a vector $|U\rangle$ is orthogonal to every vector of the vector space, then $|U\rangle = 0$: since $\langle U|\ \rangle = 0$ for all $|\ \rangle$ belongs to the vector space, so we have in particular $\langle U|U\rangle = 0$.

We will show shortly that we may construct an orthonormal basis from an arbitrary basis using a technique known as the Gram–Schmidt orthogonalization process.

Example 5.7

Let $|U\rangle = (3-4i)|1\rangle + (5-6i)|2\rangle$ and $|W\rangle = (1-i)|1\rangle + (2-3i)|2\rangle$ be two vectors expanded in terms of an orthonormal basis $|1\rangle$ and $|2\rangle$. Then we have, using Eq. (5.10):

$$\begin{aligned}
\langle U|U\rangle &= (3+4i)(3-4i) + (5+6i)(5-6i) = 86,\\
\langle W|W\rangle &= (1+i)(1-i) + (2+3i)(2-3i) = 15,\\
\langle U|W\rangle &= (3+4i)(1-i) + (5+6i)(2-3i) = 35-2i = \langle W|U\rangle^*.
\end{aligned}$$

Example 5.8

If $\tilde{A}$ and $\tilde{B}$ are two matrices, where

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \tilde{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},$$

then the following formula defines an inner product on M_{22}:

$$\langle \tilde{A}|\tilde{B}\rangle = a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22}.$$

To see this, let us first expand $\tilde{A}$ and $\tilde{B}$ in terms of the following base vectors

$$|1\rangle = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad |3\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad |4\rangle = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix};$$

$$\tilde{A} = a_{11}|1\rangle + a_{12}|2\rangle + a_{21}|3\rangle + a_{22}|4\rangle, \quad \tilde{B} = b_{11}|1\rangle + b_{12}|2\rangle + b_{21}|3\rangle + b_{22}|4\rangle.$$

The result follows easily from the defining formula (5.10).

Example 5.9

Consider the vector $|U\rangle$, in a certain orthonormal basis, with components

$$|U\rangle = \begin{pmatrix} 1+i \\ \sqrt{3}+i \end{pmatrix}, \qquad i = \sqrt{-1}.$$

We now expand it in a new orthonormal basis $|e_1\rangle, |e_2\rangle$ with components

$$|e_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \qquad |e_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

To do this, let us write

$$|U\rangle = u_1|e_1\rangle + u_2|e_2\rangle$$

and determine u_1 and u_2. To determine u_1, we take the inner product of both sides with $\langle e_1|$:

$$u_1 = \langle e_1|U\rangle = \frac{1}{\sqrt{2}}(1 \quad 1)\begin{pmatrix}1+i\\ \sqrt{3}+i\end{pmatrix} = \frac{1}{\sqrt{2}}(1+\sqrt{3}+2i);$$

likewise,

$$u_2 = \frac{1}{\sqrt{2}}(1-\sqrt{3}).$$

As a check on the calculation, let us compute the norm squared of the vector and see if it equals $|1+i|^2 + |\sqrt{3}+i|^2 = 6$. We find

$$|u_1|^2 + |u_2|^2 = \frac{1}{2}(1+3+2\sqrt{3}+4+1+3-2\sqrt{3}) = 6.$$

The Gram–Schmidt orthogonalization process

We now take up the Gram–Schmidt orthogonalization method for converting a linearly independent basis into an orthonormal one. The basic idea can be clearly illustrated in the following steps. Let $|1\rangle, |2\rangle, \ldots, |i\rangle, \ldots$ be a linearly independent basis. To get an orthonormal basis out of these, we do the following:

Step 1. Rescale the first vector by its own length, so it becomes a unit vector. This will be the first basis vector.

$$|e_1\rangle = \frac{|1\rangle}{||1\rangle|},$$

where $||1\rangle| = \sqrt{\langle 1 \mid 1\rangle}$. Clearly

$$\langle e_1 \mid e_1\rangle = \frac{\langle 1 \mid 1\rangle}{||1\rangle|} = 1.$$

Step 2. To construct the second member of the orthonormal basis, we subtract from the second vector $|2\rangle$ its projection along the first, leaving behind only the part perpendicular to the first.

$$|II\rangle = |2\rangle - |e_1\rangle\langle e_1|2\rangle.$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

Clearly

$$\langle e_1|II\rangle = \langle e_1|2\rangle - \langle e_1|e_1\rangle\langle e_1|2\rangle = 0, \quad \text{i.e.,} \quad (II|\perp|e_1\rangle.$$

Dividing $|II\rangle$ by its norm (length), we now have the second basis vector and it is orthogonal to the first base vector $|e_1\rangle$ and of unit length.

Step 3. To construct the third member of the orthonormal basis, consider

$$|III\rangle = |3\rangle - |e_1\rangle\langle e_1|III\rangle - |e_2\rangle_2|III\rangle$$

which is orthogonal to both $|e_1\rangle$ and $|e_2\rangle$. Dividing by its norm we get $|e_3\rangle$.

Continuing in this way, we will obtain an orthonormal basis $|e_1\rangle, |e_2\rangle, \ldots, |e_n\rangle$.

The Cauchy–Schwarz inequality

If **A** and **B** are non-zero vectors in E_3, then the dot product gives $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta$, where θ is the angle between the vectors. If we square both sides and use the fact that $\cos^2\theta \le 1$, we obtain the inequality

$$(\mathbf{A}\cdot\mathbf{B})^2 \le A^2B^2 \quad \text{or} \quad |\mathbf{A}\cdot\mathbf{B}| \le AB.$$

This is known as the Cauchy–Schwarz inequality. There is an inequality corresponding to the Cauchy–Schwarz inequality in any inner-product space that obeys Axioms B.1–B.4, which can be stated as

$$|\langle U|W\rangle| \le |U||W|, \quad |U| = \sqrt{\langle U|U\rangle} \quad \text{etc.,} \tag{5.13}$$

where $|U\rangle$ and $|W\rangle$ are two non-zero vectors in an inner-product space.

This can be proved as follows. We first note that, for any scalar α, the following inequality holds

$$\begin{aligned} 0 \le |\langle U+\alpha W|U+\alpha W\rangle|^2 &= \langle U+\alpha W|U+\alpha W\rangle \\ &= \langle U|U\rangle + \langle \alpha W|U\rangle + \langle U|\alpha W\rangle + \langle \alpha W|\alpha W\rangle \\ &= |U|^2 + \alpha^*\langle V|U\rangle + \alpha\langle U|W\rangle + |\alpha|^2|W|^2. \end{aligned}$$

Now let $\alpha = \lambda\langle U|W\rangle^*/|\langle U|W\rangle|$, with λ real. This is possible if $|W\rangle \ne 0$, but if $\langle U|W\rangle = 0$, then Cauchy–Schwarz inequality is trivial. Making this substitution in the above, we have

$$0 \le |U|^2 + 2\lambda|\langle U|W\rangle| + \lambda^2|W|^2.$$

This is a quadratic expression in the real variable λ with real coefficients. Therefore, the discriminant must be less than or equal to zero:

$$4|\langle U|W\rangle|^2 - 4|U|^2|W|^2 \le 0$$

or

$$|\langle U|W\rangle| \leq |U||W|,$$

which is the Cauchy–Schwarz inequality.

