1

Mathematical Introduction

The aim of this book is to provide you with an introduction to quantum mechanics, starting from its axioms. It is the aim of this chapter to equip you with the necessary mathematical machinery. All the math you will need is developed here, starting from some basic ideas on vectors and matrices that you are assumed to know. Numerous examples and exercises related to classical mechanics are given, both to provide some relief from the math and to demonstrate the wide applicability of the ideas developed here. The effort you put into this chapter will be well worth your while: not only will it prepare you for this course, but it will also unify many ideas you may have learned piecemeal. To really learn this chapter, you must, as with any other chapter, work out the problems.

1.1. Linear Vector Spaces: Basics

In this section you will be introduced to *linear vector spaces*. You are surely familiar with the arrows from elementary physics encoding the magnitude and direction of velocity, force, displacement, torque, etc. You know how to add them and multiply them by scalars and the rules obeyed by these operations. For example, you know that scalar multiplication is associative: the multiple of a sum of two vectors is the sum of the multiples. What we want to do is abstract from this simple case a set of basic features or axioms, and say that any set of objects obeying the same forms a linear vector space. The cleverness lies in deciding which of the properties to keep in the generalization. If you keep too many, there will be no other examples; if you keep too few, there will be no interesting results to develop from the axioms.

The following is the list of properties the mathematicians have wisely chosen as requisite for a vector space. As you read them, please compare them to the world of arrows and make sure that these are indeed properties possessed by these familiar vectors. But note also that conspicuously missing are the requirements that every vector have a magnitude and direction, which was the first and most salient feature drilled into our heads when we first heard about them. So you might think that dropping this requirement, the baby has been thrown out with the bath water. However, you will have ample time to appreciate the wisdom behind this choice as

you go along and see a great unification and synthesis of diverse ideas under the heading of vector spaces. You will see examples of vector spaces that involve entities that you cannot intuitively perceive as having either a magnitude or a direction. While you should be duly impressed with all this, remember that it does not hurt at all to think of these generalizations in terms of arrows and to use the intuition to prove theorems or at the very least anticipate them.

Definition 1. A linear vector space $\mathbb{V}$ *is a collection of objects* $|1\rangle$, $|2\rangle, \ldots, |V\rangle, \ldots, |W\rangle, \ldots,$ called vectors, for which there exists

1. A definite rule for forming the vector sum, denoted $|V\rangle + |W\rangle$
2. A definite rule for multiplication by scalars $a, b, \ldots$, denoted $a|V\rangle$ with the following features:

- The result of these operations is another element of the space, a feature called *closure*: $|V\rangle + |W\rangle \in \mathbb{V}$.
- Scalar multiplication is *distributive in the vectors*: $a(|V\rangle + |W\rangle) = a|V\rangle + a|W\rangle$.
- Scalar multiplication is *distributive in the scalars*: $(a+b)|V\rangle = a|V\rangle + b|V\rangle$.
- Scalar multiplication is *associative*: $a(b|V\rangle) = ab|V\rangle$.
- Addition is *commutative*: $|V\rangle + |W\rangle = |W\rangle + |V\rangle$.
- Addition is *associative*: $|V\rangle + (|W\rangle + |Z\rangle) = (|V\rangle + |W\rangle) + |Z\rangle$.
- There exist a *null vector* $|0\rangle$ obeying $|V\rangle + |0\rangle = |V\rangle$.
- For every vector $|V\rangle$ there exists an *inverse under addition*, $|-V\rangle$, such that $|V\rangle + |-V\rangle = |0\rangle$.

There is a good way to remember all of these; *do what comes naturally*.

Definition 2. The numbers $a, b, \ldots$ are called the *field* over which the vector space is defined.

If the field consists of all real numbers, we have a *real vector space*, if they are complex, we have a *complex vector space*. The vectors themselves are neither real or complex; the adjective applies only to the scalars.
  Let us note that the above axioms imply

- $|0\rangle$ is unique, i.e., if $|0'\rangle$ has all the properties of $|0\rangle$, then $|0\rangle = |0'\rangle$.
- $0|V\rangle = |0\rangle$.
- $|-V\rangle = -|V\rangle$.
- $|-V\rangle$ is the unique additive inverse of $|V\rangle$.

The proofs are left as to the following exercise. You don't have to know the proofs, but you do have to know the statements.

Exercise 1.1.1. Verify these claims. For the first consider $|0\rangle + |0'\rangle$ and use the advertised properties of the two null vectors in turn. For the second start with $|0\rangle = (0+1)|V\rangle + |-V\rangle$. For the third, begin with $|V\rangle + (-|V\rangle) = 0|V\rangle = |0\rangle$. For the last, let $|W\rangle$ also satisfy $|V\rangle + |W\rangle = |0\rangle$. Since $|0\rangle$ is unique, this means $|V\rangle + |W\rangle = |V\rangle + |-V\rangle$. Take it from here.
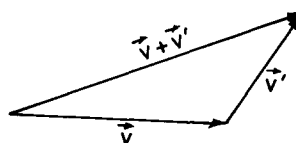
Figure 1.1. The rule for vector addition. Note that it obeys axioms (i)–(iii).

Exercise 1.1.2. Consider the set of all entities of the form (a, b, c) where the entries are real numbers. Addition and scalar multiplication are defined as follows:

$$(a, b, c) + (d, e, f) = (a, +d, b+e, c+f)$$

$$\alpha(a, b, c) = (\alpha a, \alpha b, \alpha c).$$

Write down the null vector and inverse of (a, b, c). Show that vectors of the form $(a, b, 1)$ do not form a vector space.

Observe that we are using a new symbol $|V\rangle$ to denote a generic vector. This object is called *ket V* and this nomenclature is due to Dirac whose notation will be discussed at some length later. We do not purposely use the symbol $\vec{V}$ to denote the vectors as the first step in weaning you away from the limited concept of the vector as an arrow. You are however not discouraged from associating with $|V\rangle$ the arrow-like object till you have seen enough vectors that are not arrows and are ready to drop the crutch.

You were asked to verify that the set of arrows qualified as a vector space as you read the axioms. Here are some of the key ideas you should have gone over. The vector space consists of arrows, typical ones being $\vec{V}$ and $\vec{V}'$. The rule for addition is familiar: take the tail of the second arrow, put it on the tip of the first, and so on as in Fig. 1.1.

Scalar multiplication by a corresponds to stretching the vector by a factor a. This is a real vector space since stretching by a complex number makes no sense. (If a is negative, we interpret it as changing the direction of the arrow as well as rescaling it by $|a|$.) Since these operations acting on arrows give more arrows, we have closure. Addition and scalar multiplication clearly have all the desired associative and distributive features. The null vector is the arrow of zero length, while the inverse of a vector is the vector reversed in direction.

So the set of all arrows qualifies as a vector space. But we cannot tamper with it. For example, the set of all arrows with positive z-components do not form a vector space: there is no inverse.

Note that so far, no reference has been made to magnitude or direction. The point is that while the arrows have these qualities, members of a vector space need not. This statement is pointless unless I can give you examples, so here are two.

Consider the set of all 2×2 matrices. We know how to add them and multiply them by scalars (multiply all four matrix elements by that scalar). The corresponding rules obey closure, associativity, and distributive requirements. The null matrix has all zeros in it and the inverse under *addition* of a matrix is the matrix with all elements negated. You must agree that here we have a genuine vector space consisting of things which don't have an obvious length or direction associated with them. When we want to highlight the fact that the matrix M is an element of a vector space, we may want to refer to it as, say, ket number 4 or: $|4\rangle$.

As a second example, consider all functions $f(x)$ defined in an interval $0 \le x \le L$. We define scalar multiplication by a simply as $af(x)$ and addition as pointwise addition: the sum of two functions f and g has the value $f(x) + g(x)$ at the point x. The null function is zero everywhere and the additive inverse of f is $-f$.

Exercise 1.1.3. Do functions that vanish at the end points $x = 0$ and $x = L$ form a vector space? How about *periodic functions* obeying $f(0) = f(L)$? How about functions that obey $f(0) = 4$? If the functions do not qualify, list the things that go wrong.

The next concept is that of *linear independence* of a set of vectors $|1\rangle, |2\rangle \ldots |n\rangle$. First consider a linear relation of the form

$$\sum_{i=1}^{n} a_i |i\rangle = |0\rangle \tag{1.1.1}$$

We may assume without loss of generality that the left-hand side does not contain any multiple of $|0\rangle$, for if it did, it could be shifted to the right, and combined with the $|0\rangle$ there to give $|0\rangle$ once more. (We are using the fact that any multiple of $|0\rangle$ equals $|0\rangle$.)

Definition 3. The set of vectors is said to be *linearly independent* if the only such linear relation as Eq. (1.1.1) is the trivial one with all $a_i = 0$. If the set of vectors is not linearly independent, we say they are *linearly dependent*.

Equation (1.1.1) tells us that it is not possible to write any member of the linearly independent set in terms of the others. On the other hand, if the set of vectors is linearly dependent, such a relation will exist, and it must contain at least two nonzero coefficients. Let us say $a_3 \ne 0$. Then we could write

$$|3\rangle = \sum_{i=1, \ne 3}^{n} \frac{-a_i}{a_3} |i\rangle \tag{1.1.2}$$

thereby expressing $|3\rangle$ in terms of the others.

As a concrete example, consider two nonparallel vectors $|1\rangle$ and $|2\rangle$ in a plane. These form a linearly independent set. There is no way to write one as a multiple of the other, or equivalently, no way to combine them to get the null vector. On the other hand, if the vectors are parallel, we can clearly write one as a multiple of the other or equivalently play them against each other to get 0.

Notice I said 0 and not $|0\rangle$. This is, strictly speaking, incorrect since a set of vectors can only add up to a vector and not a number. It is, however, common to represent the null vector by 0.

Suppose we bring in a third vector $|3\rangle$ also in the plane. If it is parallel to either of the first two, we already have a linearly dependent set. So let us suppose it is not. But even now the three of them are *linearly dependent*. This is because we can write one of them, say $|3\rangle$, as a linear combination of the other two. To find the combination, draw a line from the tail of $|3\rangle$ in the direction of $|1\rangle$. Next draw a line antiparallel to $|2\rangle$ from the tip of $|3\rangle$. These lines will intersect since $|1\rangle$ and $|2\rangle$ are

not parallel by assumption. The intersection point P will determine how much of $|1\rangle$ and $|2\rangle$ we want: we go from the tail of $|3\rangle$ to P using the appropriate multiple of $|1\rangle$ and go from P to the tip of $|3\rangle$ using the appropriate multiple of $|2\rangle$.

Exercise 1.1.4. Consider three elements from the vector space of real 2×2 matrices:

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

Are they linearly independent? Support your answer with details. (Notice we are calling these matrices vectors and using kets to represent them to emphasize their role as elements of a vector space.

Exercise 1.1.5. Show that the following row vectors are linearly dependent: $(1, 1, 0)$, $(1, 0, 1)$, and $(3, 2, 1)$. Show the opposite for $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$.

Definition 4. A vector space has *dimension n* if it can accommodate a maximum of n linearly independent vectors. It will be denoted by $\mathbb{V}^n(R)$ if the field is real and by $\mathbb{V}^n(C)$ if the field is complex.

In view of the earlier discussions, the plane is two-dimensional and the set of all arrows not limited to the plane define a three-dimensional vector space. How about 2×2 matrices? They form a four-dimensional vector space. Here is a proof. The following vectors are linearly independent:

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

since it is impossible to form linear combinations of any three of them to give the fourth any three of them will have a zero in the one place where the fourth does not. So the space is at least four-dimensional. Could it be bigger? No, since any arbitrary 2×2 matrix can be written in terms of them:

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

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*.

Theorem 1. Any vector $|V\rangle$ in an n-dimensional space can be written as a linearly combination of n linearly independent vectors $|1\rangle \ldots |n\rangle$.

The proof is as follows: if there were a vector $|V\rangle$ for which this were not possible, it would join the given set of vectors and form a set of $n+1$ linearly independent vectors, which is not possible in an n-dimensional space by definition.

Definition 5. A set of n linearly independent vectors in an n-dimensional space is called a *basis*.

Thus we can write, on the strength of the above

$$|V\rangle = \sum_{i=1}^{n} v_i |i\rangle \qquad (1.1.3)$$

where the vectors $|i\rangle$ form a basis.

Definition 6. The coefficients of expansion v_i of a vector in terms of a linearly independent basis ($|i\rangle$) are called the *components of the vector in that basis*.

Theorem 2. The expansion in Eq. (1.1.1) is unique.

Suppose the expansion is not unique. We must then have a second expansion:

$$|V\rangle = \sum_{i=1}^{n} v_i' |i\rangle \qquad (1.1.4)$$

Subtracting Eq. (1.1.4) from Eq. (1.1.3) (i.e., multiplying the second by the scalar -1 and adding the two equations) we get

$$|0\rangle = \sum_{i} (v_i - v_i') |i\rangle \qquad (1.1.5)$$

which implies that

$$v_i = v_i' \qquad (1.1.6)$$

since the basis vectors are linearly independent and only a trivial linear relation between them can exist. Note that given a basis the components are unique, but if we change the basis, the components will change. We refer to $|V\rangle$ as the vector in the abstract, having an existence of its own and satisfying various relations involving other vectors. When we choose a basis the vectors assume concrete forms in terms of their components and the relation between vectors is satisfied by the components. Imagine for example three arrows in the plane, $\vec{A}$, $\vec{B}$, $\vec{C}$ satisfying $\vec{A} + \vec{B} = \vec{C}$ according to the laws for adding arrows. So far no basis has been chosen and we do not need a basis to make the statement that the vectors from a closed triangle. Now we choose a basis and write each vector in terms of the components. The components will satisfy $C_i = A_i + B_i$, $i = 1, 2$. If we choose a different basis, the components will change in numerical value, but the relation between them expressing the equality of $\vec{C}$ to the sum of the other two will still hold between the new set of components.

In the case of nonarrow vectors, adding them in terms of components proceeds
as in the elementary case thanks to the axioms. If

7

MATHEMATICAL
INTRODUCTION

$$|V\rangle = \sum_i v_i |i\rangle \quad \text{and} \qquad (1.1.7)$$

$$|W\rangle = \sum_i w_i |i\rangle \quad \text{then} \qquad (1.1.8)$$

$$|V\rangle + |W\rangle = \sum_i (v_i + w_i)|i\rangle \qquad (1.1.9)$$

where we have used the axioms to carry out the regrouping of terms. Here is the
conclusion:

To add two vectors, add their components.

There is no reference to taking the tail of one and putting it on the tip of the
other, etc., since in general the vectors have no head or tail. Of course, if we are
dealing with arrows, we can add them either using the tail and tip routine or by
simply adding their components in a basis.

In the same way, we have:

$$a|V\rangle = a \sum_i v_i |i\rangle = \sum_i a v_i |i\rangle \qquad (1.1.10)$$

In other words,

To multiply a vector by a scalar, multiply all its components by the scalar.

1.2. Inner Product Spaces

The matrix and function examples must have convinced you that we can have
a vector space with no preassigned definition of length or direction for the elements.
However, we can make up quantities that have the same properties that the lengths
and angles do in the case of arrows. The first step is to define a sensible analog of
the dot product, for in the case of arrows, from the dot product

$$\vec{A} \cdot \vec{B} = |A||B| \cos \theta \qquad (1.2.1)$$

we can read off the length of say $\vec{A}$ as $\sqrt{|A| \cdot |A|}$ and the cosine of the angle between
two vectors as $\vec{A} \cdot \vec{B}/|A||B|$. Now you might rightfully object: how can you use the dot
product to define the length and angles, if the dot product itself requires knowledge of
the lengths and angles? The answer is this. Recall that the dot product has a second
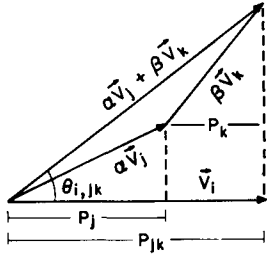
Figure 1.2. Geometrical proof that the dot product obeys axiom (iii) for an inner product. The axiom requires that the projections obey $P_k + P_j = P_{jk}$.

equivalent expression in terms of the components:

$$\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y + A_z B_z \qquad (1.2.2)$$

Our goal is to define a similar formula for the general case where we *do* have the notion of components in a basis. To this end we recall the main features of the above dot product:

1. $\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A}$ (symmetry)
2. $\vec{A} \cdot \vec{A} \geq 0$ 0 *iff* $\vec{A} = 0$ (positive semidefiniteness)
3. $\vec{A} \cdot (b\vec{B} + c\vec{C}) = b\vec{A} \cdot \vec{B} + c\vec{A} \cdot \vec{C}$ (linearity)

    The linearity of the dot product is illustrated in Fig. 1.2.

    We want to invent a generalization called the *inner product* or *scalar product* between any two vectors $|V\rangle$ and $|W\rangle$. We denote it by the symbol $\langle V|W\rangle$. It is once again a number (generally complex) dependent on the two vectors. We demand that it obey the following axioms:

- $\langle V|W\rangle = \langle W|V\rangle^*$ (skew-symmetry)
- $\langle V|V\rangle \geq 0$ 0 *iff* $|V\rangle = |0\rangle$ (positive semidefiniteness)
- $\langle V|(a|W\rangle + b|Z\rangle) \equiv \langle V|aW + bZ\rangle = a\langle V|W\rangle + b\langle V|Z\rangle$ (linearity in ket)

Definition 7. A vector space with an inner product is called an *inner product space.*

    Notice that we have not yet given an explicit rule for actually evaluating the scalar product, we are merely demanding that any rule we come up with must have these properties. With a view to finding such a rule, let us familiarize ourselves with the axioms. The first differs from the corresponding one for the dot product and makes the inner product sensitive to the order of the two factors, with the two choices leading to complex conjugates. In a real vector space this axioms states the symmetry of the dot product under exchange of the two vectors. For the present, let us note that this axiom ensures that $\langle V|V\rangle$ is real.

    The second axiom says that $\langle V|V\rangle$ is not just real but also positive semidefinite, vanishing only if the vector itself does. If we are going to define the length of the vector as the square root of its inner product with itself (as in the dot product) this quantity had better be real and positive for all nonzero vectors.

The last axiom expresses the linearity of the inner product when a linear super-position $a|W\rangle + b|Z\rangle \equiv |aW+bZ\rangle$ appears as the second vector in the scalar product. We have discussed its validity for the arrows case (Fig. 1.2).

What if the first factor in the product is a linear superposition, i.e., what is $\langle aW+bZ|V\rangle$? This is determined by the first axiom:

$$\langle aW+bZ|V\rangle = \langle V|aW+bZ\rangle^* \quad \text{by } BI$$

$$= (a\langle V|W\rangle + b\langle V|Z\rangle)^*$$

$$= a^*\langle V|W\rangle^* + b^*\langle V|Z\rangle^*$$

$$= a^*\langle W|V\rangle + b^*\langle Z|V\rangle \qquad (1.2.3)$$

which expresses the *antilinearity* of the inner product with respect to the first factor in the inner product. In other words, the inner product of a linear superposition with another vector is the corresponding superposition of inner products if the superposition occurs in the second factor, while it is the superposition with all coefficients conjugated if the superposition occurs in the first factor. This asymmetry, unfamiliar in real vector spaces, is here to stay and you will get used to it as you go along.

Let us continue with inner products. Even though we are trying to shed the restricted notion of a vector as an arrow and seeking a corresponding generalization of the dot product, we still use some of the same terminology.

Definition 8. We say that two vectors are *orthogonal* or perpendicular if their inner product vanishes.

Definition 9. We will refer to $\sqrt{\langle V|V\rangle} \equiv |V|$ as the *norm* or length of the vector. A *normalized vector* has unit norm.

Definition 10. A set of basis vectors all of unit norm, which are pairwise orthogonal will be called an *orthonormal basis*.

We will also frequently refer to the inner or scalar product as the dot product.

