CHAPTER 1

Abstract Group Theory

The concept of groups had its origin more than 150 years ago, in the beginning of the nineteenth century. The early development of the theory of groups was due to the famous mathematicians Gauss, Cauchy, Abel, Hamilton, Galois, Sylvester, Cayley, and many others.[1] However, till the advent of modern quantum mechanics in 1925, it did not find much use in physics. The advantages of group theory in physics were soon recognized and the new tool was put to use in the calculations of the atomic structures and spectra by, to name only a few, H.A. Bethe, E.P. Wigner and others. Group theory has now become indispensable in most branches of physics and physical chemistry.

Although a mathematician is generally more interested in the formal development of abstract group theory, a physicist finds the representation theory of groups of direct use in quantum physics and other branches of physics. In this chapter, we shall discuss only those aspects of abstract group theory which will be needed for understanding the representation theory; this will be taken up in Chapter 3 for finite groups and in Chapter 4 for continuous groups.

1.1 What is a Group?

Consider the set I of all integers, $I=\{\ldots, -3, -2, -1, 0, 1, 2, \ldots\}$, and consider the following four properties of this set: (a) The sum of any two elements of the set I is again an integer and hence belongs

[1]Bell (1965).

to the set I. (b) The set contains an element 0, called *zero*, which has the property that for any element $m \in I$, $m+0=0+m=m$. (c) For every element m of I, there exists a unique element n also belonging to I, such that $m+n=n+m=0$; evidently, $n=-m$. (d) If m, n and p are any three elements of I, $m+(n+p)=(m+n)+p$; this means that the law of addition is associative.

Consider another set, the set $U(n)$ of all unitary matrices of order n, where n is a fixed finite positive integer. This set has the following four properties: (a) If U and V are any two unitary matrices of order n, their product UV is again a unitary matrix of order n and hence belongs to the set $U(n)$. (b) The set contains the unit matrix I which has the property $UI=IU=U$ for every $U \in U(n)$. (c) If U is an element of $U(n)$, there exists a unique element V also in $U(n)$ such that $UV=VU=I$. (d) If U, V and W are any three elements of the set, $U(VW)=(UV)W$.

It will be noticed that the four properties satisfied by the above two sets are very much similar in nature. In fact, these properties define a *group* and both the sets discussed above are examples of a group.

Abstractly, a group is a set of *distinct* elements, $G \equiv \{E, A, B, C, D, \ldots\}$, endowed with a law of composition (such as addition, multiplication, matrix multiplication, etc.), such that the following properties are satisfied:

(a) The composition of any two elements A and B of G under the given law results in an element which also belongs to G. Thus,

$$A \circ B \in G, \quad B \circ A \in G, \tag{1.1}$$

where we have denoted the composition of two elements of G by the symbol $\circ$. Symbolically,

$$A \circ B \in G \;\forall\; A, B \in G.$$

This property is known as the *closure* property of the group and the set is said to be closed under the given law of composition.

(b) There exists an identity element $E \in G$ such that for all $A \in G$,

$$E \circ A = A \circ E = A. \tag{1.2}$$

Symbolically,

$$\exists\, E \in G \ni E \circ A = A \circ E = A \;\forall\; A \in G.$$

E is known as the *identity element* of G.

(c) For any element $A \in G$, there exists a unique element $B \in G$ such that

$$A \circ B = B \circ A = E. \tag{1.3}$$

Symbolically,

$$\forall\; A \in G \,\exists\, B \in G \ni A \circ B = B \circ A = E.$$

B is called the *inverse* of A, and vice versa.

(d) The law of composition of the group elements is *associative*, i.e., for any A, B, $C \in G$,

$$A \circ (B \circ C) = (A \circ B) \circ C. \tag{1.4}$$

Symbolically,

$$A \circ (B \circ C) = (A \circ B) \circ C \; \forall \; A, B, C \in G.$$

The number of elements in a group is called its *order*. A group containing a finite number of elements is called a *finite group*; a group containing an infinite number of elements is called an *infinite group*. An infinite group may further be either discrete or continuous: if the number of the elements in a group is denumerably infinite (such as the number of all integers), the group is *discrete*; if the number of the elements in a group is nondenumerably infinite (such as the number of all real numbers), the group is *continuous*.

Some more examples of a group are:

(i) The group of order two consisting of the real numbers 1, —1, with ordinary multiplication as the law of composition.