From the Cauchy–Schwarz inequality follows another important inequality, known as the triangle inequality,

$$|U + W| \leq |U| + |W|. \tag{5.14}$$

The proof of this is very straightforward. For any pair of vectors, we have

$$\begin{aligned} |U + W|^2 = \langle U + W|U + W\rangle &= |U|^2 + |W|^2 + \langle U|W\rangle + \langle W|U\rangle \\ &\leq |U|^2 + |W|^2 + 2|\langle U|W\rangle| \\ &\leq |U|^2 + |W|^2 + 2|U||W|(|U|^2 + |W|^2) \end{aligned}$$

from which it follows that

$$|U + W| \leq |U| + |W|.$$

If V denotes the vector space of real continuous functions on the interval $a \leq x \leq b$, and f and g are any real continuous functions, then the following is an inner product on V:

$$\langle f|g\rangle = \int_a^b f(x)g(x)dx.$$

The Cauchy–Schwarz inequality now gives

$$\left(\int_a^b f(x)g(x)dx\right)^2 \leq \int_a^b f^2(x)dx \int_a^b g^2(x)dx$$

or in Dirac notation

$$|\langle f|g\rangle|^2 \leq |f|^2|g|^2.$$

Dual vectors and dual spaces

We begin with a technical point regarding the inner product $\langle u|v\rangle$. If we set

$$|v\rangle = \alpha|w\rangle + \beta|z\rangle,$$

then

$$\langle u|v\rangle = \alpha\langle u|w\rangle + \beta\langle u|z\rangle$$

is a linear function of α and β. However, if we set

$$|u\rangle = \alpha|w\rangle + \beta|z\rangle,$$

then

$$\langle u|v\rangle = \langle v|u\rangle^* = \alpha^*\langle v|w\rangle^* + \beta^*\langle v|z\rangle^* = \alpha^*\langle w|v\rangle + \beta^*\langle z|v\rangle$$

is no longer a linear function of α and β. To remove this asymmetry, we can introduce, besides the ket vectors $|\,\rangle$, bra vectors $\langle\,|$ which form a different vector space. We will assume that there is a one-to-one correspondence between ket vectors $|\,\rangle$, and bra vectors $\langle\,|$. Thus there are two vector spaces, the space of kets and a dual space of bras. A pair of vectors in which each is in correspondence with the other will be called a pair of dual vectors. Thus, for example, $\langle v|$ is the dual vector of $|v\rangle$. Note they always carry the same identification label.

We now define the multiplication of ket vectors by bra vectors by requiring

$$\langle u|\cdot|v\rangle \equiv \langle u|v\rangle.$$

Setting

$$\langle u| = \langle w|\alpha^* + \langle z|\beta^*,$$

we have

$$\langle u|v\rangle = \alpha^*\langle w|v\rangle + \beta^*\langle z|v\rangle,$$

the same result we obtained above, and we see that $\langle w|\alpha^* + \langle z|\beta^*$ is the dual vector of $\alpha|w\rangle + \beta|z\rangle$.

From the above discussion, it is obvious that inner products are really defined only between bras and kets and hence from elements of two distinct but related vector spaces. There is a basis of vectors $|i\rangle$ for expanding kets and a similar basis $\langle i|$ for expanding bras. The basis ket $|t\rangle$ is represented in the basis we are using by a column vector with all zeros except for a 1 in the ith row, while the basis $\langle i|$ is a row vector with all zeros except for a 1 in the ith column.

Linear operators

A useful concept in the study of linear vector spaces is that of a linear transformation, from which the concept of a linear operator emerges naturally. It is instructive first to review the concept of transformation or mapping. Given vector spaces V and W and function $\underset{\sim}{T}$, if $\underset{\sim}{T}$ associates each vector in V with a unique vector in W, we say $\underset{\sim}{T}$ *maps* V into W, and write $\underset{\sim}{T}\colon V \to W$. If $\underset{\sim}{T}$ associates the vector $|w\rangle$ in W with the vector $|v\rangle$ in V, we say that $|w\rangle$ is the *image* of $|v\rangle$ under $\underset{\sim}{T}$ and write $|w\rangle = \underset{\sim}{T}|v\rangle$. Further, $\underset{\sim}{T}$ is a linear transformation if:

(a) $\underset{\sim}{T}(|u\rangle + |v\rangle) = \underset{\sim}{T}|u\rangle + \underset{\sim}{T}|v\rangle$ for all vectors $|u\rangle$ and $|v\rangle$ in V.
(b) $\underset{\sim}{T}(k|v\rangle) = k\underset{\sim}{T}|v\rangle$ for all vectors $|v\rangle$ in V and all scalars k.

We can illustrate this with a very simple example. If $|v\rangle = (x, y)$ is a vector in E_2, then $\underset{\sim}{T}(|v\rangle) = (x, x+y, x-y)$ defines a function (a transformation) that maps

E_2 into E_3. In particular, if $|v\rangle = (1,1)$, then the image of $|v\rangle$ under $\underset{\sim}{T}$ is $\underset{\sim}{T}(|v\rangle) = (1,2,0)$. It is easy to see that the transformation is linear. If $|u\rangle = (x_1, y_1)$ and $|v\rangle = (x_2, y_2)$, then

$$|u\rangle + |v\rangle = (x_1 + x_2, y_1 + y_2),$$

so that

$$\begin{aligned}\underset{\sim}{T}(|u\rangle + |v\rangle) &= (x_1 + x_2, (x_1 + x_2) + (y_1 + y_2), (x_1 + x_2) - (y_1 + y_2)) \\ &= (x_1, x_1 + y_1, x_1 - y_1) + (x_2, x_2 + y_2, x_2 - y_2) \\ &= \underset{\sim}{T}(|u\rangle) + \underset{\sim}{T}(|v\rangle)\end{aligned}$$

and if k is a scalar, then

$$\underset{\sim}{T}(k|u\rangle) = (kx_1, kx_1 + ky_1, kx_1 - ky_1) = k(x_1, x_1 + y_1, x_1 - y_1) = k\underset{\sim}{T}(|u\rangle).$$

Thus $\underset{\sim}{T}$ is a linear transformation.

If $\underset{\sim}{T}$ maps the vector space onto itself ($\underset{\sim}{T}$: $V \to V$), then it is called a linear operator on V. In E_3 a rotation of the entire space about a fixed axis is an example of an operation that maps the space onto itself. We saw in Chapter 3 that rotation can be represented by a matrix with elements λ_{ij} $(i, j = 1, 2, 3)$; if x_1, x_2, x_3 are the components of an arbitrary vector in E_3 before the transformation and x_1', x_2', x_3' the components of the transformed vector, then

$$\left.\begin{aligned} x_1' &= \lambda_{11}x_1 + \lambda_{12}x_2 + \lambda_{13}x_3, \\ x_2' &= \lambda_{21}x_1 + \lambda_{22}x_2 + \lambda_{23}x_3, \\ x_3' &= \lambda_{31}x_1 + \lambda_{32}x_2 + \lambda_{33}x_3. \end{aligned}\right\} \tag{5.15}$$

In matrix form we have

$$\tilde{x}' = \tilde{\lambda}(\theta)\tilde{x}, \tag{5.16}$$

where θ is the angle of rotation, and

$$\tilde{x}' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix}, \quad \tilde{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \text{and} \quad \tilde{\lambda}(\theta) = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix}.$$

In particular, if the rotation is carried out about x_3-axis, $\tilde{\lambda}(\theta)$ has the following form:

$$\tilde{\lambda}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

Eq. (5.16) determines the vector $\mathbf{x}'$ if the vector $\mathbf{x}$ is given, and $\tilde{\lambda}(\theta)$ is the operator (matrix representation of the rotation operator) which turns $\mathbf{x}$ into $\mathbf{x}'$.

Loosely speaking, an operator is any mathematical entity which operates on any vector in V and turns it into another vector in V. Abstractly, an operator $\underset{\sim}{L}$ is a mapping that assigns to a vector $|v\rangle$ in a linear vector space V another vector $|u\rangle$ in V: $|u\rangle = \underset{\sim}{L}|v\rangle$. The set of vectors $|v\rangle$ for which the mapping is defined, that is, the set of vectors $|v\rangle$ for which $\underset{\sim}{L}|v\rangle$ has meaning, is called the domain of $\underset{\sim}{L}$. The set of vectors $|u\rangle$ in the domain expressible as $|u\rangle = \underset{\sim}{L}|v\rangle$ is called the range of the operator. An operator $\underset{\sim}{L}$ is linear if the mapping is such that for any vectors $|u\rangle, |w\rangle$ in the domain of $\underset{\sim}{L}$ and for arbitrary scalars α, β, the vector $\alpha|u\rangle + \beta|w\rangle$ is in the domain of $\underset{\sim}{L}$ and

$$\underset{\sim}{L}(\alpha|u\rangle + \beta|w\rangle) = \alpha\underset{\sim}{L}|u\rangle + \beta\underset{\sim}{L}|w\rangle.$$

A linear operator is bounded if its domain is the entire space V and if there exists a single constant C such that

$$|\underset{\sim}{L}|v\rangle| < C||v\rangle|$$

for all $|v\rangle$ in V. We shall consider linear bounded operators only.