We are now ready to obtain a concrete formula for the inner product in terms of the components. Given $|V\rangle$ and $|W\rangle$

$$|V\rangle = \sum_i |i\rangle$$

$$|W\rangle = \sum_j |j\rangle$$

we follow the axioms obeyed by the inner product to obtain:

$$\langle V|W\rangle = \sum_i \sum_j v_i^* w_j \langle i|j\rangle \qquad (1.2.4)$$

To go any further we have to know $\langle i|j\rangle$, the inner product between basis vectors. That depends on the details of the basis vectors and all we know for sure is that

they are linearly independent. This situation exists for arrows as well. Consider a two-dimensional problem where the basis vectors are two linearly independent but nonperpendicular vectors. If we write all vectors in terms of this basis, the dot product of any two of them will likewise be a double sum with four terms (determined by the four possible dot products between the basis vectors) as well as the vector components. However, if we use an orthonormal basis such as $\vec{i}, \vec{j}$, only diagonal terms like $\langle i | i \rangle$ will survive and we will get the familiar result $\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y$ depending only on the components.

For the more general nonarrow case, we invoke Theorem 3.

Theorem 3 (Gram-Schmidt). Given a linearly independent basis we can form linear combinations of the basis vectors to obtain an orthonormal basis.

Postponing the proof for a moment, *let us assume that the procedure has been implemented and that the current basis is orthonormal*:

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

where δ_{ij} is called the *Kronecker delta symbol.* Feeding this into Eq. (1.2.4) we find the double sum collapses to a single one due to the Kronecker delta, to give

$$\langle V | W \rangle = \sum_i v_i^* w_i \qquad (1.2.5)$$

This is the form of the inner product we will use from now on.

You can now appreciate the first axiom; but for the complex conjugation of the components of the first vector, $\langle V | V \rangle$ would not even be real, not to mention positive. But now it is given by

$$\langle V | V \rangle = \sum_i |v_i|^2 \geq 0 \qquad (1.2.6)$$

and vanishes only for the null vector. This makes it sensible to refer to $\langle V | V \rangle$ as the length or norm squared of a vector.

Consider Eq. (1.2.5). Since the vector $|V\rangle$ is uniquely specified by its components in a given basis, we may, in this basis, write it as a column vector:

$$|V\rangle \rightarrow \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \text{in this basis} \qquad (1.2.7)$$

$$|W\rangle \rightarrow \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad \text{in this basis} \tag{1.2.8}$$

The inner product $\langle V|W\rangle$ is given by the matrix product of the transpose conjugate of the column vector representing $|V\rangle$ with the column vector representing $|W\rangle$:

$$\langle V|W\rangle = [v_1^*, v_2^*, \ldots, v_n^*] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \tag{1.2.9}$$

1.3. Dual Spaces and the Dirac Notation

There is a technical point here. The inner product is a number we are trying to generate from two kets $|V\rangle$ and $|W\rangle$, which are both represented by column vectors in some basis. Now there is no way to make a number out of two columns by direct matrix multiplication, but there is a way to make a number by matrix multiplication of a row times a column. Our trick for producing a number out of two columns has been to associate a unique row vector with one column (its transpose conjugate) and form its matrix product with the column representing the other. This has the feature that the answer depends on which of the two vectors we are going to convert to the row, the two choices ($\langle V|W\rangle$ and $\langle W|V\rangle$) leading to answers related by complex conjugation as per axiom 1(h).

But one can also take the following alternate view. Column vectors are concrete manifestations of an abstract vector $|V\rangle$ or ket in a basis. We can also work backward and go from the column vectors to the abstract kets. But then it is similarly possible to work backward and associate with each *row vector* an abstract object $\langle W|$, called *bra-W*. Now we can name the bras as we want but let us do the following. Associated with every ket $|V\rangle$ is a column vector. Let us take its *adjoint*, or transpose conjugate, and form a row vector. The abstract bra associated with this will bear the same label, i.e., it be called $\langle V|$. Thus there are two vector spaces, the space of kets and a dual space of bras, with a ket for every bra and vice versa (the components being related by the adjoint operation). 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 $|i\rangle$ 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 bra $\langle i|$ is a row vector with all zeros except for a 1 in the ith column.

All this may be summarized as follows:

$$|V\rangle \leftrightarrow \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \leftrightarrow [v_1^*, v_2^*, \dots v_n^*] \leftrightarrow \langle V| \tag{1.3.1}$$

where $\leftrightarrow$ means "within a basis."

There is, however, nothing wrong with the first viewpoint of associating a scalar product with a pair of columns or kets (making no reference to another dual space) and living with the asymmetry between the first and second vector in the inner product (which one to transpose conjugate?). If you found the above discussion heavy going, you can temporarily ignore it. The only thing you must remember is that in the case of a general nonarrow vector space:

- Vectors can still be assigned components in some orthonormal basis, just as with arrows, but these may be complex.
- The inner product of any two vectors is given in terms of these components by Eq. (1.2.5). This product obeys all the axioms.

1.3.1. Expansion of Vectors in an Orthonormal Basis

Suppose we wish to expand a vector $|V\rangle$ in an orthonormal basis. To find the components that go into the expansion we proceed as follows. We take the dot product of both sides of the assumed expansion with $|j\rangle$: (or $\langle j|$ if you are a purist)

$$|V\rangle = \sum_i v_i |i\rangle \tag{1.3.2}$$

$$\langle j|V\rangle = \sum_i v_i \underbrace{\langle j|i\rangle}_{\delta_{ij}} \tag{1.3.3}$$

$$= v_j \tag{1.3.4}$$

i.e., the find the jth component of a vector we take the dot product with the jth unit vector, exactly as with arrows. Using this result we may write

$$|V\rangle = \sum_i |i\rangle\langle i|V\rangle \tag{1.3.5}$$

Let us make sure the basis vectors look as they should. If we set $|V\rangle = |j\rangle$ in Eq. (1.3.5), we find the correct answer: the ith component of the jth basis vector is δ_{ij}. Thus for example the column representing basis vector number 4 will have a 1 in the 4th row and zero everywhere else. The abstract relation

$$|V\rangle = \sum_i v_i |i\rangle \tag{1.3.6}$$

becomes in this basis

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots v_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{1.3.7}$$

1.3.2. Adjoint Operation

We have seen that we may pass from the column representing a ket to the row representing the corresponding bra by the adjoint operation, i.e., transpose conjugation. Let us now ask: if $\langle V|$ is the bra corresponding to the ket $|V\rangle$ what bra corresponds to $a|V\rangle$ where a is some scalar? By going to any basis it is readily found that

$$a|V\rangle \to \begin{bmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{bmatrix} \to [a^*v_1^*, a^*v_2^*, \ldots, a^*v_n^*] \to \langle V|a^* \tag{1.3.8}$$

It is customary to write $a|V\rangle$ as $|aV\rangle$ and the corresponding bra as $\langle aV|$. What we have found is that

$$\langle aV| = \langle V|a^* \tag{1.3.9}$$

Since the relation between bras and kets is linear we can say that if we have an equation among kets such as

$$a|V\rangle = b|W\rangle + c|Z\rangle + \cdots \tag{1.3.10}$$

this implies another one among the corresponding bras:

$$\langle V|a^* = \langle W|b^* + \langle Z|c^* + \cdots \tag{1.3.11}$$

The two equations above are said to be *adjoints of each other*. Just as any equation involving complex numbers implies another obtained by taking the complex conjugates of both sides, an equation between (bras) kets implies another one between (kets) bras. If you think in a basis, you will see that this follows simply from the fact that if two columns are equal, so are their transpose conjugates.

Here is the rule for taking the adjoint:

To take the adjoint of a linear equation relating kets (bras), replace every ket (bra) by its bra (ket) and complex conjugate all coefficients.

We can extend this rule as follows. Suppose we have an expansion for a vector:

$$|V\rangle = \sum_{i=1} v_i |i\rangle \qquad (1.3.12)$$

in terms of basis vectors. The adjoint is

$$\langle V| = \sum_{i=1} \langle i| v_i^*$$

Recalling that $v_i = \langle i|V\rangle$ and $v_i^* = \langle V|i\rangle$, it follows that the adjoint of

$$|V\rangle = \sum_{i=1} |i\rangle\langle i|V\rangle \qquad (1.3.13)$$

is

$$\langle V| = \sum_{i=1} \langle V|i\rangle\langle i| \qquad (1.3.14)$$

from which comes the rule:

To take the adjoint of an equation involving bras and kets and coefficients, reverse the order of all factors, exchanging bras and kets and complex conjugating all coefficients.

Gram–Schmidt Theorem

Let us now take up the Gram–Schmidt procedure for converting a linearly independent basis into an orthonormal one. The basic idea can be seen by a simple example. Imagine the two-dimensional space of arrows in a plane. Let us take two nonparallel vectors, which qualify as a basis. To get an orthonormal basis out of these, we do the following:

● Rescale the first by its own length, so it becomes a unit vector. This will be the first basis vector.
● Subtract from the second vector its projection along the first, leaving behind only the part perpendicular to the first. (Such a part will remain since by assumption the vectors are nonparallel.)
● Rescale the left over piece by its own length. We now have the second basis vector: it is orthogonal to the first and of unit length.

This simple example tells the whole story behind this procedure, which will now be discussed in general terms in the Dirac notation.

Let $|I\rangle$, $|II\rangle$, ... be a linearly independent basis. The first vector of the orthonormal basis will be

$$|1\rangle = \frac{|I\rangle}{|I|} \quad \text{where} \quad |I| = \sqrt{\langle I|I\rangle}$$

Clearly

$$\langle 1|1\rangle = \frac{\langle I|I\rangle}{|I|^2} = 1$$

As for the second vector in the basis, consider

$$|2'\rangle = |II\rangle - |1\rangle\langle 1|II\rangle$$

which is $|II\rangle$ minus the part pointing along the first unit vector. (Think of the arrow example as you read on.) Not surprisingly it is orthogonal to the latter:

$$\langle 1|2'\rangle = \langle 1|II\rangle - \langle 1|1\rangle\langle 1|II\rangle = 0$$

We now divide $|2'\rangle$ by its norm to get $|2\rangle$ which will be orthogonal to the first and normalized to unity. Finally, consider

$$|3'\rangle = |III\rangle - |1\rangle\langle 1|III\rangle - |2\rangle\langle 2|III\rangle$$

which is orthogonal to both $|1\rangle$ and $|2\rangle$. Dividing by its norm we get $|3\rangle$, the third member of the orthogonal basis. There is nothing new with the generation of the rest of the basis.

Where did we use the linear independence of the original basis? What if we had started with a linearly dependent basis? Then at some point a vector like $|2'\rangle$ or $|3'\rangle$ would have vanished, putting a stop to the whole procedure. On the other hand, linear independence will assure us that such a thing will never happen since it amounts to having a nontrivial linear combination of linearly independent vectors that adds up the null vector. (Go back to the equations for $|2'\rangle$ or $|3'\rangle$ and satisfy yourself that these are linear combinations of the old basis vectors.)

Exercise 1.3.1. Form an orthogonal basis in two dimensions starting with $\vec{A} = 3\vec{i} + 4\vec{j}$ and $\vec{B} = 2\vec{i} - 6\vec{j}$. Can you generate another orthonormal basis starting with these two vectors? If so, produce another.

Exercise 1.3.2. Show how to go from the basis

$$|I\rangle = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \qquad |II\rangle = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \qquad |III\rangle = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}$$

to the orthonormal basis

$$|1\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad |2\rangle = \begin{bmatrix} 0 \\ 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \qquad |III\rangle = \begin{bmatrix} 0 \\ -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

When we first learn about dimensionality, we associate it with the number of perpendicular directions. In this chapter we defined in terms of the maximum number of linearly independent vectors. The following theorem connects the two definitions.

Theorem 4. The dimensionality of a space equals $n_\perp$, the maximum number of mutually orthogonal vectors in it.

To show this, first note that any mutually orthogonal set is also linearly independent. Suppose we had a linear combination of orthogonal vectors adding up to zero. By taking the dot product of both sides with any one member and using the orthogonality we can show that the coefficient multiplying that vector had to vanish. This can clearly be done for all the coefficients, showing the linear combination is trivial.

Now $n_\perp$ can only be equal to, greater than or lesser than n, the dimensionality of the space. The Gram–Schmidt procedure eliminates the last case by explicit construction, while the linear independence of the perpendicular vectors rules out the penultimate option.

Schwarz and Triangle Inequalities

Two powerful theorems apply to any inner product space obeying our axioms:

Theorem 5. The Schwarz Inequality

$$|\langle V | W \rangle| \leq |V| |W| \tag{1.3.15}$$

Theorem 6. The Triangle Inequality

$$|V + W| \leq |V| + |W| \tag{1.3.16}$$

The proof of the first will be provided so you can get used to working with bras and kets. The second will be left as an exercise.

Before proving anything, note that the results are obviously true for arrows: the *Schwarz inequality* says that the dot product of two vectors cannot exceed the product of their lengths and the *triangle inequality* says that the length of a sum cannot exceed the sum of the lengths. This is an example which illustrates the merits of thinking of abstract vectors as arrows and guessing what properties they might share with arrows. The proof will of course have to rely on just the axioms.

To prove the Schwarz inequality, consider axiom 1(i) applied to

$$|Z\rangle = |V\rangle - \frac{\langle W|V\rangle}{|W|^2}|W\rangle \qquad (1.3.17)$$

We get

$$\langle Z|Z\rangle = \langle V - \frac{\langle W|V\rangle}{|W|^2}W|V - \frac{\langle W|V\rangle}{|W|^2}W\rangle$$

$$= \langle V|V\rangle - \frac{\langle W|V\rangle\langle V|W\rangle}{|W|^2} - \frac{\langle W|V\rangle^*\langle W|V\rangle}{|W|^2}$$

$$+ \frac{\langle W|V\rangle^*\langle W|V\rangle\langle W|W\rangle}{|W|^4}$$

$$\geq 0 \qquad (1.3.18)$$

where we have used the antilinearity of the inner product with respect to the bra. Using

$$\langle W|V\rangle^* = \langle V|W\rangle$$

we find

$$\langle V|V\rangle > \frac{\langle W|V\rangle\langle V|W\rangle}{|W|^2} \qquad (1.3.19)$$

Cross-multiplying by $|W|^2$ and taking square roots, the result follows.

Exercise 1.3.3. When will this inequality be satisfied? Does this agree with you experience with arrows?

Exercise 1.3.4. Prove the triangle inequality starting with $|V+W|^2$. You must use $\mathrm{Re}\langle V|W\rangle \leq |\langle V|W\rangle|$ and the Schwarz inequality. Show that the final inequality becomes an equality only if $|V\rangle = a|W\rangle$ where a is a real positive scalar.

1.4. Subspaces

Definition 11. Given a vector space $\mathbb{V}$, a subset of its elements that form a vector space among themselves‡ is called a *subspace*. We will denote a particular subspace i of dimensionality n_i by $\mathbb{V}_i^{n_i}$.

‡ Vector addition and scalar multiplication are defined the same way in the subspace as in $\mathbb{V}$.

Example 1.4.1. In the space $\mathbb{V}^3(R)$, the following are some example of subspaces: (a) all vectors along the x axis, the space $\mathbb{V}_x^1$; (b) all vectors along the y axis, the space $\mathbb{V}_y^1$; (c) all vectors in the $x - y$ plane, the space $\mathbb{V}_{xy}^2$. Notice that all subspaces contain the null vector and that each vector is accompanied by its inverse to fulfill axioms for a vector space. Thus the set of all vectors along the positive x axis alone do not form a vector space. □

Definition 12. Given two subspaces $\mathbb{V}_i^{n_i}$ and $\mathbb{V}_j^{m_j}$, we define their sum $\mathbb{V}_i^{n_i} \oplus \mathbb{V}_j^{m_j} = \mathbb{V}_k^{m_k}$ as the set containing (1) all elements of $\mathbb{V}_i^{n_i}$, (2) all elements of $\mathbb{V}_j^{m_j}$, (3) all possible linear combinations of the above. But for the elements (3), closure would be lost.

Example 1.4.2. If, for example, $\mathbb{V}_x^1 \oplus \mathbb{V}_y^1$ contained only vectors along the x and y axes, we could, be adding two elements, one from each direction, generate one along neither. On the other hand, if we also included all linear combinations, we would get the correct answer, $\mathbb{V}_x^1 \oplus \mathbb{V}_y^1 = \mathbb{V}_{xy}^2$. □

Exercise 1.4.1. * In a space $\mathbb{V}^n$, prove that the set of all vectors $\{|V_\perp^1\rangle, |V_\perp^2\rangle, \dots\}$, orthogonal to any $|V\rangle \neq 0\rangle$, form a subspace $\mathbb{V}^{n-1}$.

Exercise 1.4.2. Suppose $\mathbb{V}_1^{n_1}$ and $\mathbb{V}_2^{n_2}$ are two subspaces such that any element of $\mathbb{V}_1$ is orthogonal to any element of $\mathbb{V}_2$. Show that the dimensionality of $\mathbb{V}_1 \oplus \mathbb{V}_2$ is $n_1 + n_2$. (Hint: Theorem 6.)

1.5. Linear Operators

An operator Ω is an instruction for transforming any given vector $|V\rangle$ into another, $|V'\rangle$. The action of the operator is represented as follows:

$$\Omega|V\rangle = |V'\rangle \tag{1.5.1}$$

One says that the operator Ω has transformed the ket $|V\rangle$ into the ket $|V'\rangle$. We will restrict our attention throughout to operators Ω that do not take us out of the vector space, i.e., if $|V\rangle$ is an element of a space $\mathbb{V}$, so is $|V'\rangle = \Omega|V\rangle$.

Operators can also act on bras:

$$\langle V'|\Omega = \langle V''| \tag{1.5.2}$$

We will only be concerned with *linear operators*, i.e., ones that obey the following rules:

$$\Omega\alpha|V_i\rangle = \alpha\Omega|V_i\rangle \tag{1.5.3a}$$

$$\Omega\{\alpha|V_i\rangle + \beta|V_j\rangle\} = \alpha\Omega|V_i\rangle + \beta\Omega|V_j\rangle \tag{1.5.3b}$$

$$\langle V_i|\alpha\Omega = \langle V_i|\Omega\alpha \tag{1.5.4a}$$

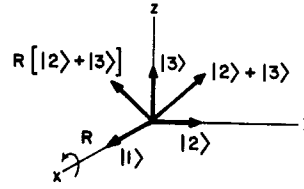$$(\langle V_i|\alpha + \langle V_j|\beta)\Omega = \alpha\langle V_i|\Omega + \beta\langle V_j|\Omega \tag{1.5.4b}$$

Figure 1.3. Action of the operator $R(\frac{1}{2}\pi\mathbf{i})$. Note that $R[|2\rangle+|3\rangle]=R|2\rangle+R|3\rangle$ as expected of a linear operator. (We will often refer to $R(\frac{1}{2}\pi\mathbf{i})$ as R if no confusion is likely.)