Matrix representation of operators

Linear bounded operators may be represented by matrix. The matrix will have a finite or an infinite number of rows according to whether the dimension of V is finite or infinite. To show this, let $|1\rangle, |2\rangle, \ldots$ be an orthonormal basis in V; then every vector $|\varphi\rangle$ in V may be written in the form

$$|\varphi\rangle = \alpha_1|1\rangle + \alpha_2|2\rangle + \cdots.$$

Since $\underset{\sim}{L}|\rangle$ is also in V, we may write

$$\underset{\sim}{L}|\varphi\rangle = \beta_1|1\rangle + \beta_2|2\rangle + \cdots.$$

But

$$\underset{\sim}{L}|\varphi\rangle = \alpha_1\underset{\sim}{L}|1\rangle + \alpha_2\underset{\sim}{L}|2\rangle + \cdots,$$

so

$$\beta_1|1\rangle + \beta_2|2\rangle + \cdots = \alpha_1\underset{\sim}{L}|1\rangle + \alpha_2\underset{\sim}{L}|2\rangle + \cdots.$$

Taking the inner product of both sides with $\langle 1|$ we obtain

$$\beta_1 = \langle 1|\underset{\sim}{L}|1\rangle\alpha_1 + \langle 1|\underset{\sim}{L}|2\rangle\alpha_2 = \gamma_{11}\alpha_1 + \gamma_{12}\alpha_2 + \cdots;$$

Similarly

$$\beta_2 = \langle 2|\underset{\sim}{L}|1\rangle\alpha_1 + \langle 2|\underset{\sim}{L}|2\rangle\alpha_2 = \gamma_{21}\alpha_1 + \gamma_{22}\alpha_2 + \cdots,$$

$$\beta_3 = \langle 3|\underset{\sim}{L}|1\rangle\alpha_1 + \langle 3|\underset{\sim}{L}|2\rangle\alpha_2 = \gamma_{31}\alpha_1 + \gamma_{32}\alpha_2 + \cdots.$$

In general, we have

$$\beta_i = \sum_j \gamma_{ij}\alpha_j,$$

where

$$\gamma_{ij} = \langle i|\underset{\sim}{L}|j\rangle. \tag{5.17}$$

Consequently, in terms of the vectors $|1\rangle, |2\rangle, \ldots$ as a basis, operator $\underset{\sim}{L}$ is represented by the matrix whose elements are γ_{ij}.

A matrix representing $\underset{\sim}{L}$ can be found by using any basis, not necessarily an orthonormal one. Of course, a change in the basis changes the matrix representing $\underset{\sim}{L}$.

The algebra of linear operators

Let $\underset{\sim}{A}$ and $\underset{\sim}{B}$ be two operators defined in a linear vector space V of vectors $|\,\rangle$. The equation $\underset{\sim}{A} = \underset{\sim}{B}$ will be understood in the sense that

$$\underset{\sim}{A}|\,\rangle = \underset{\sim}{B}|\,\rangle \quad \text{for all } |\,\rangle \in V.$$

We define the addition and multiplication of linear operators as

$$\underset{\sim}{C} = \underset{\sim}{A} + \underset{\sim}{B} \quad \text{and} \quad \underset{\sim}{D} = \underset{\sim}{A}\,\underset{\sim}{B}$$

if for any $|\,\rangle$

$$\underset{\sim}{C}|\,\rangle = (\underset{\sim}{A} + \underset{\sim}{B})|\,\rangle = \underset{\sim}{A}|\,\rangle + \underset{\sim}{B}|\,\rangle,$$

$$\underset{\sim}{D}|\,\rangle = (\underset{\sim}{A}\underset{\sim}{B})|\,\rangle = \underset{\sim}{A}(\underset{\sim}{B}|\,\rangle).$$

Note that $\underset{\sim}{A} + \underset{\sim}{B}$ and $\underset{\sim}{A}\,\underset{\sim}{B}$ are themselves linear operators.

Example 5.10

(*a*) $(\underset{\sim}{A}\,\underset{\sim}{B})(\alpha|u\rangle + \beta|v\rangle) = \underset{\sim}{A}[\alpha(\underset{\sim}{B}|u\rangle) + \beta(\underset{\sim}{B}|v\rangle)] = \alpha(\underset{\sim}{A}\,\underset{\sim}{B})|u\rangle + \beta(\underset{\sim}{A}\,\underset{\sim}{B})|v\rangle$,
(*b*) $\underset{\sim}{C}(\underset{\sim}{A} + \underset{\sim}{B})|v\rangle = \underset{\sim}{C}(\underset{\sim}{A}|v\rangle + \underset{\sim}{B}|v\rangle) = \underset{\sim}{C}\,\underset{\sim}{A}|\,\rangle + \underset{\sim}{C}\,\underset{\sim}{B}|\,\rangle$,

which shows that

$$\underset{\sim}{C}(\underset{\sim}{A} + \underset{\sim}{B}) = \underset{\sim}{C}\,\underset{\sim}{A} + \underset{\sim}{C}\,\underset{\sim}{B}.$$

In general $\underset{\sim}{A}\,\underset{\sim}{B} \neq \underset{\sim}{B}\,\underset{\sim}{A}$. The difference $\underset{\sim}{A}\,\underset{\sim}{B}-\underset{\sim}{B}\,\underset{\sim}{A}$ is called the commutator of $\underset{\sim}{A}$ and $\underset{\sim}{B}$ and is denoted by the symbol $[\underset{\sim}{A}, \underset{\sim}{B}]$:

$$[\underset{\sim}{A}, \underset{\sim}{B}] \equiv \underset{\sim}{A}\,\underset{\sim}{B}-\underset{\sim}{B}\,\underset{\sim}{A}\,. \tag{5.18}$$

An operator whose commutator vanishes is called a commuting operator.

The operator equation

$$\underset{\sim}{B} = \alpha\underset{\sim}{A} = \underset{\sim}{A}\,\alpha$$

is equivalent to the vector equation

$$\underset{\sim}{B}|\,\rangle = \alpha\underset{\sim}{A}|\,\rangle \quad \text{for any } |\,\rangle.$$

And the vector equation

$$\underset{\sim}{A}|\,\rangle = \alpha|\,\rangle$$

is equivalent to the operator equation

$$\underset{\sim}{A} = \alpha\underset{\sim}{E}$$

where $\underset{\sim}{E}$ is the identity (or unit) operator:

$$\underset{\sim}{E}|\,\rangle = |\,\rangle \qquad \text{for any } |\,\rangle.$$

It is obvious that the equation $\underset{\sim}{A} = \alpha$ is meaningless.

Example 5.11

To illustrate the non-commuting nature of operators, let $\underset{\sim}{A} = x, \underset{\sim}{B} = d/dx$. Then

$$\underset{\sim}{A}\,\underset{\sim}{B}f(x) = x\frac{d}{dx}\,f(x),$$

and

$$\underset{\sim}{B}\,\underset{\sim}{A}f(x) = \frac{d}{dx}xf(x) = \left(\frac{dx}{dx}\right)f + x\frac{df}{dx} = f + x\frac{df}{dx} = (\underset{\sim}{E}+\underset{\sim}{A}\,\underset{\sim}{B})f.$$

Thus,

$$(\underset{\sim}{A}\,\underset{\sim}{B}-\underset{\sim}{B}\,\underset{\sim}{A})f(x) = -\underset{\sim}{E}f(x)$$

or

$$\left[x, \frac{d}{dx}\right] = x\frac{d}{dx} - \frac{d}{dx}x = -\underset{\sim}{E}\,.$$

Having defined the product of two operators, we can also define an operator raised to a certain power. For example

$$\underset{\sim}{A}^m|\,\rangle = \underbrace{\underset{\sim}{A}\quad\underset{\sim}{A}\quad\cdots\quad\underset{\sim}{A}}_{m\,\text{factor}}|\,\rangle.$$

By combining the operations of addition and multiplication, functions of operators can be formed. We can also define functions of operators by their power series expansions. For example, $e^{\underset{\sim}{A}}$ formally means

$$e^{\underset{\sim}{A}} \equiv 1 + \underset{\sim}{A} + \frac{1}{2!}\underset{\sim}{A}^2 + \frac{1}{3!}\underset{\sim}{A}^3 + \cdots.$$

A function of a linear operator is a linear operator.

Given an operator $\underset{\sim}{A}$ that acts on vector $|\,\rangle$, we can define the action of the same operator on vector $\langle\,|$. We shall use the convention of operating on $\langle\,|$ from the right. Then the action of $\underset{\sim}{A}$ on a vector $\langle\,|$ is defined by requiring that for any $|u\rangle$ and $\langle v|$, we have

$$\{\langle u|\underset{\sim}{A}\}|v\rangle \equiv \langle u|\{\underset{\sim}{A}|v\rangle\} = \langle u|\underset{\sim}{A}|v\rangle.$$

We may write $\alpha|v\rangle = |\alpha v\rangle$ and the corresponding bra as $\langle \alpha v|$. However, it is important to note that $\langle \alpha v| = A^*\langle v|$.

Eigenvalues and eigenvectors of an operator

The result of operating on a vector with an operator $\underset{\sim}{A}$ is, in general, a different vector. But there may be some vector $|v\rangle$ with the property that operating with $\underset{\sim}{A}$ on it yields the same vector $|v\rangle$ multiplied by a scalar, say α:

$$\underset{\sim}{A}\,|v\rangle = \alpha|v\rangle.$$

This is called the eigenvalue equation for the operator $\underset{\sim}{A}$, and the vector $|v\rangle$ is called an eigenvector of $\underset{\sim}{A}$ belonging to the eigenvalue α. A linear operator has, in general, several eigenvalues and eigenvectors, which can be distinguished by a subscript

$$\underset{\sim}{A}\,|v_k\rangle = \alpha_k|v_k\rangle.$$

The set $\{\alpha_k\}$ of all the eigenvalues taken together constitutes the spectrum of the operator. The eigenvalues may be discrete, continuous, or partly discrete and partly continuous. In general, an eigenvector belongs to only one eigenvalue. If several linearly independent eigenvectors belong to the same eigenvalue, the eigenvalue is said to be degenerate, and the degree of degeneracy is given by the number of linearly independent eigenvectors.

Some special operators

Certain operators with rather special properties play very important roles in physics. We now consider some of them below.

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

The inverse of an operator

The operator $\underset{\sim}{X}$ satisfying $\underset{\sim}{X}\underset{\sim}{A} = \underset{\sim}{E}$ is called the left inverse of $\underset{\sim}{A}$ and we denote it by $\underset{\sim}{A}_L^{-1}$. Thus, $\underset{\sim}{A}_L^{-1}\underset{\sim}{A} \equiv \underset{\sim}{E}$. Similarly, the right inverse of $\underset{\sim}{A}$ is defined by the equation

$$\underset{\sim}{A}\,\underset{\sim}{A}_R^{-1} \equiv \underset{\sim}{E}\,.$$

In general, $\underset{\sim}{A}_L^{-1}$ or $\underset{\sim}{A}_R^{-1}$, or both, may not be unique and even may not exist at all. However, if both $\underset{\sim}{A}_L^{-1}$ and $\underset{\sim}{A}_R^{-1}$ exist, then they are unique and equal to each other:

$$\underset{\sim}{A}_L^{-1} = \underset{\sim}{A}_R^{-1} \equiv \underset{\sim}{A}^{-1},$$

and

$$\underset{\sim}{A}\underset{\sim}{A}^{-1} = \underset{\sim}{A}^{-1}\underset{\sim}{A} = \underset{\sim}{E}\,. \tag{5.19}$$

$\underset{\sim}{A}^{-1}$ is called the operator inverse to $\underset{\sim}{A}$. Obviously, an operator is the inverse of another if the corresponding matrices are.

An operator for which an inverse exists is said to be non-singular, whereas one for which no inverse exists is singular. A necessary and sufficient condition for an operator $\underset{\sim}{A}$ to be non-singular is that corresponding to each vector $|u\rangle$, there should be a unique vector $|v\rangle$ such that $|u\rangle = \underset{\sim}{A}\,|v\rangle$.

The inverse of a linear operator is a linear operator. The proof is simple: let

$$|u_1\rangle = \underset{\sim}{A}\,|v_1\rangle, \quad |u_2\rangle = \underset{\sim}{A}\,|v_2\rangle.$$

Then

$$|v_1\rangle = \underset{\sim}{A}^{-1}|u_1\rangle, \quad |v_2\rangle = \underset{\sim}{A}^{-1}|u_2\rangle$$

so that

$$c_1|v_1\rangle = c_1\underset{\sim}{A}^{-1}|u_1\rangle, \quad c_2|v_2\rangle = c_2\underset{\sim}{A}^{-1}|u_2\rangle.$$

Thus,

$$\begin{aligned}\underset{\sim}{A}^{-1}[c_1|u_1\rangle + c_2|u_2\rangle] &= \underset{\sim}{A}^{-1}[c_1\underset{\sim}{A}\,|v_1\rangle + c_2\underset{\sim}{A}\,|v_2\rangle] \\ &= \underset{\sim}{A}^{-1}\underset{\sim}{A}[c_1|v_1\rangle + c_2|v_2\rangle] \\ &= c_1|v_1\rangle + c_2|v_2\rangle\end{aligned}$$

or

$$\underset{\sim}{A}^{-1}[c_1|u_1\rangle + c_2|u_2\rangle] = c_1\underset{\sim}{A}^{-1}|u_1\rangle + c_2\underset{\sim}{A}^{-1}|u_2\rangle.$$

The inverse of a product of operators is the product of the inverse in the reverse order

$$(\underset{\sim}{A}\underset{\sim}{B})^{-1} = \underset{\sim}{B}^{-1}\underset{\sim}{A}^{-1}. \tag{5.20}$$

The proof is straightforward: we have

$$\underset{\sim}{A}\underset{\sim}{B}(\underset{\sim}{A}\underset{\sim}{B})^{-1} = \underset{\sim}{E}.$$

Multiplying successively from the left by $\underset{\sim}{A}^{-1}$ and $\underset{\sim}{B}^{-1}$, we obtain

$$(\underset{\sim}{A}\underset{\sim}{B})^{-1} = \underset{\sim}{B}^{-1}\underset{\sim}{A}^{-1},$$

which is identical to Eq. (5.20).