Example 1.5.1. The simplest operator is the identity operator, I, which carries the instruction:

$$I\rightarrow\text{Leave the vector alone!}$$

Thus,

$$I|V\rangle=|V\rangle \quad \text{for all kets } |V\rangle \tag{1.5.5}$$

and

$$\langle V|I=\langle V| \quad \text{for all bras } \langle V| \tag{1.5.6}$$

We next pass on to a more interesting operator on $\mathbb{V}^3(R)$:

$$R(\tfrac{1}{2}\pi\mathbf{i})\rightarrow\text{Rotate vector by } \tfrac{1}{2}\pi \text{ about the unit vector } \mathbf{i}$$

[More generally, $R(\boldsymbol{\theta})$ stands for a rotation by an angle $\theta=|\boldsymbol{\theta}|$ about the axis parallel to the unit vector $\hat{\theta}=\boldsymbol{\theta}/\theta$.] Let us consider the action of this operator on the three unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, which in our notation will be denoted by $|1\rangle$, $|2\rangle$, and $|3\rangle$ (see Fig. 1.3). From the figure it is clear that

$$R(\tfrac{1}{2}\pi\mathbf{i})|1\rangle=|1\rangle \tag{1.5.7a}$$

$$R(\tfrac{1}{2}\pi\mathbf{i})|2\rangle=|3\rangle \tag{1.5.7b}$$

$$R(\tfrac{1}{2}\pi\mathbf{i})|3\rangle=-|2\rangle \tag{1.5.7c}$$

Clearly $R(\frac{1}{2}\pi\mathbf{i})$ is linear. For instance, it is clear from the same figure that $R[|2\rangle+|3\rangle]=R|2\rangle+R|3\rangle$. $\qquad\square$

The nice feature of linear operators is that once their action on the basis vectors is known, their action on any vector in the space is determined. If

$$\Omega|i\rangle=|i'\rangle$$

for a basis $|1\rangle$, $|2\rangle$, ..., $|n\rangle$ in $\mathbb{V}^n$, then for any $|V\rangle=\sum v_i|i\rangle$

$$\Omega|V\rangle=\sum_i\Omega v_i|i\rangle=\sum_i v_i\Omega|i\rangle=\sum_i v_i|i'\rangle \tag{1.5.8}$$

This is the case in the example $\Omega = R(\frac{1}{2}\pi\mathbf{i})$. If

$$|V\rangle = v_1|1\rangle + v_2|2\rangle + v_3|3\rangle$$

is any vector, then

$$R|V\rangle = v_1 R|1\rangle + v_2 R|2\rangle + v_3 R|3\rangle = v_1|1\rangle + v_2|3\rangle - v_3|2\rangle$$

The *product of two operators* stands for the instruction that the instructions corresponding to the two operators be carried out in sequence

$$\Lambda\Omega|V\rangle = \Lambda(\Omega|V\rangle) = \Lambda|\Omega V\rangle \tag{1.5.9}$$

where $|\Omega V\rangle$ is the ket obtained by the action of Ω on $|V\rangle$. The order of the operators in a product is very important: in general,

$$\Omega\Lambda - \Lambda\Omega \equiv [\Omega, \Lambda]$$

called the *commutator* of Ω and Λ isn't zero. For example $R(\frac{1}{2}\pi\mathbf{i})$ and $R(\frac{1}{2}\pi\mathbf{j})$ do not commute, i.e., their commutator is nonzero.

Two useful identities involving commutators are

$$[\Omega, \Lambda\theta] = \Lambda[\Omega, \theta] + [\Omega, \Lambda]\theta \tag{1.5.10}$$

$$[\Lambda\Omega, \theta] = \Lambda[\Omega, \theta] + [\Lambda, \theta]\Omega \tag{1.5.11}$$

Notice that apart from the emphasis on ordering, these rules resemble the chain rule in calculus for the derivative of a product.

The *inverse* of Ω, denoted by Ω^{-1}, satisfies‡

$$\Omega\Omega^{-1} = \Omega^{-1}\Omega = I \tag{1.5.12}$$

Not every operator has an inverse. The condition for the existence of the inverse is given in Appendix A.1. The operator $R(\frac{1}{2}\pi\mathbf{i})$ has an inverse: it is $R(-\frac{1}{2}\pi\mathbf{i})$. The inverse of a product of operators is the product of the inverses in reverse:

$$(\Omega\Lambda)^{-1} = \Lambda^{-1}\Omega^{-1} \tag{1.5.13}$$

for only then do we have

$$(\Omega\Lambda)(\Omega\Lambda)^{-1} = (\Omega\Lambda)(\Lambda^{-1}\Omega^{-1}) = \Omega\Lambda\Lambda^{-1}\Omega^{-1} = \Omega\Omega^{-1} = I$$

1.6. Matrix Elements of Linear Operators

We are now accustomed to the idea of an abstract vector being represented in a basis by an n-tuple of numbers, called its components, in terms of which all vector

‡ In $\mathbb{V}^n(C)$ with n finite, $\Omega^{-1}\Omega = I \Leftrightarrow \Omega\Omega^{-1} = I$. Prove this using the ideas introduced toward the end of Theorem A.1.1., Appendix A.1.

operations can be carried out. We shall now see that in the same manner a linear operator can be represented in a basis by a set of n^2 numbers, written as an $n \times n$ matrix, and called its *matrix elements* in that basis. Although the matrix elements, just like the vector components, are basis dependent, they facilitate the computation of all basis-independent quantities, by rendering the abstract operator more tangible.

Our starting point is the observation made earlier, that the action of a linear operator is fully specified by its action on the basis vectors. If the basis vectors suffer a change

$$\Omega|i\rangle = |i'\rangle$$

(where $|i'\rangle$ is known), then any vector in this space undergoes a change that is readily calculable:

$$\Omega|V\rangle = \Omega \sum_i v_i |i\rangle = \sum_i v_i \Omega|i\rangle = \sum_i v_i |i'\rangle$$

When we say $|i'\rangle$ is known, we mean that its components in the original basis

$$\langle j|i'\rangle = \langle j|\Omega|i\rangle \equiv \Omega_{ji} \tag{1.6.1}$$

are known. The n^2 numbers, Ω_{ij}, are the *matrix elements* of Ω in this basis. If

$$\Omega|V\rangle = |V'\rangle$$

then the components of the transformed ket $|V'\rangle$ are expressable in terms of the Ω_{ij} and the components of $|V'\rangle$:

$$v_i' = \langle i|V'\rangle = \langle i|\Omega|V\rangle = \langle i|\Omega\left(\sum_j v_j|j\rangle\right)$$

$$= \sum_j v_j \langle i|\Omega|j\rangle$$

$$= \sum_j \Omega_{ij} v_j \tag{1.6.2}$$

Equation (1.6.2) can be cast in matrix form:

$$\begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_n' \end{bmatrix} = \begin{bmatrix} \langle 1|\Omega|1\rangle & \langle 1|\Omega|2\rangle & \cdots & \langle 1|\Omega|n\rangle \\ \langle 2|\Omega|1\rangle & & & \\ \vdots & & & \vdots \\ \langle n|\Omega|1\rangle & \cdots & & \langle n|\Omega|n\rangle \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \tag{1.6.3}$$

A mnemonic: the elements of the first column are simply the components of the first transformed basis vector $|1'\rangle = \Omega|1\rangle$ in the given basis. *Likewise, the elements of the jth column represent the image of the jth basis vector after Ω acts on it.*

Convince yourself that the same matrix Ω_{ij} acting to the *left* on the row vector corresponding to any $\langle v'|$ gives the row vector corresponding to $\langle v''| = \langle v'|\Omega$.

Example 1.6.1. Combining our mnemonic with the fact that the operator $R(\frac{1}{2}\pi\mathbf{i})$ has the following effect on the basis vectors:

$$R(\tfrac{1}{2}\pi\mathbf{i})|1\rangle = |1\rangle$$

$$R(\tfrac{1}{2}\pi\mathbf{i})|2\rangle = |3\rangle$$

$$R(\tfrac{1}{2}\pi\mathbf{i})|3\rangle = -|2\rangle$$

we can write down the matrix that represents it in the $|1\rangle$, $|2\rangle$, $|3\rangle$ basis:

$$R(\tfrac{1}{2}\pi\mathbf{i}) \leftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{1.6.4}$$

For instance, the -1 in the third column tells us that R rotates $|3\rangle$ into $-|2\rangle$. One may also ignore the mnemonic altogether and simply use the definition $R_{ij} = \langle i|R|j\rangle$ to compute the matrix. $\qquad\square$

Exercise 1.6.1. An operator Ω is given by the matrix

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

What is its action?

Let us now consider certain specific operators and see how they appear in matrix form.

(1) The Identity Operator I.

$$I_{ij} = \langle i|I|j\rangle = \langle i|j\rangle = \delta_{ij} \tag{1.6.5}$$

Thus I is represented by a diagonal matrix with 1's along the diagonal. You should verify that our mnemonic gives the same result.

(2) The Projection Operators. Let us first get acquainted with *projection operators*. Consider the expansion of an arbitrary ket $|V\rangle$ in a basis:

$$|V\rangle = \sum_{i=1}^{n} |i\rangle\langle i|V\rangle$$

In terms of the objects $|i\rangle\langle i|$, which are linear operators, and which, by definition, act on $|V\rangle$ to give $|i\rangle\langle i|V\rangle$, we may write the above as

$$|V\rangle = \left(\sum_{i=1}^{n} |i\rangle\langle i| \right)|V\rangle \tag{1.6.6}$$

Since Eq. (1.6.6) is true for all $|V\rangle$, the object in the brackets must be identified with the identity (operator)

$$I = \sum_{i=1}^{n} |i\rangle\langle i| = \sum_{i=1}^{n} \mathbb{P}_i \tag{1.6.7}$$

The object $\mathbb{P}_i = |i\rangle\langle i|$ is called the *projection operator* for the ket $|i\rangle$. Equation (1.6.7), which is called the *completeness relation*, expresses the identity as a sum over projection operators and will be invaluable to us. (If you think that any time spent on the identity, which seems to do nothing, is a waste of time, just wait and see.)

Consider

$$\mathbb{P}_i|V\rangle = |i\rangle\langle i|V\rangle = |i\rangle v_i \tag{1.6.8}$$

Clearly $\mathbb{P}_i$ is linear. Notice that whatever $|V\rangle$ is, $\mathbb{P}_i|V\rangle$ is a multiple of $|i\rangle$ with a coefficient (v_i) which is the component of $|V\rangle$ along $|i\rangle$. Since $\mathbb{P}_i$ projects out the component of any ket $|V\rangle$ along the direction $|i\rangle$, it is called a *projection operator*. The completeness relation, Eq. (1.6.7), says that the sum of the projections of a vector along all the n directions equals the vector itself. Projection operators can also act on bras in the same way:

$$\langle V|\mathbb{P}_i = \langle V|i\rangle\langle i| = v_i^*\langle 1| \tag{1.6.9}$$

Pojection operators corresponding to the basis vectors obey

$$\mathbb{P}_i\mathbb{P}_j = |i\rangle\langle i|j\rangle\langle j| = \delta_{ij}\mathbb{P}_j \tag{1.6.10}$$

This equation tells us that (1) once $\mathbb{P}_i$ projects out the part of $|V\rangle$ along $|i\rangle$, further applications of $\mathbb{P}_i$ make no difference; and (2) the subsequent application of $\mathbb{P}_j(j \neq i)$ will result in zero, since a vector entirely along $|i\rangle$ cannot have a projection along a perpendicular direction $|j\rangle$.
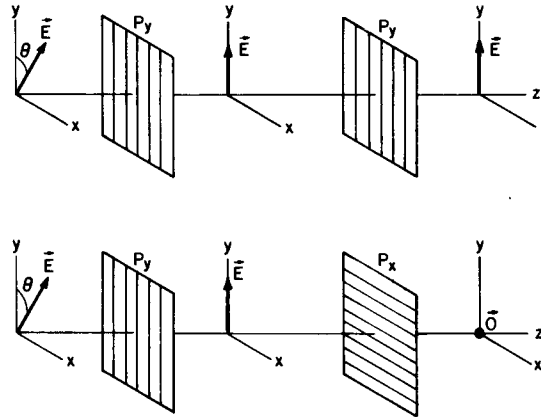
Figure 1.4. P_x and P_y are polarizers placed in the way of a beam traveling along the z axis. The action of the polarizers on the electric field **E** obeys the law of combination of projection operators: $P_i P_j = \delta_{ij} P_j$.

The following example from optics may throw some light on the discussion. Consider a beam of light traveling along the z axis and polarized in the $x - y$ plane at an angle θ with respect to the y axis (see Fig. 1.4). If a polarizer P_y, that only admits light polarized along the y axis, is placed in the way, the projection $E \cos \theta$ along the y axis is transmitted. An additional polarizer P_y placed in the way has no further effect on the beam. We may equate the action of the polarizer to that of a projection operator $\mathbb{P}_y$ that acts on the electric field vector **E**. If P_y is followed by a polarizer P_x the beam is completely blocked. Thus the polarizers obey the equation $P_i P_j = \delta_{ij} P_j$ expected of projection operators.

Let us next turn to the matrix elements of $\mathbb{P}_i$. There are two approaches. The first one, somewhat indirect, gives us a feeling for what kind of an object $|i\rangle\langle i|$ is. We know

$$
|i\rangle \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}
$$

and

$$
\langle i| \leftrightarrow (0, 0, \ldots, 1, 0, 0, \ldots, 0)
$$

so that

$$|i\rangle\langle i| \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (0, 0, \ldots, 1, 0, \ldots, 0) = \begin{bmatrix} 0 & & \cdots & & 0 \\ & \ddots & & & \\ & & 0 & & \\ \vdots & & & 1 & \\ & & & & 0 \\ & & & & & \ddots \\ 0 & & & & & 0 \end{bmatrix} \qquad (1.6.11)$$

by the rules of matrix multiplication. Whereas $\langle V|V'\rangle = (1 \times n$ matrix) $\times$ ($n \times 1$ matrix) $= (1 \times 1$ matrix) is a scalar, $|V\rangle\langle V'| = (n \times 1$ matrix) $\times (1 \times n$ matrix) $=$ ($n \times n$ matrix) is an operator. The inner product $\langle V|V'\rangle$ represents a bra and ket which have found each other, while $|V\rangle\langle V'|$, sometimes called the *outer product*, has the two factors looking the other way for a bra or a ket to dot with.

The more direct approach to the matrix elements gives

$$(\mathbb{P}_i)_{kl} = \langle k|i\rangle\langle i|l\rangle = \delta_{ki}\delta_{il} = \delta_{kl}\delta_{li} \qquad (1.6.12)$$

which is of course identical to Eq. (1.6.11). The same result also follows from mnemonic. Each projection operator has only one nonvanishing matrix element, a 1 at the ith element on the diagonal. The completeness relation, Eq. (1.6.7), says that when all the $\mathbb{P}_i$ are added, the diagonal fills out to give the identity. If we form the sum over just some of the projection operators, we get the operator which projects a given vector into the subspace spanned by just the corresponding basis vectors.

Matrices Corresponding to Products of Operators

Consider next the matrices representing a product of operators. These are related to the matrices representing the individual operators by the application of Eq. (1.6.7):

$$(\Omega\Lambda)_{ij} = \langle i|\Omega\Lambda|j\rangle = \langle i|\Omega I\Lambda|j\rangle$$
$$= \sum_k \langle i|\Omega|k\rangle\langle k|\Lambda|j\rangle = \sum_k \Omega_{ik}\Lambda_{kj} \qquad (1.6.13)$$

Thus the matrix representing the product of operators is the product of the matrices representing the factors.

The Adjoint of an Operator

Recall that given a ket $a|V\rangle \equiv |aV\rangle$ the corresponding bra is

$$\langle aV| = \langle V|a^* \quad \text{(and \textit{not} } \langle V|a)$$

In the same way, given a ket

$$\Omega|V\rangle = |\Omega V\rangle$$

the corresponding bra is

$$\langle \Omega V| = \langle V|\Omega^\dagger \qquad\qquad (1.6.14)$$

which *defines* the operator $\Omega^\dagger$. One may state this equation in words: if Ω turns a ket $|V\rangle$ to $|V'\rangle$, then $\Omega^\dagger$ turns the bra $\langle V|$ into $\langle V'|$. Just as α and α^*, $|V\rangle$ and $\langle V|$ are related but distinct objects, so are Ω and $\Omega^\dagger$. The relation between Ω, and $\Omega^\dagger$, called the *adjoint* of Ω or "omega dagger," is best seen in a basis:

$$(\Omega^\dagger)_{ij} = \langle i|\Omega^\dagger|j\rangle = \langle \Omega i|j\rangle$$
$$= \langle j|\Omega i\rangle^* = \langle j|\Omega|i\rangle^*$$

so

$$\Omega^\dagger_{ij} = \Omega^*_{ji} \qquad\qquad (1.6.15)$$

In other words, the matrix representing $\Omega^\dagger$ is the transpose conjugate of the matrix representing Ω. (Recall that the row vector representing $\langle V|$ is the transpose conjugate of the column vector representing $|V\rangle$. *In a given basis, the adjoint operation is the same as taking the transpose conjugate.*)

The adjoint of a product is the product of the adjoints in reverse:

$$(\Omega\Lambda)^\dagger = \Lambda^\dagger\Omega^\dagger \qquad\qquad (1.6.16)$$

To prove this we consider $\langle \Omega\Lambda V|$. First we treat $\Omega\Lambda$ as one operator and get

$$\langle \Omega\Lambda V| = \langle (\Omega\Lambda)V| = \langle V|(\Omega\Lambda)^\dagger$$

Next we treat (ΛV) as just another vector, and write

$$\langle \Omega\Lambda V| = \langle \Omega(\Lambda V)| = \langle \Lambda V|\Omega^\dagger$$

We next pull out Λ, pushing $\Omega^\dagger$ further out:

$$\langle \Lambda V|\Omega^\dagger = \langle V|\Lambda^\dagger\Omega^\dagger$$

Comparing this result with the one obtained a few lines above, we get the desired result.

Consider now an equation consisting of kets, scalars, and operators, such as

$$\alpha_1|V_1\rangle = \alpha_2|V_2\rangle + \alpha_3|V_3\rangle\langle V_4|V_5\rangle + \alpha_4\Omega\Lambda|V_6\rangle \qquad (1.6.17a)$$

What is its adjoint? Our old rule tells us that it is

$$\langle V_1 | \alpha_a^* = \langle V_2 | \alpha_2^* + \langle V_5 | V_4 \rangle \langle V_3 | \alpha_3^* + \langle \Omega \Lambda V_6 | \alpha_4^*$$

In the last term we can replace $\langle \Omega \Lambda V_6 |$ by

$$\langle V_6 | (\Omega \Lambda)^\dagger = \langle V_6 | \Lambda^\dagger \Omega^\dagger$$

so that finally we have the adjoint of Eq. (1.6.17a):

$$\langle V_1 | \alpha_1^* = \langle V_2 | \alpha_2^* + \langle V_5 | V_4 \rangle \langle V_3 | \alpha_3^* + \langle V_6 | \Lambda^\dagger \Omega^\dagger \alpha_4^* \qquad (1.6.17b)$$

The final rule for taking the adjoint of the most general equation we will ever encounter is this:

When a product of operators, bras, kets, and explicit numerical coefficients is encountered, reverse the order of all factors and make the substitutions $\Omega \leftrightarrow \Omega^\dagger$, $| \rangle \leftrightarrow \langle |$, $\alpha \leftrightarrow \alpha^*$.

(Of course, there is no real need to reverse the location of the scalars α except in the interest of uniformity.)

Hermitian, Anti-Hermitian, and Unitary Operators

We now turn our attention to certain special classes of operators that will play a major role in quantum mechanics.

Definition 13. An operator Ω is *Hermitian* if $\Omega^\dagger = \Omega$.

Definition 14. An operator Ω is *anti-Hermitian* if $\Omega^\dagger = -\Omega$.

The adjoint is to an operator what the complex conjugate is to numbers. Hermitian and anti-Hermitian operators are like pure real and pure imaginary numbers. Just as every number may be decomposed into a sum of pure real and pure imaginary parts,

$$\alpha = \frac{\alpha + \alpha^*}{2} + \frac{\alpha - \alpha^*}{2}$$

we can decompose every operator into its Hermitian and anti-Hermitian parts:

$$\Omega = \frac{\Omega + \Omega^\dagger}{2} + \frac{\Omega - \Omega^\dagger}{2} \qquad (1.6.18)$$

*Exercise 1.6.2.** Given Ω and Λ are Hermitian what can you say about (1) $\Omega \Lambda$; (2) $\Omega \Lambda + \Lambda \Omega$; (3) $[\Omega, \Lambda]$; and (4) $i[\Omega, \Lambda]$?

Definition 15. An operator U is *unitary* if

$$UU^\dagger = I \qquad (1.6.19)$$

This equation tells us that U and $U^\dagger$ are inverses of each other. Consequently, from Eq. (1.5.12),

$$U^\dagger U = I \qquad (1.6.20)$$

Following the analogy between operators and numbers, unitary operators are like complex numbers of unit modulus, $u = e^{i\theta}$. Just as $u^*u = 1$, so is $U^\dagger U = I$.

Exercise 1.6.3. * Show that a product of unitary operators is unitary.

Theorem 7. Unitary operators preserve the inner product between the vectors they act on.

Proof. Let

$$|V_1'\rangle = U|V_1\rangle$$

and

$$|V_2'\rangle = U|V_2\rangle$$

Then

$$\langle V_2'|V_1'\rangle = \langle UV_2|UV_1\rangle$$
$$= \langle V_2|U^\dagger U|V_1\rangle = \langle V_2|V_1\rangle \qquad (1.6.21)$$

$$(\text{Q.E.D.})$$

Unitary operators are the generalizations of rotation operators from $\mathbb{V}^3(R)$ to $\mathbb{V}^n(C)$, for just like rotation operators in three dimensions, they preserve the lengths of vectors and their dot products. In fact, on a real vector space, the unitarity condition becomes $U^{-1} = U^T$ (T means transpose), which defines an *orthogonal* or rotation matrix. $[R(\frac{1}{2}\pi\mathbf{i})$ is an example.]

Theorem 8. If one treats the columns of an $n \times n$ unitary matrix as components of n vectors, these vectors are orthonormal. In the same way, the rows may be interpreted as components of n orthonormal vectors.

Proof 1. According to our mnemonic, the jth column of the matrix representing U is the image of the jth basis vector after U acts on it. Since U preserves inner products, the rotated set of vectors is also orthonormal. Consider next the rows. We now use the fact that $U^\dagger$ is also a rotation. (How else can it neutralize U to give $U^\dagger U = I$?) Since the rows of U are the columns of $U^\dagger$ (but for an overall complex

conjugation which does not affect the question of orthonormality), the result we already have for the columns of a unitary matrix tells us the rows of U are orthonormal.

Proof 2. Since $U^\dagger U = I$,

$$\delta_{ij} = \langle i|I|j\rangle = \langle i|U^\dagger U|j\rangle$$

$$= \sum_k \langle i|U^\dagger|k\rangle\langle k|U|j\rangle$$

$$= \sum_k U_{ik}^\dagger U_{kj} = \sum_k U_{ki}^* U_{kj} \qquad (1.6.22)$$

which proves the theorem for the columns. A similar result for the rows follows if we start with the equation $UU^\dagger = I$.   Q.E.D.

Note that $U^\dagger U = I$ and $UU^\dagger = I$ are not independent conditions.

Exercise 1.6.4.* It is assumed that you know (1) what a *determinant* is, (2) that det $\Omega^T =$ det Ω (T denotes transpose), (3) that the determinant of a product of matrices is the product of the determinants. [If you do not, verify these properties for a two-dimensional case

$$\Omega = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

with det $\Omega = (\alpha\delta - \beta\gamma)$.] Prove that the determinant of a unitary matrix is a complex number of unit modulus.

Exercise 1.6.5.* Verify that $R(\tfrac{1}{2}\pi\mathbf{i})$ is unitary (orthogonal) by examining its matrix.

Exercise 1.6.6. Verify that the following matrices are unitary:

$$\frac{1}{2^{1/2}}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}, \qquad \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}$$

Verify that the determinant is of the form $e^{i\theta}$ in each case. Are any of the above matrices Hermitian?

1.7. Active and Passive Transformations

Suppose we subject all the vectors $|V\rangle$ in a space to a unitary transformation

$$|V\rangle \to U|V\rangle \qquad (1.7.1)$$

Under this transformation, the matrix elements of any operator Ω are modified as follows:

$$\langle V'|\Omega|V\rangle \to \langle UV'|\Omega|UV\rangle = \langle V'|U^\dagger\Omega U|V\rangle \qquad (1.7.2)$$

It is clear that the same change would be effected if we left the vectors alone and subjected all operators to the change

$$\Omega \to U^{\dagger}\Omega U \tag{1.7.3}$$

The first case is called an *active transformation* and the second a *passive transformation*. The present nomenclature is in reference to the vectors: they are affected in an active transformation and left alone in the passive case. The situation is exactly the opposite from the point of view of the operators.

Later we will see that the physics in quantum theory lies in the matrix elements of operators, and that active and passive transformations provide us with two equivalent ways of describing the same physical transformation.

Exercise 1.7.1. * The *trace* of a matrix is defined to be the sum of its diagonal matrix elements

$$\text{Tr}\,\Omega = \sum_{i}\Omega_{ii}$$

Show that

(1) $\text{Tr}(\Omega\Lambda) = \text{Tr}(\Lambda\Omega)$
(2) $\text{Tr}(\Omega\Lambda\theta) = \text{Tr}(\Lambda\theta\Omega) = \text{TR}(\theta\Omega\Lambda)$ (The permutations are *cyclic*).
(3) The trace of an operator is unaffected by a unitary change of basis $|i\rangle \to U|i\rangle$. [Equivalently, show $\text{Tr}\,\Omega = \text{Tr}(U^{\dagger}\Omega U)$.]

Exercise 1.7.2. Show that the determinant of a matrix is unaffected by a unitary change of basis. [Equivalently show $\det\Omega = \det(U^{\dagger}\Omega U)$.]

1.8. The Eigenvalue Problem

Consider some linear operator Ω acting on an arbitrary *nonzero* ket $|V\rangle$:

$$\Omega|V\rangle = |V'\rangle \tag{1.8.1}$$

Unless the operator happens to be a trivial one, such as the identity or its multiple, the ket will suffer a nontrivial change, i.e., $|V'\rangle$ will not be simply related to $|V\rangle$. So much for an arbitrary ket. Each operator, however, has certain kets of its own, called its *eigenkets*, on which its action is simply that of rescaling:

$$\Omega|V\rangle = \omega|V\rangle \tag{1.8.2}$$

Equation (1.8.2) is an eigenvalue equation: $|V\rangle$ is an *eigenket* of Ω with *eigenvalue* ω. In this chapter we will see how, given an operator Ω, one can systematically determine all its eigenvalues and eigenvectors. How such an equation enters physics will be illustrated by a few examples from mechanics at the end of this section, and once we get to quantum mechanics proper, it will be eigen, eigen, eigen all the way.

Example 1.8.1. To illustrate how easy the eigenvalue problem really is, we will begin with a case that will be completely solved: the case $\Omega = I$. Since

$$I|V\rangle = |V\rangle$$

for all $|V\rangle$, we conclude that

(1) the only eigenvalue of I is 1;
(2) all vectors are its eigenvectors with this eigenvalue. $\square$

Example 1.8.2. After this unqualified success, we are encouraged to take on a slightly more difficult case: $\Omega = \mathbb{P}_V$, the projection operator associated with a *normalized* ket $|V\rangle$. Clearly

(1) any ket $\alpha|V\rangle$, parallel to $|V\rangle$ is an eigenket with eigenvalue 1:

$$\mathbb{P}_V|\alpha V\rangle = |V\rangle\langle V|\alpha V\rangle = \alpha|V\rangle|V|^2 = 1 \cdot |\alpha V\rangle$$

(2) any ket $|V_\perp\rangle$, perpendicular to $|V\rangle$, is an eigenket with eigenvalue 0:

$$\mathbb{P}_V|V_\perp\rangle = |V\rangle\langle V|V_\perp\rangle = 0 = 0|V_\perp\rangle$$

(3) kets that are neither, i.e., kets of the form $\alpha|V\rangle + \beta|V_\perp\rangle$, are simply not eigenkets:

$$\mathbb{P}_V(\alpha|V\rangle + \beta|V_\perp\rangle) = |\alpha V\rangle \neq \gamma(\alpha|V\rangle + \beta|V_\perp\rangle)$$

Since every ket in the space falls into one of the above classes, we have found all the eigenvalues and eigenvectors. $\square$

Example 1.8.3. Consider now the operator $R(\frac{1}{2}\pi\mathbf{i})$. We already know that it has one eigenket, the basis vector $|1\rangle$ along the x axis:

$$R(\tfrac{1}{2}\pi\mathbf{i})|1\rangle = |1\rangle$$

Are there others? Of course, any vector $\alpha|1\rangle$ along the x axis is also unaffected by the x rotation. This is a general feature of the eigenvalue equation and reflects the linearity of the operator:

if

$$\Omega|V\rangle = \omega|V\rangle$$

then

$$\Omega\alpha|V\rangle = \alpha\Omega|V\rangle = \alpha\omega|V\rangle = \omega\alpha|V\rangle$$

for any multiple α. Since the eigenvalue equation fixes the eigenvector only up to an overall scale factor, we will not treat the multiples of an eigenvector as distinct eigenvectors. With this understanding in mind, let us ask if $R(\frac{1}{2}\pi\mathbf{i})$ has any eigenvectors besides $|1\rangle$. Our intuition says no, for any vector not along the x axis necessarily gets rotated by $R(\frac{1}{2}\pi\mathbf{i})$ and cannot possibly transform into a multiple of itself. Since every vector is either parallel to $|1\rangle$ or isn't, we have fully solved the eigenvalue problem.

The trouble with this conclusion is that it is wrong! $R(\frac{1}{2}\pi\mathbf{i})$ has two other eigenvectors besides $|1\rangle$. But our intuition is not to be blamed, for these vectors are in $\mathbb{V}^3(C)$ and not $\mathbb{V}^3(R)$. It is clear from this example that we need a reliable and systematic method for solving the eigenvalue problem in $\mathbb{V}^n(C)$. We now turn our attention to this very question. $\qquad\qquad\square$

The Characteristic Equation and the Solution to the Eigenvalue Problem

We begin by rewriting Eq. (1.8.2) as

$$(\Omega - \omega I)|V\rangle = |0\rangle \qquad\qquad (1.8.3)$$

Operating both sides with $(\Omega - \omega I)^{-1}$, assuming it exists, we get

$$|V\rangle = (\Omega - \omega I)^{-1}|0\rangle \qquad\qquad (1.8.4)$$

Now, any finite operator (an operator with finite matrix elements) acting on the null vector can only give us a null vector. It therefore seems that in asking for a nonzero eigenvector $|V\rangle$, we are trying to get something for nothing out of Eq. (1.8.4). This is impossible. It follows that our assumption that the operator $(\Omega - \omega I)^{-1}$ exists (as a finite operator) is false. So we ask when this situation will obtain. Basic matrix theory tells us (see Appendix A.1) that the inverse of any matrix M is given by

$$M^{-1} = \frac{\text{cofactor } M^T}{\det M} \qquad . \qquad\qquad (1.8.5)$$

Now the cofactor of M is finite if M is. Thus what we need is the vanishing of the determinant. The condition for nonzero eigenvectors is therefore

$$\det(\Omega - \omega I) = 0 \qquad\qquad (1.8.6)$$

This equation will determine the eigenvalues ω. To find them, we project Eq. (1.8.3) onto a basis. Dotting both sides with a basis bra $\langle i|$, we get

$$\langle i|\Omega - \omega I|V\rangle = 0$$

and upon introducing the representation of the identity [Eq. (1.6.7)], to the left of $|V\rangle$, we get the following image of Eq. (1.8.3):

$$\sum_j (\Omega_{ij} - \omega \delta_{ij}) v_j = 0 \tag{1.8.7}$$

Setting the determinant to zero will give us an expression of the form

$$\sum_{m=0}^n c_m \omega^m = 0 \tag{1.8.8}$$

Equation (1.8.8) is called the *characteristic equation* and

$$P^n(\omega) = \sum_{m=0}^n c_m \omega^m \tag{1.8.9}$$

is called the *characteristic polynomial*. Although the polynomial is being determined in a particular basis, the eigenvalues, which are its roots, are basis independent, for they are defined by the abstract Eq. (1.8.3), which makes no reference to any basis.

Now, a fundamental result in analysis is that every nth-order polynomial has n roots, not necessarily distinct and not necessarily real. Thus every operator in $\mathbb{V}^n(\mathbb{C})$ has n eigenvalues. Once the eigenvalues are known, the eigenvectors may be found, at least for Hermitian and unitary operators, using a procedure illustrated by the following example. [Operators on $\mathbb{V}^n(C)$ that are not of the above variety may not have n eigenvectors—see Exercise 1.8.4. Theorems 10 and 12 establish that Hermitian and unitary operators on $\mathbb{V}^n(C)$ will have n eigenvectors.]

Example 1.8.4. Let us use the general techniques developed above to find all the eigenvectors and eigenvalues of $R(\frac{1}{2}\pi\mathbf{i})$. Recall that the matrix representing it is

$$R(\tfrac{1}{2}\pi\mathbf{i}) \leftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Therefore the characteristic equation is

$$\det(R - \omega I) = \begin{vmatrix} 1 - \omega & 0 & 0 \\ 0 & -\omega & -1 \\ 0 & 1 & -\omega \end{vmatrix} = 0$$

i.e.,

$$(1 - \omega)(\omega^2 + 1) = 0 \tag{1.8.10}$$

with roots $\omega = 1$, $\pm i$. We know that $\omega = 1$ corresponds to $|1\rangle$. Let us see this come out of the formalism. Feeding $\omega = 1$ into Eq. (1.8.7) we find that the components x_1, x_2, and x_3 of the corresponding eigenvector must obey the equations

$$\begin{bmatrix} 1-1 & 0 & 0 \\ 0 & 0-1 & -1 \\ 0 & 1 & 0-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \left. \begin{array}{c} 0 = 0 \\ -x_2 - x_3 = 0 \\ x_2 - x_3 = 0 \end{array} \right\} \rightarrow x_2 = x_3 = 0$$

Thus any vector of the form

$$x_1 |1\rangle \leftrightarrow \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}$$

is acceptable, as expected. It is conventional to use the freedom in scale to normalize the eigenvectors. Thus in this case a choice is

$$|\omega = 1\rangle = |1\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

I say *a* choice, and not *the* choice, since the vector may be multiplied by a number of modulus unity without changing the norm. There is no universally accepted convention for eliminating this freedom, except perhaps to choose the vector with real components when possible.

Note that of the three simultaneous equations above, the first is not a real equation. In general, there will be only $(n-1)$ LI equations. This is the reason the norm of the vector is not fixed and, as shown in Appendix A.1, the reason the determinant vanishes.

Consider next the equations corresponding to $\omega = i$. The components of the eigenvector obey the equations

$$(1 - i)x_1 = 0 \qquad (\text{i.e., } x_1 = 0)$$

$$-ix_2 - x_3 = 0 \qquad (\text{i.e., } x_2 = ix_3)$$

$$x_2 - ix_3 = 0 \qquad (\text{i.e., } x_2 = ix_3)$$

Notice once again that we have only $n-1$ useful equations. A properly normalized solution to the above is

$$|\omega = i\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 0 \\ i \\ 1 \end{bmatrix}$$

$$|\omega = -i\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 0 \\ -i \\ 1 \end{bmatrix} \qquad \square$$

In the above example we have introduced a popular convention: labeling the eigenvectors by the eigenvalue. For instance, the ket corresponding to $\omega = \omega_i$ is labeled $|\omega = \omega_i\rangle$ or simply $|\omega_i\rangle$. This notation presumes that to each ω_i there is just one vector labeled by it. Though this is not always the case, only a slight change in this notation will be needed to cover the general case.

The phenomenon of a single eigenvalue representing more than one eigenvector is called *degeneracy* and corresponds to repeated roots for the characteristic polynomial. In the face of degeneracy, we need to modify not just the labeling, but also the procedure used in the example above for finding the eigenvectors. Imagine that instead of $R(\frac{1}{2}\pi i)$ we were dealing with another operator Ω on $\mathbb{V}^3(R)$ with roots ω_1 and $\omega_2 = \omega_3$. It appears as if we can get two eigenvectors, by the method described above, one for each distinct ω. How do we get a third? Or is there no third? These equations will be answered in all generality shortly when we examine the question of degeneracy in detail. We now turn our attention to two central theorems on Hermitian operators. These play a vital role in quantum mechanics.

Theorem 9. The eigenvalues of a Hermitian operator are real.

Proof. Let

$$\Omega|\omega\rangle = \omega|\omega\rangle$$

Dot both sides with $\langle\omega|$:

$$\langle\omega|\Omega|\omega\rangle = \omega\langle\omega|\omega\rangle \qquad (1.8.11)$$

Take the adjoint to get

$$\langle\omega|\Omega^\dagger|\omega\rangle = \omega^*\langle\omega|\omega\rangle$$

Since $\Omega = \Omega^\dagger$, this becomes

$$\langle\omega|\Omega|\omega\rangle = \omega^*\langle\omega|\omega\rangle$$

Subtracting from Eq. (1.8.11)

$$0 = (\omega - \omega^*)\langle\omega|v\rangle$$

$$\omega = \omega^* \quad \text{Q.E.D.}$$

Theorem 10. To every Hermitian operator Ω, there exists (at least) a basis consisting of its orthonormal eigenvectors. It is diagonal in this eigenbasis and has its eigenvalues as its diagonal entries.

Proof. Let us start with the characteristic equation. It must have at least one root, call it ω_1. Corresponding to ω_1 there must exist at least one nonzero eigenvector $|\omega_1\rangle$. [If not, Theorem (A.1.1) would imply that $(\Omega - \omega_1 I)$ is invertible.] Consider the subspace $\mathbb{V}_{\perp 1}^{n-1}$ of all vectors orthogonal to $|\omega_1\rangle$. Let us choose as our basis the vector $|\omega_1\rangle$ (normalized to unity) and any $n-1$ orthonormal vectors $\{V_{\perp 1}^1, V_{\perp 1}^2, \ldots, V_{\perp 1}^{n-1}\}$ in $\mathbb{V}_{\perp i}^{n-1}$. In this basis Ω has the following form:

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & & & & & & \\ 0 & & & \boxed{} & & & \\ \vdots & & & & & & \\ 0 & & & & & & \end{bmatrix} \tag{1.8.12}$$

The first column is just the image of $|\omega_1\rangle$ after Ω has acted on it. Given the first column, the first row follows from the Hermiticity of Ω.

The characteristic equation now takes the form

$$(\omega_1 - \omega) \cdot (\text{determinant of boxed submatrix}) = 0$$

$$(\omega_1 - \omega) \sum_0^{n-1} c_m \omega^m = (\omega_1 - \omega) P^{n-1}(\omega) = 0$$

Now the polynomial P^{n-1} must also generate one root, ω_2, and a normalized eigenvector $|\omega_2\rangle$. Define the subspace $\mathbb{V}_{\perp 1,2}^{n-2}$ of vectors in $\mathbb{V}_{\perp 1}^{n-1}$ orthogonal to $|\omega_2\rangle$ (and automatically to $|\omega_2\rangle$) and repeat the same procedure as before. Finally, the matrix Ω becomes, in the basis $|\omega_1\rangle, |\omega_2\rangle, \ldots, |\omega_n\rangle$,

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2 & 0 & & 0 \\ 0 & 0 & \omega_3 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & & \omega_n \end{bmatrix}$$

Since every $|\omega_i\rangle$ was chosen from a space that was orthogonal to the previous ones, $|\omega_1\rangle, |\omega_2\rangle, \ldots, |\omega_{i-1}\rangle$; the basis of eigenvectors is orthonormal. (Notice that nowhere did we have to assume that the eigenvalues were all distinct.)   Q.E.D.

[The analogy between real numbers and Hermitian operators is further strengthened by the fact that in a certain basis (of eigenvectors) the Hermitian operator can be represented by a matrix with all real elements.]

In stating Theorem 10, it was indicated that there might exist more than one basis of eigenvectors that diagonalized Ω. This happens if there is any degeneracy. Suppose $\omega_1 = \omega_2 = \omega$. Then we have two orthonormal vectors obeying

$$\Omega|\omega_1\rangle = \omega|\omega_1\rangle$$

$$\Omega|\omega_2\rangle = \omega|\omega_2\rangle$$

It follows that

$$\Omega[\alpha|\omega_1\rangle + \beta|\omega_2\rangle] = \alpha\omega|\omega_1\rangle + \beta\omega|\omega_2\rangle = \omega[\alpha|\omega_1\rangle + \beta|\omega_2\rangle]$$

for any α and β. Since the vectors $|\omega_1\rangle$ and $|\omega_2\rangle$ are orthogonal (and hence LI), we find that there is a whole two-dimensional subspace spanned by $|\omega_1\rangle$ and $|\omega_2\rangle$, the elements of which are eigenvectors of Ω with eigenvalue ω. One refers to this space as an *eigenspace* of Ω with eigenvalue ω. Besides the vectors $|\omega_1\rangle$ and $|\omega_2\rangle$, there exists an infinity of orthonormal pairs $|\omega_1'\rangle$, $|\omega_2'\rangle$, obtained by a rigid rotation of $|\omega_1\rangle$, $|\omega_2\rangle$, from which we may select any pair in forming the eigenbasis of Ω. In general, if an eigenvalue occurs m_i times, that is, if the characteristic equation has m_i of its roots equal to some ω_i, there will be an eigenspace $\mathbb{V}_{\omega_i}^{m_i}$ from which we may choose any m_i orthonormal vectors to form the basis referred to in Theorem 10.