The adjoint operators

Assuming that V is an inner-product space, then the operator $\underset{\sim}{X}$ satisfying the relation

$$\langle u|\underset{\sim}{X}|v\rangle = \langle v|\underset{\sim}{A}|u\rangle^* \quad \text{for any } |u\rangle, \quad |v\rangle \in V$$

is called the adjoint operator of $\underset{\sim}{A}$ and is denoted by $\underset{\sim}{A}^+$. Thus

$$\langle u|\underset{\sim}{A}^+|v\rangle \equiv \langle v|\underset{\sim}{A}|u\rangle^* \quad \text{for any } |u\rangle, \quad |v\rangle \in V. \tag{5.21}$$

We first note that $\langle\ |\underset{\sim}{A}^+$ is a dual vector of $\underset{\sim}{A}|\ \rangle$. Next, it is obvious that

$$(\underset{\sim}{A}^+)^+ = \underset{\sim}{A}. \tag{5.22}$$

To see this, let $\underset{\sim}{A}^+ = \underset{\sim}{B}$, then $(\underset{\sim}{A}^+)^+$ becomes $\underset{\sim}{B}^+$, and from Eq. (5.21) we find

$$\langle v|\underset{\sim}{B}^+|u\rangle = \langle u|\underset{\sim}{B}|v\rangle^*, \quad \text{for any } |u\rangle,\ |v\rangle \in V.$$

But

$$\langle u|\underset{\sim}{B}|v\rangle^* = \langle u|\underset{\sim}{A}^+|v\rangle^* = \langle v|\underset{\sim}{A}|u\rangle.$$

Thus

$$\langle v|\underset{\sim}{B}^+|u\rangle = \langle u|\underset{\sim}{B}|v\rangle^* = \langle v|\underset{\sim}{A}|u\rangle$$

from which we find

$$(\underset{\sim}{A}^+)^+ = \underset{\sim}{A}.$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

It is also easy to show that

$$(\underset{\sim}{A}\underset{\sim}{B})^{+} = \underset{\sim}{B}^{+}\underset{\sim}{A}^{+}. \tag{5.23}$$

For any $|u\rangle, |v\rangle$, $\langle v|\underset{\sim}{B}^{+}$ and $\underset{\sim}{B}\,|v\rangle$ is a pair of dual vectors; $\langle u|\underset{\sim}{A}^{+}$ and $\underset{\sim}{A}\,|u\rangle$ is also a pair of dual vectors. Thus we have

$$\langle v|\underset{\sim}{B}^{+}\underset{\sim}{A}^{+}|u\rangle = \{\langle v|\underset{\sim}{B}^{+}\}\{\underset{\sim}{A}^{+}|u\rangle\} = [\{\langle u|\underset{\sim}{A}\}\{\underset{\sim}{B}|v\rangle\}]^{*}$$

$$= \langle u|\underset{\sim}{A}\,\underset{\sim}{B}|v\rangle^{*} = \langle v|(\underset{\sim}{A}\,\underset{\sim}{B})^{+}|u\rangle$$

and therefore

$$(\underset{\sim}{A}\,\underset{\sim}{B})^{+} = \underset{\sim}{B}^{+}\underset{\sim}{A}^{+}.$$

Hermitian operators

An operator $\underset{\sim}{H}$ that is equal to its adjoint, that is, that obeys the relation

$$\underset{\sim}{H} = \underset{\sim}{H}^{+} \tag{5.24}$$

is called Hermitian or self-adjoint. And $\underset{\sim}{H}$ is anti-Hermitian if

$$\underset{\sim}{H} = -\underset{\sim}{H}^{+}.$$

Hermitian operators have the following important properties:

(1) The eigenvalues are real: Let $\underset{\sim}{H}$ be the Hermitian operator and let $|v\rangle$ be an eigenvector belonging to the eigenvalue α:

$$\underset{\sim}{H}\,|v\rangle = \alpha|v\rangle.$$

By definition, we have

$$\langle v|\underset{\sim}{A}|v\rangle = \langle v|\underset{\sim}{A}|v\rangle^{*},$$

that is,

$$(\alpha^{*} - \alpha)\langle v|v\rangle = 0.$$

Since $\langle v|v\rangle \neq 0$, we have

$$\alpha^{*} = \alpha.$$

(2) Eigenvectors belonging to different eigenvalues are orthogonal: Let $|u\rangle$ and $|v\rangle$ be eigenvectors of $\underset{\sim}{H}$ belonging to the eigenvalues α and β respectively:

$$\underset{\sim}{H}\,|u\rangle = \alpha|u\rangle, \quad \underset{\sim}{H}\,|v\rangle = \beta\,|v\rangle.$$

Then

$$\langle u|\underset{\sim}{H}|v\rangle = \langle v|\underset{\sim}{H}|u\rangle^*.$$

That is,

$$(\alpha - \beta)\langle v|u\rangle = 0 \quad (\text{since } \alpha^* = \alpha).$$

But $\alpha \neq \beta$, so that

$$\langle v|u\rangle = 0.$$

(3) The set of all eigenvectors of a Hermitian operator forms a complete set: The eigenvectors are orthogonal, and since we can normalize them, this means that the eigenvectors form an orthonormal set and serve as a basis for the vector space.

Unitary operators

A linear operator $\underset{\sim}{U}$ is unitary if it preserves the Hermitian character of an operator under a similarity transformation:

$$(\underset{\sim}{U}\,\underset{\sim}{A}\,\underset{\sim}{U}^{-1})^+ = \underset{\sim}{U}\,\underset{\sim}{A}\,\underset{\sim}{U}^{-1},$$

where

$$\underset{\sim}{A}^+ = \underset{\sim}{A}.$$

But, according to Eq. (5.23)

$$(\underset{\sim}{U}\,\underset{\sim}{A}\,\underset{\sim}{U}^{-1})^+ = (\underset{\sim}{U}^{-1})^+\underset{\sim}{A}\underset{\sim}{U}^+,$$

thus, we have

$$(\underset{\sim}{U}^{-1})^+\underset{\sim}{A}\,\underset{\sim}{U}^+ = \underset{\sim}{U}\,\underset{\sim}{A}\,\underset{\sim}{U}^{-1}.$$

Multiplying from the left by $\underset{\sim}{U}^+$ and from the right by $\underset{\sim}{U}$, we obtain

$$\underset{\sim}{U}^+(\underset{\sim}{U}^{-1})^+\underset{\sim}{A}\,\underset{\sim}{U}^+\underset{\sim}{U} = \underset{\sim}{U}+\underset{\sim}{U}\,\underset{\sim}{A};$$

this reduces to

$$\underset{\sim}{A}(\underset{\sim}{U}^+\underset{\sim}{U}) = (\underset{\sim}{U}+\underset{\sim}{U})\underset{\sim}{A},$$

since

$$\underset{\sim}{U}^+(\underset{\sim}{U}^{-1})^+ = (\underset{\sim}{U}^{-1}\underset{\sim}{U})^+ = \underset{\sim}{E}.$$

Thus

$$\underset{\sim}{U}^+\underset{\sim}{U} = \underset{\sim}{E}$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

or

$$\underset{\sim}{U}^{+} = \underset{\sim}{U}^{-1}. \tag{5.25}$$

We often use Eq. (5.25) for the definition of the unitary operator.

Unitary operators have the remarkable property that transformation by a unitary operator preserves the inner product of the vectors. This is easy to see: under the operation $\underset{\sim}{U}$, a vector $|v\rangle$ is transformed into the vector $|v'\rangle = \underset{\sim}{U}|v\rangle$. Thus, if two vectors $|v\rangle$ and $|u\rangle$ are transformed by the same unitary operator $\underset{\sim}{U}$, then

$$\langle u'|v'\rangle = \langle \underset{\sim}{U} u|\underset{\sim}{U} v\rangle = \langle u|\underset{\sim}{U}^{+}\underset{\sim}{U} v\rangle = \langle u|v\rangle,$$

that is, the inner product is preserved. In particular, it leaves the norm of a vector unchanged. Thus, a unitary transformation in a linear vector space is analogous to a rotation in the physical space (which also preserves the lengths of vectors and the inner products).

Corresponding to every unitary operator $\underset{\sim}{U}$, we can define a Hermitian operator $\underset{\sim}{H}$ and vice versa by

$$\underset{\sim}{U} = e^{i\varepsilon \underset{\sim}{H}}, \tag{5.26}$$

where ε is a parameter. Obviously

$$\underset{\sim}{U}^{+} = e^{(i\varepsilon \underset{\sim}{H})^{+}} / e^{-i\varepsilon \underset{\sim}{H}} = \underset{\sim}{U}^{-1}.$$

A unitary operator possesses the following properties:

(1) The eigenvalues are unimodular; that is, if $\underset{\sim}{U}|v\rangle = \alpha|v\rangle$, then $|\alpha| = 1$.
(2) Eigenvectors belonging to different eigenvalues are orthogonal.
(3) The product of unitary operators is unitary.