In the absence of degeneracy, we can prove Theorem 9 and 10 very easily. Let us begin with two eigenvectors:

$$\Omega|\omega_i\rangle = \omega_i|\omega_i\rangle \tag{1.8.13a}$$

$$\Omega|\omega_j\rangle = \omega_j|\omega_j\rangle \tag{1.8.13b}$$

Dotting the first with $\langle\omega_j|$ and the second with $\langle\omega_i|$, we get

$$\langle\omega_j|\Omega|\omega_i\rangle = \omega_i\langle\omega_j|\omega_i\rangle \tag{1.8.14a}$$

$$\langle\omega_i|\Omega|\omega_j\rangle = \omega_j\langle\omega_i|\omega_j\rangle \tag{1.8.14b}$$

Taking the adjoint of the last equation and using the Hermitian nature of Ω, we get

$$\langle\omega_j|\Omega|\omega_i\rangle = \omega_j^*\langle\omega_j|\omega_i\rangle$$

Subtracting this equation from Eq. (1.8.14a), we get

$$0 = (\omega_i - \omega_j^*)\langle\omega_j|\omega_i\rangle \tag{1.8.15}$$

If $i=j$, we get, since $\langle\omega_i|\omega_i\rangle \neq 0$,

$$\omega_i = \omega_i^* \tag{1.8.16}$$

If $i \neq j$, we get

$$\langle \omega_i | \omega_j \rangle = 0 \qquad (1.8.17)$$

since $\omega_i - \omega_j^* = \omega_i - \omega_j \neq 0$ by assumption. That the proof of orthogonality breaks down for $\omega_i = \omega_j$ is not surprising, for two vectors labeled by a degenerated eigenvalue could be any two members of the degenerate space which need not necessarily be orthogonal. The modification of this proof in this case of degeneracy calls for arguments that are essentially the ones used in proving Theorem 10. The advantage in the way Theorem 10 was proved first is that it suffers no modification in the degenerate case.

Degeneracy

We now address the question of degeneracy as promised earlier. Now, our general analysis of Theorem 10 showed us that in the face of degeneracy, we have not one, but an infinity of orthonormal eigenbases. Let us see through an example how this variety manifests itself when we look for eigenvectors and how it is to be handled.

Example 1.8.5. Consider an operator Ω with matrix elements

$$\Omega \leftrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

in some basis. The characteristic equation is

$$(\omega - 2)^2 \omega = 0$$

i.e.,

$$\omega = 0, 2, 2$$

The vector corresponding to $\omega = 0$ is found by the usual means to be

$$|\omega = 0\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

The case $\omega = 2$ leads to the following equations for the components of the eigenvector:

$$-x_1 + x_3 = 0$$

$$0 = 0$$

$$x_1 - x_2 = 0$$

Now we have just one equation, instead of the two $(n-1)$ we have grown accustomed to! This is a reflection of the degeneracy. For every extra appearance (besides the first) a root makes, it takes away one equation. Thus degeneracy permits us extra degrees of freedom besides the usual one (of normalization). The conditions

$$x_1 = x_3$$

$$x_2 \text{ arbitrary}$$

define an ensemble of vectors that are perpendicular to the first, $|\omega=0\rangle$, i.e., lie in a plane perpendicular to $|\omega=0\rangle$. This is in agreement with our expectation that a twofold degeneracy should lead to a two-dimensional eigenspace. The freedom in x_2 (or more precisely, the ratio x_2/x_3) corresponds to the freedom of orientation in this plane. Let us arbitrarily choose $x_2=1$, to get a normalized eigenvector corresponding to $\omega=2$:

$$|\omega=2\rangle \leftrightarrow \frac{1}{3^{1/2}}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

The third vector is now chosen to lie in this plane and to be orthogonal to the second (being in this plane automatically makes it perpendicular to the first $|\omega=0\rangle$):

$$|\omega=2, \text{second one}\rangle \leftrightarrow \frac{1}{6^{1/2}}\begin{bmatrix}1\\-2\\1\end{bmatrix}$$

Clearly each distinct choice of the ratio, x_2/x_3, gives us a distinct doublet of orthonormal eigenvectors with eigenvalue 2. $\qquad\square$

Notice that in the face of degeneracy, $|\omega_i\rangle$ no longer refers to a single ket but to a generic element of the eigenspace $\mathbb{V}_{\omega_i}^{m_i}$. To refer to a particular element, we must use the symbol $|\omega_i, \alpha\rangle$, where α labels the ket within the eigenspace. A natural choice of the label α will be discussed shortly.

We now consider the analogs of Theorems 9 and 10 for unitary operators.

Theorem 11. The eigenvalues of a unitary operator are complex numbers of unit modulus.

Theorem 12. The eigenvectors of a unitary operator are mutually orthogonal. (We assume there is no degeneracy.)

Proof of Both Theorems (assuming no degeneracy).   Let

$$U|u_i\rangle = u_i|u_i\rangle \qquad (1.8.18a)$$

and

$$U|u_j\rangle = u_j|u_j\rangle \qquad (1.8.18b)$$

If we take the adjoint of the second equation and dot each side with the corresponding side of the first equation, we get

$$\langle u_j| U^\dagger U|u_i\rangle = u_i u_j^* \langle u_j|u_i\rangle$$

so that

$$(1 - u_i u_j^*)\langle u_j|u_i\rangle = 0 \qquad (1.8.19)$$

If $i=j$, we get, since $\langle u_i|u_i\rangle \neq 0$,

$$u_i u_i^* = 1 \qquad (1.8.20a)$$

while if $i \neq j$,

$$\langle u_i|u_j\rangle = 0 \qquad (1.8.20b)$$

since $|u_i\rangle \neq |u_j\rangle \Rightarrow u_i \neq u_j \Rightarrow u_i u_j^* \neq u_i u_i^* \Rightarrow u_i u_j^* \neq 1$.   (Q.E.D.)

If U is degenerate, we can carry out an analysis parallel to that for the Hermitian operator Ω, with just one difference. Whereas in Eq. (1.8.12), the zeros of the first row followed from the zeros of the first column and $\Omega^\dagger = \Omega$, here they follow from the requirement that the sum of the modulus squared of the elements in each row adds up to 1. Since $|u_1| = 1$, all the other elements in the first row must vanish.

Diagonalization of Hermitian Matrices

Consider a Hermitian operator Ω on $\mathbb{V}^n(C)$ represented as a matrix in some orthonormal basis $|1\rangle, \ldots, |i\rangle, \ldots, |n\rangle$. If we trade this basis for the eigenbasis $|\omega_1\rangle, \ldots, |\omega_i\rangle, \ldots, |\omega_n\rangle$, the matrix representing Ω will become diagonal. Now the operator U inducing the change of basis

$$|\omega_i\rangle = U|i\rangle \qquad (1.8.21)$$

is clearly unitary, for it "rotates" one orthonormal basis into another. (If you wish you may apply our mnemonic to U and verify its unitary nature: its columns contain the components of the eigenvectors $|\omega_i\rangle$ that are orthonormal.) This result is often summarized by the statement:

Every Hermitian matrix on $\mathbb{V}^n(C)$ may be diagonalized by a unitary change of basis.

We may restate this result in terms of passive transformations as follows:

If Ω is a Hermitian matrix, there exists a unitary matrix U (built out of the eigenvectors of Ω) such that $U^\dagger \Omega U$ is diagonal.

Thus the problem of finding a basis that diagonalizes Ω is equivalent to solving its eigenvalue problem.

Exercise 1.8.1. (1) Find the eigenvalues and normalized eigenvectors of the matrix

$$\Omega = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 4 \end{bmatrix}$$

(2) Is the matrix Hermitian? Are the eigenvectors orthogonal?

*Exercise 1.8.2.** Consider the matrix

$$\Omega = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

(1) Is it Hermitian?
(2) Find its eigenvalues and eigenvectors.
(3) Verify that $U^\dagger \Omega U$ is diagonal, U being the matrix of eigenvectors of Ω.

*Exercise 1.8.3.** Consider the Hermitian matrix

$$\Omega = \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

(1) Show that $\omega_1 = \omega_2 = 1$; $\omega_3 = 2$.
(2) Show that $|\omega = 2\rangle$ is any vector of the form

$$\frac{1}{(2a^2)^{1/2}} \begin{bmatrix} 0 \\ a \\ -a \end{bmatrix}$$

(3) Show that the $\omega = 1$ eigenspace contains all vectors of the form

$$\frac{1}{(b^2 + 2c^2)^{1/2}} \begin{bmatrix} b \\ c \\ c \end{bmatrix}$$

either by feeding $\omega = 1$ into the equations or by requiring that the $\omega = 1$ eigenspace be orthogonal to $|\omega = 2\rangle$.

Exercise 1.8.4. An arbitrary $n \times n$ matrix need not have n eigenvectors. Consider as an example

$$\Omega = \begin{bmatrix} 4 & 1 \\ -1 & 2 \end{bmatrix}$$

(1) Show that $\omega_1 = \omega_2 = 3$.
(2) By feeding in this value show we get only one eigenvector of the form

$$\frac{1}{(2a^2)^{1/2}} \begin{bmatrix} +a \\ -a \end{bmatrix}$$

We cannot find another one that is LI.

*Exercise 1.8.5.** Consider the matrix

$$\Omega = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

(1) Show that it is unitary.
(2) Show that its eigenvalues are $e^{i\theta}$ and $e^{-i\theta}$.
(3) Find the corresponding eigenvectors; show that they are orthogonal.
(4) Verify that $U^\dagger \Omega U = $ (diagonal matrix), where U is the matrix of eigenvectors of Ω.

*Exercise 1.8.6.** (1) We have seen that the determinant of a matrix is unchanged under a unitary change of basis. Argue now that

$$\det \Omega = \text{product of eigenvalues of } \Omega = \prod_{i=1}^{n} \omega_i$$

for a Hermitian or unitary Ω.
(2) Using the invariance of the trace under the same transformation, show that

$$\text{Tr } \Omega = \sum_{i=1}^{n} \omega_i$$

Exercise 1.8.7. By using the results on the trace and determinant from the last problem, show that the eigenvalues of the matrix

$$\Omega = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

are 3 and -1. Verify this by explicit computation. Note that the Hermitian nature of the matrix is an essential ingredient.

$$M^i M^j + M^j M^i = 2\delta^{ij} I, \qquad i, j = 1, \ldots, 4$$

(1) Show that the eigenvalues of M^i are ± 1. (Hint: go to the eigenbasis of M^i, and use the equation for $i = j$.)

(2) By considering the relation

$$M^i M^j = -M^j M^i \quad \text{for } i \neq j$$

show that M^i are traceless. [Hint: $\mathrm{Tr}(ACB) = \mathrm{Tr}(CBA)$.]

(3) Show that they cannot be odd-dimensional matrices.

Exercise 1.8.9. A collection of masses m_α, located at $\mathbf{r}_\alpha$ and rotating with angular velocity ω around a common axis has an angular momentum

$$\mathbf{l} = \sum_\alpha m_\alpha (\mathbf{r}_\alpha \times \mathbf{v}_\alpha)$$

where $\mathbf{v}_\alpha = \omega \times \mathbf{r}_\alpha$ is the velocity of m_α. By using the identity

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

show that each Cartesian component l_i of $\mathbf{l}$ is given by

$$l_i = \sum_j M_{ij} \omega_j$$

where

$$M_{ij} = \sum_\alpha m_\alpha [r_\alpha^2 \delta_{ij} - (\mathbf{r}_\alpha)_i (\mathbf{r}_\alpha)_j]$$

or in Dirac notation

$$|l\rangle = M|\omega\rangle$$

(1) Will the angular momentum and angular velocity always be parallel?

(2) Show that the moment of inertia matrix M_{ij} is Hermitian.

(3) Argue now that there exist three directions for ω such that $\mathbf{l}$ and ω will be parallel. How are these directions to be found?

(4) Consider the moment of inertia matrix of a sphere. Due to the complete symmetry of the sphere, it is clear that every direction is its eigendirection for rotation. What does this say about the three eigenvalues of the matrix M?

Simultaneous Diagonalization of Two Hermitian Operators

Let us consider next the question of simultaneously diagonalizing two Hermitian operators.

Theorem 13. If Ω and Λ are two commuting Hermitian operators, there exists (at least) a basis of common eigenvectors that diagonalizes them both.

Proof. Consider first the case where at least one of the operators is nondegenerate, i.e., to a given eigenvalue, there is just one eigenvector, up to a scale. Let us assume Ω is nondegenerate. Consider any one of its eigenvectors:

$$\Omega|\omega_i\rangle = \omega_i|\omega_i\rangle$$

$$\Lambda\Omega|\omega_i\rangle = \omega_i\Lambda|\omega_i\rangle$$

Since $[\Lambda, \Omega] = 0$,

$$\Omega\Lambda|\omega_i\rangle = \omega_i\Lambda|\omega_i\rangle \tag{1.8.22}$$

i.e., $\Lambda/\omega_i\rangle$ is an eigenvector of Ω with eigenvalue ω_i. Since this vector is unique up to a scale,

$$\Lambda|\omega_i\rangle = \lambda_i|\omega_i\rangle \tag{1.8.23}$$

Thus $|\omega_i\rangle$ is also an eigenvector of Λ with eigenvalue λ_i. Since every eigenvector of Ω is an eigenvector of Λ, it is evident that the basis $|\omega_i\rangle$ will diagonalize both operators. Since Ω is nondegenerate, there is only one basis with this property.

What if both operators are degenerate? By ordering the basis vectors such that the elements of each eigenspace are adjacent, we can get one of them, say Ω, into the form (Theorem 10)

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & & & & & & & \\ & \omega_1 & & & & & & \\ & & \ddots & & & & & \\ & & & \omega_1 & & & & \\ & & & & \omega_2 & & & \\ & & & & & \ddots & & \\ & & & & & & \omega_m & \\ & & & & & & & \omega_m \end{bmatrix}$$

Now this basis is not unique: in every eigenspace $\mathbb{V}_{\omega_i}^{m_i} \equiv \mathbb{V}_i^{m_i}$ corresponding to the eigenvalue ω_i, there exists an infinity of bases. Let us arbitrarily pick in $\mathbb{V}_{\omega_i}^{m_i}$ a set $|\omega_i, \alpha\rangle$ where the additional label α runs from 1 to m_i.

How does Λ appear in the basis? Although we made no special efforts to get Λ into a simple form, it already has a simple form by virtue of the fact that it commutes with Ω. Let us start by mimicking the proof in the nondegenerate case:

$$\Omega\Lambda|\omega_i, \alpha\rangle = \Lambda\Omega|\omega_i, \alpha\rangle = \omega_i\Lambda|\omega_i, \alpha\rangle$$

However, due to the degeneracy of Ω, we can only conclude that

$$\Lambda|\omega_i, \alpha\rangle \text{ lies in } \mathbb{V}_i^{m_i}$$

Now, since vectors from different eigenspaces are orthogonal [Eq. (1.8.15)],

$$\langle \omega_j, \beta|\Lambda|\omega_i, \alpha\rangle = 0$$

if $|\omega_i, \alpha\rangle$ and $|\omega_j, \beta\rangle$ are basis vectors such that $\omega_i \neq \omega_j$. Consequently, in this basis,

$$\Lambda \leftrightarrow \begin{bmatrix} \boxed{\Lambda_1} & & & 0 \\ & \boxed{\Lambda_2} & & \\ & & \ddots & \\ 0 & & & \boxed{\Lambda_k} \end{bmatrix}$$

which is called a *block diagonal matrix* for obvious reasons. The block diagonal form of Λ reflects the fact that when Λ acts on some element $|\omega_i, \alpha\rangle$ of the eigenspace $\mathbb{V}_i^{m_i}$, it turns it into another element of $\mathbb{V}_i^{m_i}$. Within each subspace i, Λ is given by a matrix Λ_i, which appears as a block in the equation above. Consider a matrix Λ_i in $\mathbb{V}_i^{m_i}$. It is Hermitian since Λ is. It can obviously be diagonalized by trading the basis $|\omega_i, 1\rangle, |\omega_i, 2\rangle, \ldots, |\omega_i, m_i\rangle$ in $\mathbb{V}_i^{m_i}$ that we started with, for the eigenbasis of Λ_i. Let us make such a change of basis in each eigenspace, thereby rendering Λ diagonal. Meanwhile what of Ω? It remains diagonal of course, since it is indifferent to the choice of orthonormal basis in each degenerate eigenspace. If the eigenvalues of Λ_i are $\lambda_i^{(1)} \lambda_i^{(2)}, \ldots, \lambda_i^{(m_i)}$ then we end up with

$$\Lambda \leftrightarrow \begin{bmatrix} \lambda_1^{(1)} & & & & & & \\ & \lambda_1^{(2)} & & & & & \\ & & \ddots & & & & \\ & & & \lambda_1^{(m_1)} & & & \\ & & & & \lambda_2^{(1)} & & \\ & & & & & \ddots & \\ & & & & & & \lambda_k^{(m_k)} \end{bmatrix},$$

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & & & & & & \\ & \omega_1 & & & & & \\ & & \ddots & & & & \\ & & & \omega_1 & & & \\ & & & & \omega_2 & & \\ & & & & & \ddots & \\ & & & & & & \omega_m \end{bmatrix}$$

Q.E.D.

If Λ is not degenerate *within any given subspace*, $\lambda_i^{(k)} \neq \lambda_i^{(l)}$, for any k, l, and i, the basis we end up with is unique: the freedom Ω gave us in each eigenspace is fully eliminated by Λ. The elements of this basis may be named uniquely by the pair of indices ω and λ as $|\omega, \lambda\rangle$, with λ playing the role of the extra label α. If Λ is degenerate within an eigenspace of Ω, if say $\lambda_1^{(1)} = \lambda_1^{(2)}$, there is a two-dimensional eigenspace from which we can choose any two orthonormal vectors for the common basis. It is then necessary to bring in a third operator Γ, that commutes with both Ω and Λ, and which will be nondegenerate in this subspace. In general, one can always find, for finite n, a set of operators $\{\Omega, \Lambda, \Gamma, \ldots\}$ that commute with each other and that nail down a unique, common, eigenbasis, the elements of which may be labeled unambiguously as $|\omega, \lambda, \gamma, \ldots\rangle$. In our study of quantum mechanics it will be assumed that such a *complete set of commuting operators* exists if n is infinite.

*Exercise 1.8.10.** By considering the commutator, show that the following Hermitian matrices may be simultaneously diagonalized. Find the eigenvectors common to both and verify that under a unitary transformation to this basis, both matrices are diagonalized.

$$\Omega = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad , \quad \Lambda = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

Since Ω is degenerate and Λ is not, you must be prudent in deciding which matrix dictates the choice of basis.

Example 1.8.6. We will now discuss, in some detail, the complete solution to a problem in mechanics. It is important that you understand this example thoroughly, for it not only illustrates the use of the mathematical techniques developed in this chapter but also contains the main features of the central problem in quantum mechanics.

The mechanical system in question is depicted in Fig. 1.5. The two masses m are coupled to each other and the walls by springs of force constant k. If x_1 and x_2 measure the displacements of the masses from their equilibrium points, these coordinates obey the following equations, derived through an elementary application of Newton's laws:

$$\ddot{x}_1 = -\frac{2k}{m} x_1 + \frac{k}{m} x_2 \tag{1.8.24a}$$

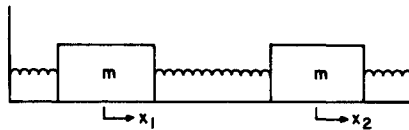$$\ddot{x}_2 = \frac{k}{m} x_1 - \frac{2k}{m} x_2 \tag{1.8.24b}$$

Figure 1.5. The coupled mass problem. All masses are m, all spring constants are k, and the displacements of the masses from equilibrium are x_1 and x_2.

The problem is to find $x_1(t)$ and $x_2(t)$ given the initial-value data, which in this case consist of the initial positions and velocities. If we restrict ourselves to the case of zero initial velocities, our problem is to find $x_1(t)$ and $x_2(t)$, given $x_1(0)$ and $x_2(0)$.

In what follows, we will formulate the problem in the language of linear vector spaces and solve it using the machinery developed in this chapter. As a first step, we rewrite Eq. (1.8.24) in matrix form:

$$\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad (1.8.25a)$$

where the elements of the *Hermitian* matrix Ω_{ij} are

$$\Omega_{11} = \Omega_{22} = -2k/m, \qquad \Omega_{12} = \Omega_{21} = k/m \qquad (1.8.25b)$$

We now view x_1 and x_2 as components of an abstract vector $|x\rangle$, and Ω_{ij} as the matrix elements of a Hermitian operator Ω. Since the vector $|x\rangle$ has two real components, it is an element of $\mathbb{V}^2(R)$, and Ω is a Hermitian operator on $\mathbb{V}^2(R)$. The abstract form of Eq. (1.8.25a) is

$$|\ddot{x}(t)\rangle = \Omega|x(t)\rangle \qquad (1.8.26)$$

Equation (1.8.25a) is obtained by projecting Eq. (1.8.26) on the basis vectors $|1\rangle$, $|2\rangle$, which have the following physical significance:

$$|1\rangle \leftrightarrow \begin{bmatrix} 1 \\ 0 \end{bmatrix} \leftrightarrow \begin{bmatrix} \text{first mass displaced by unity} \\ \text{second mass undisplaced} \end{bmatrix} \qquad (1.8.27a)$$

$$|2\rangle \leftrightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} \leftrightarrow \begin{bmatrix} \text{first mass undisplaced} \\ \text{second mass displaced by unity} \end{bmatrix} \qquad (1.8.27b)$$

An arbitrary state, in which the masses are displaced by x_1 and x_2, is given in this basis by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x_2 \qquad (1.8.28)$$

The abstract counterpart of the above equation is

$$|x\rangle = |1\rangle x_1 + |2\rangle x_2 \qquad (1.8.29)$$

It is in this $|1\rangle$, $|2\rangle$ basis that Ω is represented by the matrix appearing in Eq. (1.8.25), with elements $-2k/m$, k/m, etc.

The basis $|1\rangle$, $|2\rangle$ is very desirable physically, for the components of $|x\rangle$ in this basis (x_1 and x_2) have the simple interpretation as displacements of the masses. However, from the standpoint of finding a mathematical solution to the initial-value problem, it is not so desirable, for the components x_1 and x_2 obey the *coupled*

differential equations (1.8.24a) and (1.8.24b). The coupling is mediated by the off-diagonal matrix elements $\Omega_{12} = \Omega_{21} = k/m$.

Having identified the problem with the $|1\rangle$, $|2\rangle$ basis, we can now see how to get around it: we must switch to a basis in which Ω is diagonal. The components of $|x\rangle$ in this basis will then obey uncoupled differential equation which may be readily solved. Having found the solution, we can return to the physically preferable $|1\rangle$, $|2\rangle$ basis. This, then, is our broad strategy and we now turn to the details.

From our study of Hermitian operators we know that the basis that diagonalizes Ω is the basis of its normalized eigenvectors. Let $|I\rangle$ and $|II\rangle$ be its eigenvectors defined by

$$\Omega|I\rangle = -\omega_I^2 |I\rangle \tag{1.8.30a}$$

$$\Omega|II\rangle = -\omega_{II}^2 |II\rangle \tag{1.8.30b}$$

We are departing here from our usual notation: the eigenvalue of Ω is written as $-\omega^2$ rather than as ω in anticipation of the fact that Ω has eigenvalues of the form $-\omega^2$, with ω real. We are also using the symbols $|I\rangle$ and $|II\rangle$ to denote what should be called $|-\omega_I^2\rangle$ and $|-\omega_{II}^2\rangle$ in our convention.

It is a simple exercise (which you should perform) to solve the eigenvalue problem of Ω in the $|1\rangle$, $|2\rangle$ basis (in which the matrix elements of Ω are known) and to obtain

$$\omega_I = \left(\frac{k}{m}\right)^{1/2}, \qquad |I\rangle \leftrightarrow \frac{1}{2^{1/2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{1.8.31a}$$

$$\omega_{II} = \left(\frac{3k}{m}\right)^{1/2}, \qquad |II\rangle \leftrightarrow \frac{1}{2^{1/2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{1.8.31b}$$

If we now expand the vector $|x(t)\rangle$ in this new basis as

$$|x(t)\rangle = |I\rangle x_I(t) + |II\rangle x_{II}(t) \tag{1.8.32}$$

[in analogy with Eq. (1.8.29)], the components x_I and x_{II} will evolve as follows:

$$\begin{bmatrix} \ddot{x}_I \\ \ddot{x}_{II} \end{bmatrix} = \begin{bmatrix} -\omega_I^2 & 0 \\ 0 & -\omega_{II}^2 \end{bmatrix}\begin{bmatrix} x_I \\ x_{II} \end{bmatrix}$$

$$= \begin{bmatrix} -\omega_I^2 x_I \\ -\omega_{II}^2 x_{II} \end{bmatrix} \tag{1.8.33}$$

We obtain this equation by rewriting Eq. (1.8.24) in the $|I\rangle$, $|II\rangle$ basis in which Ω has its eigenvalues as the diagonal entries, and in which $|x\rangle$ has components x_I and

$$\frac{d^2}{dt^2} - \Omega$$

to both sides of the expansion of Eq. (1.8.32), and get

$$|0\rangle = |I\rangle(\ddot{x}_I + \omega_I^2 x_I) + |II\rangle(\ddot{x}_{II} + \omega_{II}^2 x_{II}) \qquad (1.8.34)$$

Since $|I\rangle$ and $|II\rangle$ are orthogonal, each coefficient is zero.

The solution to the *decoupled* equations

$$\ddot{x}_i + \omega_i^2 x_i = 0, \qquad i = I, II \qquad (1.8.35)$$

subject to the condition of vanishing initial velocities, is

$$x_i(t) = x_i(0) \cos \omega_i t, \qquad i = I, II \qquad (1.8.36)$$

As anticipated, the components of $|x\rangle$ in the $|I\rangle$, $|II\rangle$ basis obey decoupled equations that can be readily solved. Feeding Eq. (1.8.36) into Eq. (1.8.32) we get

$$|x(t)\rangle = |I\rangle x_I(0) \cos \omega_I t + |II\rangle x_{II}(0) \cos \omega_{II} t \qquad (1.8.37a)$$

$$= |I\rangle\langle I | x(0)\rangle \cos \omega_I t + |II\rangle\langle II | x(0)\rangle \cos \omega_{II} t \qquad (1.8.37b)$$

Equation (1.8.37) provides the explicit solution to the initial-value problem. It corresponds to the following algorithm for finding $|x(t)\rangle$ given $|x(0)\rangle$.

Step (1). Solve the eigenvalue problem of Ω.

Step (2). Find the coefficients $x_I(0) = \langle I | x(0)\rangle$ and $x_{II}(0) = \langle II | x(0)\rangle$ in the expansion

$$|x(0)\rangle = |I\rangle x_I(0) + |II\rangle x_{II}(0)$$

Step (3). Append to each coefficient $x_i(0)$ ($i = I, II$) a time dependence $\cos \omega_i t$ to get the coefficients in the expansion of $|x(t)\rangle$.

Let me now illustrate this algorithm by solving the following (general) initial-value problem: Find the future state of the system given that at $t = 0$ the masses are displaced by $x_1(0)$ and $x_2(0)$.

Step (1). We can ignore this step since the eigenvalue problem has been solved [Eq. (1.8.31)].

Step (2).

$$x_{\mathrm{I}}(0) = \langle \mathrm{I}|x(0)\rangle = \frac{1}{2^{1/2}}(1, 1)\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \frac{x_1(0) + x_2(0)}{2^{1/2}}$$

$$x_{\mathrm{II}}(0) = \langle \mathrm{II}|x(0)\rangle = \frac{1}{2^{1/2}}(1, -1)\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \frac{x_1(0) - x_2(0)}{2^{1/2}}$$

Step (3).

$$|x(t)\rangle = |\mathrm{I}\rangle \frac{x_1(0) + x_2(0)}{2^{1/2}} \cos \omega_{\mathrm{I}} t + |\mathrm{II}\rangle \frac{x_1(0) - x_2(0)}{2^{1/2}} \cos \omega_{\mathrm{II}} t$$

The explicit solution above can be made even more explicit by projecting $|x(t)\rangle$ onto the $|1\rangle$, $|2\rangle$ basis to find $x_1(t)$ and $x_2(t)$, the displacements of the masses. We get (feeding in the explicit formulas for ω_{I} and ω_{II})

$$x_1(t) = \langle 1|x(t)\rangle$$

$$= \langle 1|\mathrm{I}\rangle \frac{x_1(0) + x_2(0)}{2^{1/2}} \cos\left[\left(\frac{k}{m}\right)^{1/2} t\right] + \langle 1|\mathrm{II}\rangle \frac{x_1(0) - x_2(0)}{2^{1/2}} \cos\left[\left(\frac{3k}{m}\right)^{1/2} t\right]$$

$$= \frac{1}{2}[x_1(0) + x_2(0)] \cos\left[\left(\frac{k}{m}\right)^{1/2} t\right] + \frac{1}{2}[x_1(0) - x_2(0)] \cos\left[\left(\frac{3k}{m}\right)^{1/2} t\right] \qquad (1.8.38a)$$

using the fact that

$$\langle 1|\mathrm{I}\rangle = \langle 1|\mathrm{II}\rangle = 1/2^{1/2}$$

It can likewise be shown that

$$x_2(t) = \frac{1}{2}[x_1(0) + x_2(0)] \cos\left[\left(\frac{k}{m}\right)^{1/2} t\right] - \frac{1}{2}[x_1(0) - x_2(0)] \cos\left[\left(\frac{3k}{m}\right)^{1/2} t\right] \qquad (1.8.38b)$$

We can rewrite Eq. (1.8.38) in matrix form as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \dfrac{\cos[(k/m)^{1/2}t] + \cos[(3k/m)^{1/2}t]}{2} & \dfrac{\cos[(k/m)^{1/2}t] - \cos[(3k/m)^{1/2}t]}{2} \\ \dfrac{\cos[(k/m)^{1/2}t] - \cos[(3k/m)^{1/2}t]}{2} & \dfrac{\cos[(k/m)^{1/2}t] + \cos[(3k/m)^{1/2}t]}{2} \end{bmatrix}$$

$$\times \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} \qquad (1.8.39)$$

This completes our determination of the future state of the system given the initial state.

The Propagator

There are two remarkable features in Eq. (1.8.39):

(1) The final-state vector is obtained from the initial-state vector upon multiplication by a matrix.
(2) This matrix is independent of the initial state. We call this matrix the *propagator*. Finding the propagator is tantamount to finding the complete solution to the problem, for given any other initial state with displacements $\tilde{x}_1(0)$ and $\tilde{x}_2(0)$, we get $\tilde{x}_1(t)$ and $\tilde{x}_2(t)$ by applying the same matrix to the initial-state vector.

We may view Eq. (1.8.39) as the image in the $|1\rangle$, $|2\rangle$ basis of the abstract relation

$$|x(t)\rangle = U(t)|x(0)\rangle \qquad (1.8.40)$$

By comparing this equation with Eq. (1.8.37b), we find the abstract representation of U:

$$U(t) = |I\rangle\langle I| \cos \omega_I t + |II\rangle\langle II| \cos \omega_{II} t \qquad (1.8.41a)$$

$$= \sum_{i=I}^{II} |i\rangle\langle i| \cos \omega_i t \qquad (1.8.41b)$$

You may easily convince yourself that if we take the matrix elements of this operator in the $|1\rangle$, $|2\rangle$ basis, we regain the matrix appearing in Eq. (1.8.39). For example

$$U_{11} = \langle 1| U| 1\rangle$$

$$= \langle 1| \left\{ |I\rangle\langle I| \cos\left[\left(\frac{k}{m}\right)^{1/2} t\right] + |II\rangle\langle II| \cos\left[\left(\frac{3k}{m}\right)^{1/2} t\right] \right\} |1\rangle$$

$$= \langle 1|I\rangle\langle I|1\rangle \cos\left[\left(\frac{k}{m}\right)^{1/2} t\right] + \langle 1|II\rangle\langle II|1\rangle \cos\left[\left(\frac{3k}{m}\right)^{1/2} t\right]$$

$$= \frac{1}{2} \left\{ \cos\left[\left(\frac{k}{m}\right)^{1/2} t\right] + \cos\left[\left(\frac{3k}{m}\right)^{1/2} t\right] \right\}$$

Notice that $U(t)$ [Eq. (1.8.41)] is determined completely by the eigenvectors and eigenvalues of Ω. We may then restate our earlier algorithm as follows. To solve the equation

$$|\ddot{x}\rangle = \Omega|x\rangle$$

(1) Solve the eigenvalue problem of Ω.

(2) Construct the propagator U in terms of the eigenvalues and eigenvectors.

(3) $|x(t)\rangle = U(t)|x(0)\rangle$.

The Normal Modes

There are two initial states $|x(0)\rangle$ for which the time evolution is particularly simple. Not surprisingly, these are the eigenkets $|I\rangle$ and $|II\rangle$. Suppose we have $|x(0)\rangle = |I\rangle$. Then the state at time t is

$$|I(t)\rangle \equiv U(t)|I\rangle$$
$$= (|I\rangle\langle I| \cos \omega_I t + |II\rangle\langle II| \cos \omega_{II} t)|I\rangle$$
$$= |I\rangle \cos \omega_I t \qquad (1.8.42)$$

Thus the system starting off in $|I\rangle$ is only modified by an overall factor $\cos \omega_I t$. A similar remark holds with $I \to II$. These two modes of vibration, in which all (two) components of a vector oscillate in step are called *normal modes*.

The physics of the normal modes is clear in the $|1\rangle$, $|2\rangle$ basis. In this basis

$$|I\rangle \leftrightarrow \frac{1}{2^{1/2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and corresponds to a state in which both masses are displaced by equal amounts. The middle spring is then a mere spectator and each mass oscillates with a frequency $\omega_I = (k/m)^{1/2}$ in response to the end spring nearest to it. Consequently

$$|I(t)\rangle \leftrightarrow = \frac{1}{2^{1/2}}\begin{bmatrix} \cos[(k/m)^{1/2}t] \\ \cos[(k/m)^{1/2}t] \end{bmatrix}$$

On the other hand, if we start with

$$|II\rangle \leftrightarrow \frac{1}{2^{1/2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

the masses are displaced by equal and opposite amounts. In this case the middle spring is distorted by *twice* the displacement of each mass. If the masses are adjusted by Δ and $-\Delta$, respectively, each mass feels a restoring force of $3k\Delta$ ($2k\Delta$ from the middle spring and $k\Delta$ from the end spring nearest to it). Since the effective force constant is $k_{eff} = 3k\Delta/\Delta = 3k$, the vibrational frequency is $(3k/m)^{1/2}$ and

$$|II\rangle(t)\rangle \leftrightarrow = \frac{1}{2^{1/2}}\begin{bmatrix} \cos[(3k/m)^{1/2}t] \\ -\cos[(3k/m)^{1/2}t] \end{bmatrix}$$

If the system starts off in a linear combination of $|I\rangle$ and $|II\rangle$ it evolves into the corresponding linear combination of the normal modes $|I(t)\rangle$ and $|II(t)\rangle$. This

$$|x(t)\rangle = U(t)|x(0)\rangle$$

$$= |\text{I}\rangle\langle\text{I}|x(0)\rangle \cos \omega_{\text{I}} t + |\text{II}\rangle\langle\text{II}| \, x(0)\rangle \cos \omega_{\text{II}} t$$

$$= |\text{I}(t)\rangle \cdot \langle\text{I}|x(0)\rangle + |\text{II}(t)\rangle\langle\text{II}|x(0)\rangle$$

Another way to see the simple evolution of the initial states $|\text{I}\rangle$ and $|\text{II}\rangle$ is to determine the matrix representing U in the $|\text{I}\rangle$, $|\text{II}\rangle$ basis:

$$U \underset{\substack{\text{I,II} \\ \text{basis}}}{\longleftrightarrow} \begin{bmatrix} \cos \omega_{\text{I}} t & 0 \\ 0 & \cos \omega_{\text{II}} t \end{bmatrix} \qquad (1.8.43)$$

You should verify this result by taking the appropriate matrix elements of $U(t)$ in Eq. (1.8.41b). Since each column above is the image of the corresponding basis vectors ($|\text{I}\rangle$ or $|\text{II}\rangle$) after the action of $U(t)$, (which is to say, after time evolution), we see that the initial states $|\text{I}\rangle$ and $|\text{II}\rangle$ evolve simply in time.

The central problem in quantum mechanics is very similar to the simple example that we have just discussed. The state of the system is described in quantum theory by a ket $|\psi\rangle$ which obeys the Schrödinger equation

$$i\hbar|\dot{\psi}\rangle = H|\psi\rangle$$

where $\hbar$ is a constant related to Planck's constant h by $\hbar = h/2\pi$, and H is a *Hermitian* operator called the Hamiltonian. The problem is to find $|\psi(t)\rangle$ given $|\psi(0)\rangle$. [Since the equation is first order in t, no assumptions need be made about $|\dot{\psi}(0)\rangle$, which is determined by the Schrödinger equation to be $(-i/\hbar)H|\psi(0)\rangle$.]

In most cases, H is a time-independent operator and the algorithm one follows in solving this initial-value problem is completely analogous to the one we have just seen:

Step (1). Solve the eigenvalue problem of H.

Step (2). Find the propagator $U(t)$ in terms of the eigenvectors and eigenvalues of H.

Step (3). $|\psi(t)\rangle = U(t)|\psi(0)\rangle$.

You must of course wait till Chapter 4 to find out the physical interpretation of $|\psi\rangle$, the actual form of the operator H, and the precise relation between $U(t)$ and the eigenvalues and eigenvectors of H. ▢

Exercise 1.8.11. Consider the coupled mass problem discussed above.

(1) Given that the initial state is $|\text{I}\rangle$, in which the first mass is displaced by unity and the second is left alone, calculate $|\text{I}(t)\rangle$ by following the algorithm.

(2) Compare your result with that following from Eq. (1.8.39).

Exercise 1.8.12. Consider once again the problem discussed in the previous example. (1) Assuming that

$$|\ddot{x}\rangle = \Omega|x\rangle$$

has a solution

$$|x(t)\rangle = U(t)|x(0)\rangle$$

find the differential equation satisfied by $U(t)$. Use the fact that $|x(0)\rangle$ is arbitrary.

(2) Assuming (as is the case) that Ω and U can be simultaneously diagonalized, solve for the elements of the matrix U in this common basis and regain Eq. (1.8.43). Assume $|\dot{x}(0)\rangle = 0$.