The projection operators

A symbol of the type of $|u\rangle\langle v|$ is quite useful: it has all the properties of a linear operator, multiplied from the right by a ket $|\ \rangle$, it gives $|u\rangle$ whose magnitude is $\langle v|\ \rangle$; and multiplied from the left by a bra $\langle\ |$ it gives $\langle v|$ whose magnitude is $\langle\ |u\rangle$. The linearity of $|u\rangle\langle v|$ results from the linear properties of the inner product. We also have

$$\{|u\rangle\langle v|\}^{+} = |v\rangle\langle u|.$$

The operator $\underset{\sim}{P}_j = |j\rangle\langle j|$ is a very particular example of projection operator. To see its effect on an arbitrary vector $|u\rangle$, let us expand $|u\rangle$:

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

$$|u\rangle = \sum_{j=1}^{n} u_j|j\rangle, \quad u_j = \langle j|u\rangle. \tag{5.27}$$

We may write the above as

$$|u\rangle = \left(\sum_{j=1}^{n} |j\rangle\langle j|\right)|u\rangle,$$

which is true for all $|u\rangle$. Thus the object in the brackets must be identified with the identity operator:

$$\underset{\sim}{I} = \sum_{j=1}^{n} |j\rangle\langle j| = \sum_{j=1}^{n} \underset{\sim}{P}_j. \tag{5.28}$$

Now we will see that the effect of this particular projection operator on $|u\rangle$ is to produce a new vector whose direction is along the basis vector $|j\rangle$ and whose magnitude is $\langle j|u\rangle$:

$$\underset{\sim}{P}_j|u\rangle = |j\rangle\langle j|u\rangle = |j\rangle u_j.$$

We see that whatever $|u\rangle$ is, $\underset{\sim}{P}_j|u\rangle$ is a multiple of $|j\rangle$ with a coefficient u_j which is the component of $|u\rangle$ along $|j\rangle$. Eq. (5.28) says that the sum of the projections of a vector along all the n directions equals the vector itself.

When $\underset{\sim}{P}_j = |j\rangle\langle j|$ acts on $|j\rangle$, it reproduces that vector. On the other hand, since the other basis vectors are orthogonal to $|j\rangle$, a projection operation on any one of them gives zero (null vector). The basis vectors are therefore eigenvectors of $\underset{\sim}{P}_k$ with the property

$$\underset{\sim}{P}_k|j\rangle = \delta_{kj}|j\rangle, \qquad (j, k = 1, \ldots, n).$$

In this orthonormal basis the projection operators have the matrix form

$$\underset{\sim}{P}_1 = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \underset{\sim}{P}_2 = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \underset{\sim}{P}_N = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots_{\;1} \end{pmatrix}.$$

Projection operators can also act on bras in the same way:

$$\langle u|\underset{\sim}{P}_j = \langle u|j\rangle\langle j| = u_j^*\langle j|.$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

Change of basis

The choice of base vectors (basis) is largely arbitrary and different representations are physically equally acceptable. How do we change from one orthonormal set of base vectors $|\varphi_1\rangle, |\varphi_2\rangle, \ldots, |\varepsilon_n\rangle$ to another such set $|\xi_1\rangle, |\xi_2\rangle, \ldots, |\xi_n\rangle$? In other words, how do we generate the orthonomal set $|\xi_1\rangle, |\xi_2\rangle, \ldots, |\xi_n\rangle$ from the old set $|\varphi_1\rangle, |\varphi_2\rangle, \ldots, |\varphi_n$? This task can be accomplished by a unitary transformation:

$$|\xi_i\rangle = \underset{\sim}{U}\,|\varphi_i\rangle \quad (i = 1, 2, \ldots, n). \tag{5.29}$$

Then given a vector $|X\rangle = \sum_{i=1}^{n} a_i|\varphi_i\rangle$, it will be transformed into $|X'\rangle$:

$$|X'\rangle = \underset{\sim}{U}\,|X\rangle = \underset{\sim}{U}\sum_{i=1}^{n} a_i|\varphi_i\rangle = \sum_{i=1}^{n} \underset{\sim}{U}\,a_i|\varphi_i\rangle = \sum_{i=1}^{n} a_i|\xi_i\rangle.$$

We can see that the operator $\underset{\sim}{U}$ possesses an inverse $\underset{\sim}{U}^{-1}$ which is defined by the equation

$$|\varphi_i\rangle = \underset{\sim}{U}^{-1}|\xi_i\rangle \quad (i = 1, 2, \ldots, n).$$

The operator $\underset{\sim}{U}$ is unitary; for, if $|X\rangle = \sum_{i=1}^{n} a_i|\varphi_i\rangle$ and $|Y\rangle = \sum_{i=1}^{n} b_i|\varphi_i\rangle$, then

$$\langle X|Y\rangle = \sum_{i,j=1}^{n} a_i^* b_j \langle \varphi_i|\varphi_j\rangle = \sum_{i=1}^{n} a_i^* b_i, \qquad \langle UX|UY\rangle = \sum_{i,j=1}^{n} a_i^* b_j \langle \xi_i|\xi_j\rangle = \sum_{i=1}^{n} a_i^* b_i.$$

Hence

$$U^{-1} = U^{+}.$$

The inner product of two vectors is independent of the choice of basis which spans the vector space, since unitary transformations leave all inner products invariant. In quantum mechanics inner products give physically observable quantities, such as expectation values, probabilities, etc.

It is also clear that the matrix representation of an operator is different in a different basis. To find the effect of a change of basis on the matrix representation of an operator, let us consider the transformation of the vector $|X\rangle$ into $|Y\rangle$ by the operator $\underset{\sim}{A}$:

$$|Y\rangle = \underset{\sim}{A}\,|(X\rangle. \tag{5.30}$$

Referred to the basis $|\varphi_1\rangle$, $|\varphi_2\rangle$, $\ldots$, $|\varphi\rangle$, $|X\rangle$ and $|Y\rangle$ are given by $|X\rangle = \sum_{i=1}^{n} a_i|\varphi_i\rangle$ and $|Y\rangle = \sum_{i=1}^{n} b_i|\varphi_i\rangle$, and the equation $|Y\rangle = \underset{\sim}{A}\,|X\rangle$ becomes

$$\sum_{i=1}^{n} b_i|\varphi_i\rangle = \underset{\sim}{A}\sum_{j=1}^{n} a_j|\varphi_j\rangle.$$

Multiplying both sides from the left by the bra vector $\langle \varphi_i|$ we find

$$b_i = \sum_{j=1}^{n} a_j \langle \varphi_i | \underset{\sim}{A} | \varphi_j \rangle = \sum_{j=1}^{n} a_j A_{ij}. \tag{5.31}$$

Referred to the basis $|\xi_1\rangle, |\xi_2\rangle, \ldots, |\xi_n\rangle$ the same vectors $|X\rangle$ and $|Y\rangle$ are $|X\rangle = \sum_{i=1}^{n} a_i' |\xi_i\rangle$, and $|Y\rangle = \sum_{i=1}^{n} b_i' |\xi_i\rangle$, and Eqs. (5.31) are replaced by

$$b_i' = \sum_{j=1}^{n} a_j' \langle \xi_i | \underset{\sim}{A} | \xi_j \rangle = \sum_{j=1}^{n} a_j' A_{ij}',$$

where $A_{ij}' = \langle \xi_i | \underset{\sim}{A} | \xi_j \rangle$, which is related to A_{ij} by the following relation:

$$A_{ij}' = \langle \xi_i | \underset{\sim}{A} | \xi_j \rangle = \langle U\varphi_i | \underset{\sim}{A} | U\varphi_j \rangle = \langle \varphi_i | U^* \underset{\sim}{A}\, U | \varphi_j \rangle = (U^* \underset{\sim}{A}\, U)_{ij}$$

or using the rule for matrix multiplication

$$A_{ij}' = \langle \xi_i | \underset{\sim}{A} | \xi_j \rangle = (U^* \underset{\sim}{A}\, U)_{ij} = \sum_{r=1}^{n} \sum_{s=1}^{n} U_{ir}^* A_{rs} U_{sj}. \tag{5.32}$$

From Eqs. (5.32) we can find the matrix representation of an operator with respect to a new basis.