1.9. Functions of Operators and Related Concepts

We have encountered two types of objects that act on vectors: scalars, which commute with each other and with all operators; and operators, which do not generally commute with each other. It is customary to refer to the former as c numbers and the latter as q numbers. Now, we are accustomed to functions of c numbers such as $\sin(x)$, $\log(x)$, etc. We wish to examine the question whether functions of q numbers can be given a sensible meaning. We will restrict ourselves to those functions that can be written as a power series. Consider a series

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \tag{1.9.1}$$

where x is a c number. We define the same function of an operator or q number to be

$$f(\Omega) = \sum_{n=0}^{\infty} a_n \Omega^n \tag{1.9.2}$$

This definition makes sense only if the sum converges to a definite limit. To see what this means, consider a common example:

$$e^{\Omega} = \sum_{n=1}^{\infty} \frac{\Omega^n}{n!} \tag{1.9.3}$$

Let us restrict ourselves to Hermitian Ω. By going to the eigenbasis of Ω we can readily perform the sum of Eq. (1.9.3). Since

$$\Omega = \begin{bmatrix} \omega_1 & & & \\ & \omega_2 & & \\ & & \ddots & \\ & & & \omega_n \end{bmatrix} \tag{1.9.4}$$

$$\Omega^m = \begin{bmatrix} \omega_1^m & & & \\ & \omega_2^m & & \\ & & \ddots & \\ & & & \omega_n^m \end{bmatrix} \qquad (1.9.5)$$

$$e^{\Omega} = \begin{bmatrix} \sum\limits_{m=0}^{\infty} \dfrac{\omega_1^m}{m!} & & \\ & \ddots & \\ & & \sum\limits_{m=0}^{\infty} \dfrac{\omega_n^m}{m!} \end{bmatrix} \qquad (1.9.6)$$

Since each sum converges to the familiar limit e^{ω_i}, the operator e^{Ω} is indeed well defined by the power series in this basis (and therefore in any other).

*Exercise 1.9.1.** We know that the series

$$f(x) = \sum_{n=0}^{\infty} x^n$$

may be equated to the function $f(x) = (1-x)^{-1}$ if $|x| < 1$. By going to the eigenbasis, examine when the q number power series

$$f(\Omega) = \sum_{n=0}^{\infty} \Omega^n$$

of a Hermitian operator Ω may be identified with $(1-\Omega)^{-1}$.

*Exercise 1.9.2.** If H is a Hermitian operator, show that $U = e^{iH}$ is unitary. (Notice the analogy with c numbers: if θ is real, $u = e^{i\theta}$ is a number of unit modulus.)

Exercise 1.9.3. For the case above, show that $\det U = e^{i\mathrm{Tr}\,H}$.

Derivatives of Operators with Respect to Parameters

Consider next an operator $\theta(\lambda)$ that depends on a parameter λ. Its derivative with respect to λ is defined to be

$$\frac{d\theta(\lambda)}{d\lambda} = \lim_{\Delta\lambda \to 0} \left[\frac{\theta(\lambda + \Delta\lambda) - \theta(\lambda)}{\Delta\lambda} \right]$$

If $\theta(\lambda)$ is written as a matrix in some basis, then the matrix representing $d\theta(\lambda)/d\lambda$ is obtained by differentiating the matrix elements of $\theta(\lambda)$. A special case of $\theta(\lambda)$ we

are interested in is

$$\theta(\lambda) = e^{\lambda\Omega}$$

where Ω is Hermitian. We can show, by going to the eigenbasis of Ω, that

$$\frac{d\theta(\lambda)}{d\lambda} = \Omega e^{\lambda\Omega} = e^{\lambda\Omega}\Omega = \theta(\lambda)\Omega \qquad (1.9.7)$$

The same result may be obtained, even if Ω is not Hermitian, by working with the power series, provided it exists:

$$\frac{d}{d\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n\Omega^n}{n!} = \sum_{n=1}^{\infty}\frac{n\lambda^{n-1}\Omega^n}{n!} = \Omega\sum_{n=1}^{\infty}\frac{\lambda^{n-1}\Omega^{n-1}}{(n-1)!} = \Omega\sum_{m=0}^{\infty}\frac{\lambda^m\Omega^m}{m!} = \Omega e^{\lambda\Omega}$$

Conversely, we can say that if we are confronted with the differential Eq. (1.9.7), its solution is given by

$$\theta(\lambda) = c\exp\left(\int_0^\lambda \Omega \, d\lambda'\right) = c\exp(\Omega\lambda)$$

(It is assumed here that the exponential exists.) In the above, c is a constant (operator) of integration. The solution $\theta = e^{\Omega\lambda}$ corresponds to the choice $c = \text{I}$.

In all the above operations, we see that Ω behaves as if it were just a c number. Now, the real difference between c numbrs and q numbers is that the latter do not generally commute. However, if only one q number (or powers of it) enter the picture, everything commutes and we can treat them as c numbers. If one remembers this mnemonic, one can save a lot of time.

If, on the other hand, more than one q number is involved, the order of the factors is all important. For example, it is true that

$$e^{\alpha\Omega}e^{\beta\Omega} = e^{(\alpha+\beta)\Omega}$$

as may be verified by a power-series expansion, while it is not true that

$$e^{\alpha\Omega}e^{\beta\theta} = e^{\alpha\Omega+\beta\theta}$$

or that

$$e^{\alpha\Omega}e^{\beta\theta}e^{-\alpha\Omega} = e^{\beta\theta}$$

unless $[\Omega, \theta] = 0$. Likewise, in differentiating a product, the chain rule is

$$\frac{d}{d\lambda}e^{\lambda\Omega}e^{\lambda\theta} = \Omega e^{\lambda\Omega}e^{\lambda\theta} + e^{\lambda\Omega}e^{\lambda\theta}\theta \qquad (1.9.8)$$

We are free to move Ω through $e^{\lambda\Omega}$ and write the first term as

$$e^{\lambda\Omega}\Omega e^{\lambda\theta}$$

but not as

$$e^{\lambda\Omega}e^{\lambda\theta}\Omega$$

unless $[\Omega, \theta] = 0$.

1.10. Generalization to Infinite Dimensions

In all of the preceding discussions, the dimensionality (n) of the space was unspecified but assumed to be some finite number. We now consider the generalization of the preceding concepts to infinite dimensions.

Let us begin by getting acquainted with an infinite-dimensional vector. Consider a function defined in some interval, say, $a \le x \le b$. A concrete example is provided by the displacement $f(x, t)$ of a string clamped at $x = 0$ and $x = L$ (Fig. 1.6).

Suppose we want to communicate to a person on the moon the string's displacement $f(x)$, at some time t. One simple way is to divide the interval $0 - L$ into 20 equal parts, measure the displacement $f(x_i)$ at the 19 points $x = L/20, 2L/20, \ldots, 19L/20$, and transmit the 19 values on the wireless. Given these $f(x_i)$, our friend on the moon will be able to reconstruct the approximate picture of the string shown in Fig. 1.7.

If we wish to be more accurate, we can specify the values of $f(x)$ at a larger number of points. Let us denote by $f_n(x)$ the discrete approximation to $f(x)$ that coincides with it at n points and vanishes in between. Let us now interpret the ordered n-tuple $\{f_n(x_1), f_n(x_2), \ldots, f_n(x_n)\}$ as components of a ket $|f_n\rangle$ in a vector space $\mathbb{V}^n(R)$:

$$|f_n\rangle \leftrightarrow \begin{bmatrix} f_n(x_1) \\ f_n(x_2) \\ \vdots \\ f_n(x_n) \end{bmatrix} \qquad (1.10.1)$$
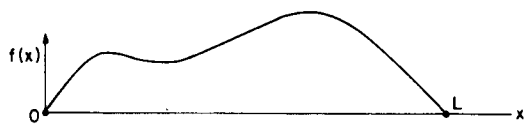

Figure 1.6. The string is clamped at $x = 0$ and $x = L$. It is free to oscillate in the plane of the paper.
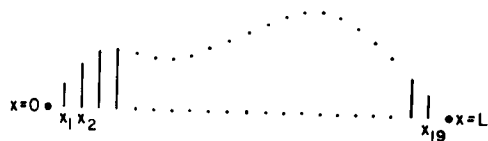

Figure 1.7. The string as reconstructed by the person on the moon.

The basis vectors in this space are

$$|x_i\rangle \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{th place} \qquad (1.10.2)$$

corresponding to the discrete function which is unity at $x = x_i$ and zero elsewhere. The basis vectors satisfy

$$\langle x_i | x_j \rangle | = \delta_{ij} \text{ (orthogonality)} \qquad (1.10.3)$$

$$\sum_{i=1}^{n} |x_i\rangle\langle x_i| = I \text{ (completeness)} \qquad (1.10.4)$$

Try to imagine a space containing n mutually perpendicular axes, one for each point x_i. Along each axis is a unit vector $|x_i\rangle$. The function $f_n(x)$ is represented by a vector whose projection along the ith direction is $f_n(x_i)$:

$$|f_n\rangle = \sum_{i=1}^{n} f_n(x_i)|x_i\rangle \qquad (1.10.5)$$

To every possible discrete approximation $g_n(x)$, $h_n(x)$, etc., there is a corresponding ket $|g_n\rangle$, $|h_n\rangle$, etc., and vice versa. You should convince yourself that if we define vector addition as the addition of the components, and scalar multiplication as the multiplication of each component by the scalar, then the set of all kets representing discrete functions that vanish at $x = 0$, L and that are specified at n points in between, forms a vector space.

We next define the inner product in this space:

$$\langle f_n | g_n \rangle = \sum_{i=1}^{n} f_n(x_i)g_n(x_i) \qquad (1.10.6)$$

Two functions $f_n(x)$ and $g_n(x)$ will be said to be orthogonal if $\langle f_n | g_n \rangle = 0$.

Let us now forget the man on the moon and consider the maximal specification of the string's displacement, by giving its value at every point in the interval $0 - L$. In this case $f_\infty(x) \equiv f(x)$ is specified by an ordered infinity of numbers: an $f(x)$ for each point x. Each function is now represented by a ket $|f_\infty\rangle$ in an infinite-dimensional vector space and vice versa. Vector addition and scalar multiplication are defined just as before. Consider, however, the inner product. For finite n it was

defined as

$$\langle f_n | g_n \rangle = \sum_{i=1}^{n} f_n(x_i) g_n(x_i)$$

in particular

$$\langle f_n | f_n \rangle = \sum_{i=1}^{n} [f_n(x_i)]^2$$

If we now let n go to infinity, so does the sum, for practically any function. What we need is the redefinition of the inner product for finite n in such a way that as n tends to infinity, a smooth limit obtains. The natural choice is of course

$$\langle f_n | g_n \rangle = \sum_{i=1}^{n} f_n(x_i) g_n(x_i) \Delta, \qquad \Delta = L/(n+1) \tag{1.10.6'}$$

If we now let n go to infinity, we get, by the usual definition of the integral,

$$\langle f | g \rangle = \int_0^L f(x) g(x) \, dx \tag{1.10.7}$$

$$\langle f | f \rangle = \int_0^L f^2(x) \, dx \tag{1.10.8}$$

If we wish to go beyond the instance of the string and consider complex functions of x as well, in some interval $a \le x \le b$, the only modification we need is in the inner product:

$$\langle f | g \rangle = \int_a^b f^*(x) g(x) \, dx \tag{1.10.9}$$

What are the basis vectors in this space and how are they normalized? We know that each point x gets a basis vector $|x\rangle$. The orthogonality of two different axes requires that

$$\langle x | x' \rangle = 0, \qquad x \ne x' \tag{1.10.10}$$

What if $x = x'$? Should we require, as in the finite-dimensional case, $\langle x | x \rangle = 1$? The answer is no, and the best way to see it is to deduce the correct normalization. We start with the natural generalization of the completeness relation Eq. (1.10.4) to the case where the kets are labeled by a continuous index x':

$$\int_a^b |x'\rangle\langle x'| \, dx' = I \tag{1.10.11}$$

where, as always, the identity is required to leave each ket unchanged. Dotting both sides of Eq. (1.10.11) with some arbitrary ket $|f\rangle$ from the right and the basis bra $\langle x|$ from the left,

$$\int_a^b \langle x|x'\rangle\langle x'|f\rangle\, dx' = \langle x|\mathrm{I}|f\rangle = \langle x|f\rangle \qquad (1.10.12)$$

Now, $\langle x|f\rangle$, the projection of $|f\rangle$ along the basis ket $|x\rangle$, is just $f(x)$. Likewise $\langle x'|f\rangle = f(x')$. Let the inner product $\langle x|x'\rangle$ be some unknown function $\delta(x, x')$. Since $\delta(x, x')$ vanishes if $x \neq x'$ we can restrict the integral to an infinitesimal region near $x' = x$ in Eq. (1.10.2):

$$\int_{x-\varepsilon}^{x+\varepsilon} \delta(x, x')f(x')\, dx' = f(x) \qquad (1.10.13)$$

In this infinitesimal region, $f(x')$ (for any reasonably smooth f) can be approximated by its value at $x' = x$, and pulled out of the integral:

$$f(x) \int_{x-\varepsilon}^{x+\varepsilon} \delta(x, x')\, dx' = f(x) \qquad (1.10.14)$$

so that

$$\int_{x-\varepsilon}^{x+\varepsilon} \delta(x, x')\, dx' = 1 \qquad (1.10.15)$$

Clearly $\delta(x, x')$ cannot be finite at $x' = x$, for then its integral over an infinitesimal region would also be infinitesimal. In fact $\delta(x, x')$ should be infinite in such a way that its integral is unity. Since $\delta(x, x')$ depends only on the difference $x - x'$, let us write it as $\delta(x - x')$. The "function," $\delta(x - x')$, with the properties

$$\delta(x - x') = 0, \qquad x \neq x'$$
$$\int_a^b \delta(x - x')\, dx' = 1, \qquad a < x < b \qquad (1.10.16)$$

is called the *Dirac delta function* and fixes the normalization of the basis vectors:

$$\langle x|x'\rangle = \delta(x - x') \qquad (1.10.17)$$

It will be needed any time the basis kets are labeled by a continuous index such as x. Note that it is defined only in the context of an integration: the integral of the delta function $\delta(x - x')$ with any smooth function $f(x')$ is $f(x)$. One sometimes calls
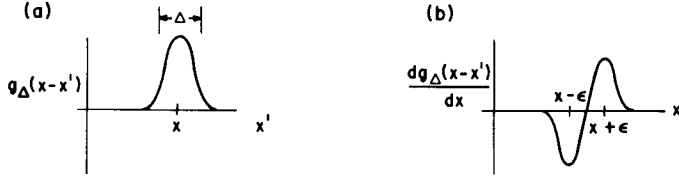
Figure 1.8. (a) The Gaussian g_Δ approaches the delta function as $\Delta \to 0$. (b) Its derivative $(dg/dx)(x-x')$ approaches $\delta'(x-x')$ as $\Delta \to 0$.

the delta function the sampling function, since it samples the value of the function $f(x')$ at one point‡

$$\int \delta(x-x')f(x')\,dx = f(x) \tag{1.10.18}$$

The delta function does not look like any function we have seen before, its values being either infinite or zero. It is therefore useful to view it as the limit of a more conventional function. Consider a Gaussian

$$g_\Delta(x-x') = \frac{1}{(\pi\Delta^2)^{1/2}} \exp\left[-\frac{(x-x')^2}{\Delta^2} \right] \tag{1.10.19}$$

as shown in Fig. 1.8a. The Gaussian is centered at $x'=x$, has wdith Δ, maximum height $(\pi\Delta^2)^{1/2}$, and unit area, independent of Δ. As Δ approaches zero, g_Δ becomes a better and better approximation to the delta function.§

It is obvious from the Gaussian model that the delta function is even. This may be verified as follows:

$$\delta(x-x') = \langle x|x'\rangle = \langle x'|x\rangle^* = \delta(x'-x)^* = \delta(x'-x)$$

since the delta function is real.

Consider next an object that is even more peculiar than the delta function: its derivative with respect to the *first* argument x:

$$\delta'(x-x') = \frac{d}{dx}\delta(x-x') = -\frac{d}{dx'}\delta(x-x') \tag{1.10.20}$$

What is the action of this function under the integral? The clue comes from the Gaussian model. Consider $dg_\Delta(x-x')/dx = -dg_\Delta(x-x')/dx'$ as a function of x'. As

‡ We will often omit the limits of integration if they are unimportant.
§ A fine point that will not concern you till Chapter 8: This formula for the delta function is valid even if Δ^2 is pure imaginary, say, equal to $i\beta^2$. First we see from Eq. (A.2.5) that g has unit area. Consider next the integral of g times $f(x')$ over a region in x' that includes x. For the most part, we get zero because f is smooth and g is wildly oscillating as $\beta \to 0$. However, at $x=x'$, the derivative of the phase of g vanishes and the oscillations are suspended. Pulling $f(x'=x)$ out of the integral, we get the desired result.

g_Δ shrinks, each bump at $\pm \varepsilon$ will become, up to a scale factor, the δ function. The first one will sample $-f(x-\varepsilon)$ and the second one $+f(x+\varepsilon)$, again up to a scale, so that

$$\int \delta'(x-x') f(x') \, dx' \propto f(x+\varepsilon) - f(x-\varepsilon) = 2\varepsilon \left. \frac{df}{dx'} \right|_{x'=x}$$

The constant of proportionality happens to be $1/2\varepsilon$ so that

$$\int \delta'(x-x') f(x') \, dx = \left. \frac{df}{dx'} \right|_{x'=x} = \frac{df(x)}{dx} \qquad (1.10.21)$$

This result may be verified as follows:

$$\int \delta'(x-x') f(x') \, dx' = \int \frac{d\delta(x-x')}{dx} f(x') \, dx' = \frac{d}{dx} \int \delta(x-x') f(x') \, dx'$$

$$= \frac{d}{dx} f(x)$$

Note that $\delta'(x-x')$ is an odd function. This should be clear from Fig. 1.8b or Eq. (1.10.20). An equivalent way to describe the action of the δ' function is by the equation

$$\delta'(x-x') = \delta(x-x') \frac{d}{dx'} \qquad (1.10.22)$$

where it is understood that both sides appear in an integral over x' and that the differential operator acts on any function that accompanies the δ' function in the integrand. In this notation we can describe the action of higher derivatives of the delta function:

$$\frac{d^n \delta(x-x')}{dx^n} = \delta(x-x') \frac{d^n}{dx'^n} \qquad (1.10.23)$$

We will now develop an alternate representation of the delta function. We know from basic Fourier analysis that, given a function $f(x)$, we may define its transform

$$f(k) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-ikx} f(x) \, dx \qquad (1.10.24)$$

$$f(x') = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{ikx'} f(k) \, dk \qquad (1.10.25)$$

Feeding Eq. (1.10.24) into Eq. (1.10.25), we get

$$f(x') = \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} dk \, e^{ik(x'-x)} \right) f(x) \, dx$$

Comparing this result with Eq. (1.10.18), we see that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} dk \, e^{ik(x'-x)} = \delta(x'-x) \qquad (1.10.26)$$

*Exercise 1.10.1.** Show that $\delta(ax) = \delta(x)/|a|$. [Consider $\int \delta(ax) \, d(ax)$. Remember that $\delta(x) = \delta(-x)$.]

*Exercise 1.10.2.** Show that

$$\delta(f(x)) = \sum_i \frac{\delta(x_i - x)}{|df/dx_i|}$$

where x_i are the zeros of $f(x)$. Hint: Where does $\delta(f(x))$ blow up? Expand $f(x)$ near such points in a Taylor series, keeping the first nonzero term.

*Exercise 1.10.3.** Consider the *theta function* $\theta(x - x')$ which vanishes if $x - x'$ is negative and equals 1 if $x - x'$ is positive. Show that $\delta(x - x') = d/dx \, \theta(x - x')$.

Operators in Infinite Dimensions

Having acquainted ourselves with the elements of this function space, namely, the kets $|f\rangle$ and the basis vectors $|x\rangle$, let us turn to the (linear) operators that act on them. Consider the equation

$$\Omega |f\rangle = |\tilde{f}\rangle$$

Since the kets are in correspondence with the functions, Ω takes the function $f(x)$ into another, $\tilde{f}(x)$. Now, one operator that does such a thing is the familiar differential operator, which, acting on $f(x)$, gives $\tilde{f}(x) = df(x)/dx$. In the function space we can describe the action of this operator as

$$D |f\rangle = |df/dx\rangle$$

where $|df/dx\rangle$ is the ket corresponding to the function df/dx. What are the matrix elements of D in the $|x\rangle$ basis? To find out, we dot both sides of the above equation

with $\langle x |$,

$$\langle x | D | f \rangle = \left\langle x \left| \frac{df}{dx} \right. \right\rangle = \frac{df(x)}{dx}$$

and insert the resolution of identity at the right place

$$\int \langle x | D | x' \rangle \langle x' | f \rangle \, dx' = \frac{df}{dx} \qquad (1.10.27)$$

Comparing this to Eq. (1.10.21), we deduce that

$$\langle x | D | x' \rangle = D_{xx'} = \delta'(x - x') = \delta(x - x') \frac{d}{dx'} \qquad (1.10.28)$$

It is worth remembering that $D_{xx'} = \delta'(x - x')$ *is to be integrated over the second index* (x') *and pulls out the derivative of f at the first index* (x). Some people prefer to integrate $\delta'(x - x')$ over the first index, in which case it pulls out $-df/dx'$. Our convention is more natural if one views $D_{xx'}$ as a matrix acting to the *right* on the components $f_{x'} \equiv f(x')$ of a vector $| f \rangle$. Thus the familiar differential operator is an infinite-dimensional matrix with the elements given above. Normally one doesn't think of D as a matrix for the following reason. Usually when a matrix acts on a vector, there is a sum over a common index. In fact, Eq. (1.10.27) contains such a sum over the index x'. If, however, we feed into this equation the value of $D_{xx'}$, the delta function renders the integration trivial:

$$\int \delta(x - x') \frac{d}{dx'} f(x') \, dx' = \frac{df}{dx'} \bigg|_{x' = x} = \frac{df}{dx}$$

Thus the action of D is simply to apply d/dx to $f(x)$ with no sum over a common index in sight. Although we too will drop the integral over the common index ultimately, we will continue to use it for a while to remind us that D, like all linear operators, is a matrix.

Let us now ask if D is Hermitian and examine its eigenvalue problem. If D were Hermitian, we would have

$$D_{xx'} = D_{x'x}^*$$

But this is not the case:

$$D_{xx'} = \delta'(x - x')$$

while

$$D_{x'x}^* = \delta'(x' - x)^* = \delta'(x' - x) = -\delta'(x - x')$$

But we can easily convert D to a Hermitian matrix by multiplying it with a pure imaginary number. Consider

$$K = -iD$$

which satisfies

$$K_{x'x}^* = [-i\delta'(x'-x)]^* = +i\delta'(x'-x) = -i\delta'(x-x') = K_{xx'}$$

It turns out that despite the above, the operator K is not guaranteed to be Hermitian, as the following analysis will indicate. Let $|f\rangle$ and $|g\rangle$ be two kets in the function space, whose images in the X basis are two functions $f(x)$ and $g(x)$ in the interval $a-b$. If K is Hermitian, it must also satisfy

$$\langle g|K|f\rangle = \langle g|Kf\rangle = \langle Kf|g\rangle^* = \langle f|K^\dagger|g\rangle^* = \langle f|K|g\rangle^*$$

So we ask

$$\int_a^b \int_a^b \langle g|x\rangle\langle x|K|x'\rangle\langle x'|f\rangle \, dx \, dx'$$

$$\overset{?}{=} \left(\int_a^b \int_a^b \langle f|x\rangle\langle x|K|x'\rangle\langle x'|g\rangle \, dx \, dx'\right)^*$$

$$\int_a^b g^*(x)\left[-\frac{i\,df(x)}{dx}\right]dx \overset{?}{=} \left\{\int_a^b f^*(x)\left[-\frac{i\,dg(x)}{dx}\right]dx\right\}^* = i\int_a^b \frac{dg^*}{dx}f(x)\,dx$$

Integrating the left-hand side by parts gives

$$-ig^*(x)f(x)\Big|_a^{b^\circ} + i\int_a^b \frac{dg^*(x)}{dx}f(x)\,dx \qquad .$$

So K is Hermitian only if the *surface term* vanishes:

$$-ig^*(x)f(x)\Big|_a^b = 0 \qquad\qquad (1.10.29)$$

In contrast to the finite-dimensional case, $K_{xx'} = K_{x'x}^*$ is not a sufficient condition for K to be Hermitian. One also needs to look at the behavior of the functions at the end points a and b. Thus K is Hermitian in the space consists of functions that obey Eq. (1.10.29). One set of functions that obey this condition are the possible configurations $f(x)$ of the string clamped at $x = 0$, L, since $f(x)$ vanishes at the end points. But condition (1.10.29) can also be fulfilled in another way. Consider functions in our own three-dimensional space, parametrized by r, θ, and ϕ (ϕ is the angle measured around the z axis). Let us require that these functions be single

valued. In particular, if we start at a certain point and go once around the z axis, returning to the original point, the function must take on its original value, i.e.,

$$f(\phi) = f(\phi + 2\pi)$$

In the space of such periodic functions, $K = -i\,d/d\phi$ is a Hermitian operator. The surface term vanishes because the contribution from one extremity cancels that from the other:

$$-ig^*(\phi)f(\phi)\Big|_0^{2\pi} = -i[g^*(2\pi)f(2\pi) - g^*(0)f(0)] = 0$$

In the study of quantum mechanics, we will be interested in functions defined over the full interval $-\infty \leq x \leq +\infty$. They fall into two classes, those that vanish as $|x| \to \infty$, and those that do not, the latter behaving as e^{ikx}, k being a real parameter that labels these functions. It is clear that $K = -i\,d/dx$ is Hermitian when sandwiched between two functions of the first class or a function from each, since in either case the surface term vanishes. When sandwiched between two functions of the second class, the Hermiticity hinges on whether

$$e^{ikx}e^{-ik'x}\Big|_{-\infty}^{\infty} \overset{?}{=} 0$$

If $k = k'$, the contribution from one end cancels that from the other. If $k \neq k'$, the answer is unclear since $e^{i(k-k')x}$ oscillates, rather than approaching a limit as $|x| \to \infty$. Now, there exists a way of defining a limit for such functions that cannot make up their minds: the limit as $|x| \to \infty$ is defined to be the average over a large interval. According to this prescription, we have, say as $x \to \infty$,

$$\lim_{x \to \infty} e^{ikx}e^{-ik'x} = \lim_{\substack{L \to \infty \\ \Delta \to \infty}} \frac{1}{\Delta} \int_L^{\lambda+\Delta} e^{i(k-k')x}\,dx = 0 \qquad \text{if } k \neq k'$$

and so K is Hermitian in this space.

We now turn to the eigenvalue problem of K. The task seems very formidable indeed, for we have now to find the roots of an infinite-order characteristic polynomial and get the corresponding eigenvectors. It turns out to be quite simple and you might have done it a few times in the past without giving yourself due credit. Let us begin with

$$K|k\rangle = k|k\rangle \tag{1.10.30}$$

$$\langle x|K|k\rangle = k\langle x|k\rangle$$

$$\int \langle x|K|x'\rangle \langle x'|k\rangle \, dx' = k\psi_k(x) \qquad (1.10.31)$$

$$-i\frac{d}{dx}\psi_k(x) = k\psi_k(x)$$

where by definition $\psi_k(x) = \langle x|k\rangle$. This equation could have been written directly had we made the immediate substitution $K = -i\,d/dx$ in the X basis. From now on we shall resort to this shortcut unless there are good reasons for not doing so.

The solution to the above equation is simply

$$\psi_k(x) = A\,e^{ikx} \qquad (1.10.32)$$

where A, the overall scale, is a free parameter, unspecified by the eigenvalue problem. So the eigenvalue problem of K is fully solved: any real number k is an eigenvalue, and the corresponding eigenfunction is given by $A\,e^{ikx}$. As usual, the freedom in scale will be used to normalize the solution. We choose A to be $(1/2\pi)^{-1/2}$ so that

$$|k\rangle \leftrightarrow \frac{1}{(2\pi)^{1/2}}e^{ikx}$$

and

$$\langle k|k'\rangle = \int_{-\infty}^{\infty} \langle k|x\rangle\langle x|k'\rangle \, dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i(k-k')x}\,dx = \delta(k-k') \quad (1.10.33)$$

(Since $\langle k|k\rangle$ is infinite, no choice of A can normalize $|k\rangle$ to unity. The delta function normalization is the natural one when the eigenvalue spectrum is continuous.)

The attentive reader may have a question at this point.

"Why was it assumed that the eigenvalue k was real? It is clear that the function $A\,e^{ikx}$ with $k = k_1 + ik_2$ also satisfies Eq. (1.10.31)."

The answer is, yes, there are eigenfunctions of K with complex eigenvalues. If, however, our space includes such functions, K must be classified a non-Hermitian operator. (The surface term no longer vanishes since e^{ikx} blows up exponentially as x tends to either $+\infty - \infty$, depending on the sign of the imaginary part k_2.) In restricting ourselves to real k we have restricted ourselves to what we will call the *physical Hilbert space*, which is of interest in quantum mechanics. This space is defined as the space of functions that can be either normalized to unity or to the Dirac delta function and plays a central role in quantum mechanics. (We use the qualifier "physical" to distinguish it from the Hilbert space as defined by mathematicians, which contain only *proper vectors*, i.e., vectors normalizable to unity. The role of the *improper vectors* in quantum theory will be clear later.)

We will assume that the theorem proved for finite dimensions, namely, that the eigenfunctions of a Hermitian operator form a complete basis, holds in the Hilbert‡ space. (The trouble with infinite-dimensional spaces is that even if you have an infinite number of orthonormal eigenvectors, you can never be sure you have them all, since adding or subtracting a few still leaves you with an infinite number of them.)

Since K is a Hermitian operator, functions that were expanded in the X basis with components $f(x) = \langle x | f \rangle$ must also have an expansion in the K basis. To find the components, we start with a ket $| f \rangle$, and do the following:

$$f(k) = \langle k | f \rangle = \int_{-\infty}^{\infty} \langle k | x \rangle \langle x | f \rangle \, dx = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-ikx} f(x) \, dx \quad (1.10.34)$$

The passage back to the X basis is done as follows:

$$f(x) = \langle x | f \rangle = \int_{-\infty}^{\infty} \langle k | k \rangle \langle k | f \rangle \, dk = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{ikx} f(k) \, dk \quad (1.10.35)$$

Thus the familiar Fourier transform is just the passage from one complete basis $|x\rangle$ to another, $|k\rangle$. Either basis may be used to expand functions that belong to the Hilbert space.

The matrix elements of K are trivial in the K basis:

$$\langle k | K | k' \rangle = k' \langle k | k' \rangle = k' \delta(k - k') \quad (1.10.36)$$

Now, we know where the K basis came from: it was generated by the Hermitian operator K. Which operator is responsible for the orthonormal X basis? Let us call it the operator X. The kets $|x\rangle$ are its eigenvectors with eigenvalue x:

$$X | x \rangle = x | x \rangle \quad (1.10.37)$$

Its matrix elements in the X basis are

$$\langle x' | X | x \rangle = x \delta(x' - x) \quad (1.10.38)$$

To find its action on functions, let us begin with

$$X | f \rangle = | \tilde{f} \rangle$$

and follow the routine:

$$\langle x | X | f \rangle = \int \langle x | X | x' \rangle \langle x' | f \rangle \, dx' = x f(x) = \langle x | \tilde{f} \rangle = \tilde{f}(x)$$

$$\therefore \quad \tilde{f}(x) = x f(x)$$

‡ Hereafter we will omit the qualifier "physical."

Thus the effect of X is to multiply $f(x)$ by x. As in the case of the K operator, one generally suppresses the integral over the common index since it is rendered trivial by the delta function. We can summarize the action of X in Hilbert space as

$$X|f(x)\rangle = |xf(x)\rangle \qquad (1.10.39)$$

where as usual $|xf(x)\rangle$ is the ket corresponding to the function $xf(x)$.

There is a nice reciprocity between the X and K operators which manifests itself if we compute the matrix elements of X in the K basis:

$$\langle k|X|k'\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\, x\, e^{ik'x}\, dx$$

$$= +i\frac{d}{dk}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(k'-k)x}\, dx\right) = i\delta'(k-k')\ddagger$$

Thus if $|g(k)\rangle$ is a ket whose image in the k basis is $g(k)$, then

$$X|g(k)\rangle = \left|\frac{i\, dg(k)}{dk}\right\rangle \qquad (1.10.40)$$

In summary then, in the X basis, X acts as x and K as $-i\, d/dx$ [on the functions $f(x)$], while in the K basis, K acts like k and X like $i\, d/dk$ [on $f(k)$]. Operators with such an interrelationship are said to be *conjugate* to each other.

The conjugate operators X and K do not commute. Their commutator may be calculated as follows. Let us operate X and K in both possible orders on some ket $|f\rangle$ and follow the action in the X basis:

$$X|f\rangle \to xf(x)$$

$$K|f\rangle \to -i\frac{df(x)}{dx}$$

So

$$XK|f\rangle \to -ix\frac{df(x)}{dx}$$

$$KX|f\rangle \to -i\frac{d}{dx}xf(x)$$

Therefore

$$[X,K]|f\rangle \to -ix\frac{df}{dx}+ix\frac{df}{dx}+if=if \to iI|f\rangle$$

$\ddagger$ In the last step we have used the fact that $\delta(k'-k)=\delta(k-k')$.

69

MATHEMATICAL
INTRODUCTION

Since $|f\rangle$ is an arbitrary ket, we now have the desired result:

$$[X, K] = iI \tag{1.10.41}$$

This brings us to the end of our discussion on Hilbert space, except for a final example. Although there are many other operators one can study in this space, we restricted ourselves to X and K since almost all the operators we will need for quantum mechanics are functions of X and $P = \hbar K$, where $\hbar$ is a constant to be defined later.

Example 1.10.1: A Normal Mode Problem in Hilbert Space. Consider a string of length L clamped at its two ends $x = 0$ and L. The displacement $\psi(x, t)$ obeys the differential equation

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{\partial^2 \psi}{\partial x^2} \tag{1.10.42}$$

Given that at $t = 0$ the displacement is $\psi(x, 0)$ and the velocity $\dot{\psi}(x, 0) = 0$, we wish to determine the time evolution of the string.

But for the change in dimensionality, the problem is identical to that of the two coupled masses encountered at the end of Section 1.8 [see Eq. (1.8.26)]. It is recommended that you go over that example once to refresh your memory before proceeding further.

We first identify $\psi(x, t)$ as components of a vector $|\psi(t)\rangle$ in a Hilbert space, the elements of which are in correspondence with possible displacements ψ, i.e., functions that are continuous in the interval $0 \leq x \leq L$ and vanish at the end points. You may verify that these functions do form a vector space.

The analog of the operator Ω in Eq. (1.8.26) is the operator $\partial^2/\partial x^2$. We recognize this to be minus the square of the operator $K \leftrightarrow -i\partial/\partial x$. Since K acts on a space in which $\psi(0) = \psi(L) = 0$, it is Hermitian, and so is K^2. Equation (1.10.42) has the abstract counterpart

$$|\ddot{\psi}(t)\rangle = -K^2 |\psi(t)\rangle \tag{1.10.43}$$

We solve the initial-value problem by following the algorithm developed in Example 1.8.6:

Step (1). Solve the eigenvalue problem of $-K^2$.

Step (2). Construct the propagator $U(t)$ in terms of the eigenvectors and eigenvalues.

Step (3).

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle \tag{1.10.44}$$

The equation to solve is

$$K^2|\psi\rangle = k^2|\psi\rangle \qquad (1.10.45)$$

In the X basis, this becomes

$$-\frac{d^2}{dx^2}\psi_k(x) = k^2\psi_k(x) \qquad (1.10.46)$$

the general solution to which is

$$\psi_k(x) = A\cos kx + B\sin kx \qquad (1.10.47)$$

where A and B are arbitrary. However, not all these solutions lie in the Hilbert space we are considering. We want only those that vanish at $x=0$ and $x=L$. At $x=0$ we find

$$\psi_k(0) = 0 = A \qquad (1.10.48a)$$

while at $x=L$ we find

$$0 = B\sin kL \qquad (1.10.48b)$$

If we do not want a trivial solution ($A=B=0$) we must demand

$$\sin kL = 0, \quad kL = m\pi, \qquad m = 1, 2, 3, \ldots \qquad (1.10.49)$$

We do not consider negative m since it doesn't lead to any further LI solutions [$\sin(-x) = -\sin x$]. The allowed eigenvectors thus form a discrete set labeled by an integer m:

$$\psi_m(x) = \left(\frac{2}{L}\right)^{1/2}\sin\left(\frac{m\pi x}{L}\right) \qquad (1.10.50)$$

where we have chosen $B = (2/L)^{1/2}$ so that

$$\int_0^L \psi_m(x)\psi_{m'}(x)\,dx = \delta_{mm'} \qquad (1.10.51)$$

Let us associate with each solution labeled by the integer m an abstract ket $|m\rangle$:

$$|m\rangle \xrightarrow[X\text{ basis}]{} (2/L)^{1/2}\sin\left(\frac{m\pi x}{L}\right) \qquad (1.10.52)$$

If we project $|\psi(t)\rangle$ on the $|m\rangle$ basis, in which K is diagonal with eigenvalues $(m\pi/L)^2$, the components $\langle m|\psi(t)\rangle$ will obey the decoupled equations

$$\frac{d^2}{dt^2}\langle m|\psi(t)\rangle = -\left(\frac{m^2\pi^2}{L^2}\right)\langle m|\psi(t)\rangle, \qquad m = 1, 2, \ldots \qquad (1.10.53)$$

in analogy with Eq. (1.8.33). These equations may be readily solved (subject to the condition of vanishing initial velocities) as

$$\langle m|\psi(t)\rangle = \langle m|\psi(0)\rangle \cos\left(\frac{m\pi t}{L}\right) \qquad (1.10.54)$$

Consequently

$$|\psi(t)\rangle = \sum_{m=1}^{\infty} |m\rangle\langle m|\psi(t)\rangle$$

$$= \sum_{m=1}^{\infty} |m\rangle\langle m|\psi(0)\rangle \cos\omega_m t, \qquad \omega_m = \frac{m\pi}{L} \qquad (1.10.55)$$

or

$$U(t) = \sum_{m=1}^{\infty} |m\rangle\langle m|\cos\omega_m t, \qquad \omega_m = \frac{m\pi}{L} \qquad (1.10.56)$$

The propagator equation

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle$$

becomes in the $|x\rangle$ basis

$$\langle x|\psi(t)\rangle = \psi(x, t)$$

$$= \langle x|U(t)|\psi(0)\rangle$$

$$= \int_0^\lambda \langle x|U(t)|x'\rangle\langle x'|\psi(0)\rangle \, dx' \qquad (1.10.57)$$

It follows from Eq. (1.10.56) that

$$\langle x|U(t)|x'\rangle = \sum_m \langle x|m\rangle\langle m|x'\rangle \cos\omega_m t$$

$$= \sum_m \left(\frac{2}{L}\right)\sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{m\pi x'}{L}\right)\cos\omega_m t \qquad (1.10.58)$$

Thus, given any $\psi(x', 0)$, we can get $\psi(x, t)$ by performing the integral in Eq. (1.10.57), using $\langle x| U(t)|x'\rangle$ from Eq. (1.10.58). If the propagator language seems too abstract, we can begin with Eq. (1.10.55). Dotting both sides with $\langle x|$, we get

$$\psi(x, t) = \sum_{m=1}^{\infty} \langle x| m\rangle \langle m| \psi(0)\rangle \cos \omega_m t$$

$$= \sum_{m=1}^{\infty} \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{m\pi x}{L}\right) \cos \omega_m t \langle m| \psi(0)\rangle \qquad (1.10.59)$$

Given $|\psi(0)\rangle$, one must then compute

$$\langle m| \psi(0)\rangle = \left(\frac{2}{L}\right)^{1/2} \int_0^L \sin\left(\frac{m\pi x}{L}\right) \psi(x, 0) \, dx$$

Usually we will find that the coefficients $\langle m| \psi(0)\rangle$ fall rapidly with m so that a few leading terms may suffice to get a good approximation. $\qquad\square$

Exercise 1.10.4. A string is displaced as follows at $t = 0$:

$$\psi(x, 0) = \frac{2xh}{L}, \qquad 0 \le x \le \frac{L}{2}$$

$$= \frac{2h}{L}(L - x), \qquad \frac{L}{2} \le x \le L$$

Show that

$$\psi(x, t) = \sum_{m=1}^{\infty} \sin\left(\frac{m\pi x}{L}\right) \cos \omega_m t \cdot \left(\frac{8h}{\pi^2 m^2}\right) \sin\left(\frac{\pi m}{2}\right)$$