If the operator $\underset{\sim}{A}$ transforms vector $|X\rangle$ into vector $|Y\rangle$ which is vector $|X\rangle$ itself multiplied by a scalar λ: $|Y\rangle = \lambda |X\rangle$, then Eq. (5.30) becomes an eigenvalue equation:

$$\underset{\sim}{A}\, | X \rangle = \lambda \, | X \rangle.$$

Commuting operators

In general, operators do not commute. But commuting operators do exist and they are of importance in quantum mechanics. As Hermitian operators play a dominant role in quantum mechanics, and the eigenvalues and the eigenvectors of a Hermitian operator are real and form a complete set, respectively, we shall concentrate on Hermitian operators. It is straightforward to prove that

> Two commuting Hermitian operators possess a complete orthonormal set of common eigenvectors, and vice versa.

If $\underset{\sim}{A}$ and $\underset{\sim}{A}\,|v\rangle = \alpha|v\rangle$ are two commuting Hermitian operators, and if

$$\underset{\sim}{A}\,|v\rangle = \alpha|v\rangle, \tag{5.33}$$

then we have to show that

$$\underset{\sim}{B}\,|v\rangle = \beta|v\rangle. \tag{5.34}$$

Multiplying Eq. (5.33) from the left by $\underset{\sim}{B}$, we obtain

$$\underset{\sim}{B}(\underset{\sim}{A}|v\rangle) = \alpha(\underset{\sim}{B}|v\rangle),$$

which using the fact $\underset{\sim}{A}\,\underset{\sim}{B} = \underset{\sim}{B}\,\underset{\sim}{A}$, can be rewritten as

$$\underset{\sim}{A}(\underset{\sim}{B}|v\rangle) = \alpha(\underset{\sim}{B}|v\rangle).$$

Thus, $\underset{\sim}{B}|v\rangle$ is an eigenvector of $\underset{\sim}{A}$ belonging to eigenvalue α. If α is non-degenerate, then $\underset{\sim}{B}|v\rangle$ should be linearly dependent on $|v\rangle$, so that

$$a(\underset{\sim}{B}|v\rangle) + b|v\rangle = 0, \quad \text{with} \quad a \neq 0 \quad \text{and} \quad b \neq 0.$$

It follows that

$$\underset{\sim}{B}|v\rangle = -(b/a)|v\rangle = \beta|v\rangle.$$

If A is degenerate, then the matter becomes a little complicated. We now state the results without proof. There are three possibilities:

(1) The degenerate eigenvectors (that is, the linearly independent eigenvectors belonging to a degenerate eigenvalue) of $\underset{\sim}{A}$ are degenerate eigenvectors of $\underset{\sim}{B}$ also.
(2) The degenerate eigenvectors of $\underset{\sim}{A}$ belong to different eigenvalues of $\underset{\sim}{B}$. In this case, we say that the degeneracy is removed by the Hermitian operator $\underset{\sim}{B}$.
(3) Every degenerate eigenvector of $\underset{\sim}{A}$ is not an eigenvector of $\underset{\sim}{B}$. But there are linear combinations of the degenerate eigenvectors, as many in number as the degrees of degeneracy, which are degenerate eigenvectors of $\underset{\sim}{A}$ but are non-degenerate eigenvectors of $\underset{\sim}{B}$. Of course, the degeneracy is removed by $\underset{\sim}{B}$.

Function spaces

We have seen that functions can be elements of a vector space. We now return to this theme for a more detailed analysis. Consider the set of all functions that are continuous on some interval. Two such functions can be added together to construct a third function $h(x)$:

$$h(x) = f(x) + g(x), \quad a \leq x \leq b,$$

where the plus symbol has the usual operational meaning of 'add the value of f at the point x to the value of g at the same point.'

A function $f(x)$ can also be multiplied by a number k to give the function $p(x)$:

$$p(x) = k \cdot f(x), \quad a \leq x \leq b.$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

The centred dot, the multiplication symbol, is again understood in the conventional meaning of 'multiply by k the value of $f(x)$ at the point x.'

It is evident that the following conditions are satisfied:

(*a*) By adding two continuous functions, we obtain a continuous function.
(*b*) The multiplication by a scalar of a continuous function yields again a continuous function.
(*c*) The function that is identically zero for $a \leq x \leq b$ is continuous, and its addition to any other function does not alter this function.
(*d*) For any function $f(x)$ there exists a function $(-1)f(x)$, which satisfies

$$f(x) + [(-1)f(x)] = 0.$$

Comparing these statements with the axioms for linear vector spaces (Axioms A.1–A.8), we see clearly that the set of all continuous functions defined on some interval forms a linear vector space; this is called a function space. We shall consider the entire set of values of a function $f(x)$ as representing a vector $|f\rangle$ of this abstract vector space F (F stands for function space). In other words, we shall treat the number $f(x)$ at the point x as the component with 'index x' of an abstract vector $|f\rangle$. This is quite similar to what we did in the case of finite-dimensional spaces when we associated a component a_i of a vector with each value of the index i. The only difference is that this index assumed a discrete set of values 1, 2, etc., up to N (for N-dimensional space), whereas the argument x of a function $f(x)$ is a continuous variable. In other words, the function $f(x)$ has an infinite number of components, namely the values it takes in the continuum of points labeled by the real variable x. However, two questions may be raised.

The first question concerns the orthonormal basis. The components of a vector are defined with respect to some basis and we do not know which basis has been (or could be) chosen in the function space. Unfortunately, we have to postpone the answer to this question. Let us merely note that, once a basis has been chosen, we work only with the components of a vector. Therefore, provided we do not change to other basis vectors, we need not be concerned about the particular basis that has been chosen.

The second question is how to define an inner product in an infinite-dimensional vector space. Suppose the function $f(x)$ describes the displacement of a string clamped at $x = 0$ and $x = L$. We divide the interval of length L into N equal parts and measure the displacements $f(x_i) \equiv f_i$ at N point $x_i, i = 1, 2, \ldots, N$. At fixed N, the functions are elements of a finite N-dimensional vector space. An inner product is defined by the expression

$$\langle f|g\rangle = \sum_{i=1}^{N} f_i g_i.$$

For a vibrating string, the space is real and there is no need to conjugate anything.

To improve the description, we can increase the number N. However, as $N \to \infty$ by increasing the number of points without limit, the inner product diverges as we subdivide further and further. The way out of this is to modify the definition by a positive prefactor $\Delta = L/N$ which does not violate any of the axioms for the inner product. But now

$$\langle f|g\rangle = \lim_{\Delta\to 0}\sum_{i=1}^{N} f_i g_i \Delta \to \int_0^L f(x)g(x)dx,$$

by the usual definition of an integral. Thus the inner product of two functions is the integral of their product. Two functions are orthogonal if this inner product vanishes, and a function is normalized if the integral of its square equals unity. Thus we can speak of an orthonormal set of functions in a function space just as in finite dimensions. The following is an example of such a set of functions defined in the interval $0 \le x \le L$ and vanishing at the end points:

$$|e_m\rangle \to m(x) = \sqrt{\frac{2}{L}}\sin\frac{m\pi x}{L}, \qquad m = 1, 2, \ldots, \infty,$$

$$\langle e_m|e_n\rangle = \frac{2}{L}\int_0^L \sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L}dx = \delta_{mn}.$$

For the details, see 'Vibrating strings' of Chapter 4.

In quantum mechanics we often deal with complex functions and our definition of the inner product must then modified. We define the inner product of $f(x)$ and $g(x)$ as

$$\langle f|g\rangle = \int_0^L f^*(x)g(x)dx,$$

where f^* is the complex conjugate of f. An orthonormal set for this case is

$$m(x) = \frac{1}{\sqrt{2\pi}}e^{imx}, \qquad m = 0, \pm 1, \pm 2, \ldots,$$

which spans the space of all functions of period 2π with finite norm. A linear vector space with a complex-type inner product is called a *Hilbert space*.

Where and how did we get the orthonormal functions? In general, by solving the eigenvalue equation of some Hermitian operator. We give a simple example here. Consider the derivative operator $D = d(\,)/dx$:

$$Df(x) = df(x)/dx, \quad D|f\rangle = d|f\rangle/dx.$$

However, D is not Hermitian, because it does not meet the condition:

$$\int_0^L f^*(x)\frac{dg(x)}{dx}dx = \left(\int_0^L g^*(x)\frac{df(x)}{dx}dx\right)^*.$$

Here is why:

$$\left(\int_0^L g^*(x)\frac{df(x)}{dx}dx\right)^* = \int_0^L g(x)\frac{df^*(x)}{dx}dx$$
$$= gf^*\Big|_0^L - \int_0^L f^*(x)\frac{dg(x)}{dx}dx.$$

It is easy to see that hermiticity of D is lost on two counts. First we have the term coming from the end points. Second the integral has the wrong sign. We can fix both of these by doing the following:

(a) Use operator $-iD$. The extra i will change sign under conjugation and kill the minus sign in front of the integral.
(b) Restrict the functions to those that are periodic: $f(0) = f(L)$.

Thus, $-iD$ is a Hermitian operator on period functions. Now we have

$$-i\frac{df(x)}{dx} = \lambda f(x),$$

where λ is the eigenvalue. Simple integration gives

$$f(x) = Ae^{i\lambda x}.$$

Now the periodicity requirement gives

$$e^{i\lambda L} = e^{i\lambda 0} = 1$$

from which it follows that

$$\lambda = 2\pi m/L, \qquad m = 0, \pm 1, \pm 2,$$

and the normalization condition gives

$$A = \frac{1}{\sqrt{L}}.$$

Hence the set of orthonormal eigenvectors is given by

$$f_m(x) = \frac{1}{\sqrt{L}}e^{2\pi imx/L}.$$

In quantum mechanics the eigenvalue equation is the Schrödinger equation and the Hermitian operator is the Hamiltonian operator. Quantum mechanically, a system with n degrees of freedom which is classically specified by n generalized coordinates $q_1, \ldots, q_2, q_n$ is specified at a fixed instant of time by a wave function $\psi(q_1, q_2, \ldots, q_n)$ whose norm is unity, that is,

$$\langle\psi|\psi\rangle = \int |\psi(q_1, q_2, \ldots, q_n)|^2 dq_1, dq_2, \ldots, dq_n = 1,$$

https://doi.org/10.1017/CBO9780511755781.006 Published online by Cambridge University Press

the integration being over the accessible values of the coordinates $q_1, q_2, \ldots, q_n$. The set of all such wave functions with unit norm spans a Hilbert space H. Every possible state of the system is represented by a function in this Hilbert space, and conversely, every vector in this Hilbert space represents a possible state of the system. In addition to depending on the coordinates $q_1, q_2, \ldots, q_n$, the wave function depends also on the time t, but the dependence on the qs and on t are essentially different. The Hilbert space H is formed with respect to the spatial coordinates $q_1, q_2, \ldots, q_n$ only, for example, the inner product is formed with respect to the qs only, and one wave function $\psi(q_1, q_2, \ldots, q_n)$ states its complete spatial dependence. On the other hand the states of the system at different instants of time $t_1, t_2, \ldots$ are given by the different wave functions $\psi_1(q_1, q_2, \ldots, q_n), \psi_2(q_1, q_2, \ldots, q_n) \ldots$ of the Hilbert space.

Problems

5.1 Prove the three main properties of the dot product given by Eq. (5.7).

5.2 Show that the points on a line V passing through the origin in E_3 form a linear vector space under the addition and scalar multiplication operations for vectors in E_3.
Hint: The points of V satisfy parametric equations of the form

$$x_1 = at, \quad x_2 = bt, \quad x_3 = ct, \qquad -\infty < t < \infty.$$

5.3 Do all Hermitian 2×2 matrices form a vector space under addition? Is there any requirement on the scalars that multiply them?

5.4 Let V be the set of all points (x_1, x_2) in E_2 that lie in the first quadrant; that is, such that $x_1 \geq 0$ and $x_2 \geq 0$. Show that the set V fails to be a vector space under the operations of addition and scalar multiplication.
Hint: Consider $\mathbf{u} = (1, 1)$ which lies in V. Now form the scalar multiplication $(-1)\mathbf{u} = (-1, -1)$; where is this point located?

5.5 Show that the set W of all 2×2 matrices having zeros on the principal diagonal is a subspace of the vector space M_{22} of all 2×2 matrices.

5.6 Show that $|W\rangle = (4, -1, 8)$ is not a linear combination of $|U\rangle = (1, 2, -1)$ and $|V\rangle = (6, 4, 2)$.

5.7 Show that the following three vectors in E_3 cannot serve as base vectors of E_3:

$$|1\rangle = (1, 1, 2), |2\rangle = (1, 0, 1), \quad \text{and} \quad |3\rangle = (2, 1, 3).$$

5.8 Determine which of the following lie in the space spanned by $|f\rangle = \cos^2 x$ and $|g\rangle = \sin^2 x$: (a) cos $2x$; $(b) 3 + x^2$; $(c) 1$; $(d) \sin x$.

5.9 Determine whether the three vectors

$$|1\rangle = (1, -2, 3), \quad |2\rangle = (5, 6, -1), \quad |3\rangle = (3, 2, 1)$$

are linearly dependent or independent.

5.10 Given the following three vectors from the vector space of real 2×2 matrices:

$$|1\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad |3\rangle = \begin{pmatrix} -2 & -1 \\ 0 & -2 \end{pmatrix},$$

determine whether they are linearly dependent or independent.

5.11 If $S = \{|1\rangle, |2\rangle, \ldots, |n\rangle\}$ is a basis for a vector space V, show that every set with more than n vectors is linearly dependent.

5.12 Show that any two bases for a finite-dimensional vector space have the same number of vectors.

5.13 Consider the vector space E_3 with the Euclidean inner product. Apply the Gram–Schmidt process to transform the basis

$$|1\rangle = (1, 1, 1), \quad |2\rangle = (0, 1, 1), \quad |3\rangle = (0, 0, 1)$$

into an orthonormal basis.

5.14 Consider the two linearly independent vectors of Example 5.10:

$$|U\rangle = (3 - 4i)|1\rangle + (5 - 6i)|2\rangle,$$

$$|W\rangle = (1 - i)|1\rangle + (2 - 3i)|2\rangle,$$

where $|1\rangle$ and $|2\rangle$ are an orthonormal basis. Apply the Gram–Schmidt process to transform the two vectors into an orthonormal basis.

5.15 Show that the eigenvalue of the square of an operator is the square of the eigenvalue of the operator.

5.16 Show that if, for a given $\underset{\sim}{A}$, both operators $\underset{\sim}{A}_L^{-1}$ and $\underset{\sim}{A}_R^{-1}$ exist, then

$$\underset{\sim}{A}_L^{-1} = \underset{\sim}{A}_R^{-1} \equiv \underset{\sim}{A}^{-1}.$$

5.17 Show that if a unitary operator $\underset{\sim}{U}$ can be written in the form $\underset{\sim}{U} = 1 + ie\underset{\sim}{F}$, where e is a real infinitesimally small number, then the operator $\underset{\sim}{F}$ is Hermitian.

5.18 Show that the differential operator

$$\underset{\sim}{p} = \frac{\hbar}{i}\frac{d}{dx}$$

is linear and Hermitian in the space of all differentiable wave functions $\phi(x)$ that, say, vanish at both ends of an interval (a, b).

5.19 The translation operator $T(a)$ is defined to be such that $T(a)\phi(x) = \phi(x + a)$. Show that:

(*a*) $T(a)$ may be expressed in terms of the operator

$$\underset{\sim}{p} = \frac{\hbar}{i}\frac{d}{dx};$$

(*b*) $T(a)$ is unitary.

5.21 Verify that:

(*a*) $\dfrac{2}{L}\displaystyle\int_0^L \sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L}\,dx = \delta_{mn}.$

(*b*) $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^{2\pi} e^{i(m-n)}\,dx = \delta_{mn}.$