(ii) The group of order four consisting of the complex numbers $1, i, -1, -i$ (where $i^2 = -1$), under multiplication.

(iii) The discrete infinite group of all real integers discussed above. The law of composition is addition and the identity element is 0.

(iv) The set of all real numbers under addition. This is a continuous group with 0 as the identity element. The inverse of a number b is its negative $-b$.

(v) The set of all positive (zero excluded) real numbers under multiplication. The identity element is 1 and the inverse of x is its reciprocal $1/x$.

(vi) The single point set containing just the unity is a group of order one under multiplication.

(vii) The set of the two matrices $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ under matrix multiplication.

(viii) The set of all nonsingular square matrices of order n (n a positive integer) under matrix multiplication.

(ix) If k is a positive integer, the set $(0, 1, 2, \ldots, k-1)$ of k integers is a group under[2] addition modulo (k). The identity element is zero and the inverse of an element r is $k-r$.

[2]A number n modulo (k) is defined as the remainder obtained on dividing n by k. Thus, 10 modulo (6) $=4$, 3 modulo (3) $=0$, etc. Let $k=6$ in Example (ix); then $3+4=1$, $5+1=0$, etc.

(x) If p is a prime number greater than 1, the set $(1, 2, \ldots, p-1)$ of $p-1$ integers is a group under multiplication[3] modulo (p). The identity element is 1 and the inverse of an element r is $(sp+1)/r$ where s is the smallest positive integer which makes $sp+1$ an integral multiple of r in the ordinary sense.

(xi) The set of all matrices of order $m \times n$ under matrix addition. The identity element is the null matrix of order $m \times n$ and the inverse of an element A is its negative $-A$.

In the above examples, we come across two basic laws of composition—addition and multiplication—each referring to scalars and matrices. When the law of composition of a group is addition, the inverse of an element is called the *additive inverse*; when it is multiplication, the inverse is called the *multiplicative inverse*. Thus, if x is a number, $-x$ is its additive inverse and $1/x$ the multiplicative inverse provided $x \neq 0$. If A is a matrix, $-A$ is its additive inverse and A^{-1} the multiplicative inverse provided A is nonsingular. Similarly, in the case of a group of numbers, 0 is the *additive identity* and 1 the *multiplicative identity*; in the case of a group of matrices, the null matrix (of appropriate order) is the additive identity while the unit matrix (of appropriate order) is the multiplicative identity.

Hereafter, the symbol $\circ$ will be dropped and, for example, AB will be written for $A \circ B$. Similarly, we shall often replace the word 'composition' by 'multiplication' or 'product' of group elements.

The product of the group elements is not necessarily commutative, i.e., in general, $AB \neq BA$. If all the elements of a group commute with each other, it is said to be an *abelian group*. Such groups have important consequences as will be seen later. All the groups considered above, except the group $U(n)$ of all unitary matrices of order n and the group of all nonsingular matrices of order n, are abelian groups.

1.1.1 Group of transformations. The groups of particular interest to a physicist are the groups of transformations[4] of physical systems. A transformation which leaves a physical system invariant is called a *symmetry transformation* of the system. Thus any rotation of a circle about an axis passing through its centre and perpendicular to the plane of the circle is a symmetry transformation for it. A permutation of two identical atoms in a molecule is a symmetry transformation for the molecule.

[3]See footnote 2. In this Example, if $p=7$, then $3.4=5$, $2.5=3$, etc.; the inverse of 4 is 2, since $4.2=1$.

[4]Such as rotations, reflections, permutations, translations, etc.

We shall now show that the set of all symmetry transformations of a system is a group. First we observe that if we perform two symmetry transformations of the system successively, the system remains invariant. Thus the composition of any two symmetry transformations of the system is again a symmetry transformation of the system, i.e., the set considered is closed under the law of successive transformations. We can define an identity transformation which leaves the system unchanged; and this obviously belongs to the set. Given a symmetry transformation, we see that there exists an inverse transformation which also belongs to the set. Finally, the successive transformation of the system obeys the associative law. This proves that the set considered is a group.

The group of all symmetry transformations of a system is called the *group of symmetry* of the system.

1.1.2 The group of symmetry of a square. Suppose we have a square cut out in a piece of cardboard as shown in Fig. (1.1). Let us label the various points of the square as shown in the figure: the corners by a, b, c, d; the centres of the edges by e, f, g, h; and the centre of the square by o. The points marked 1, 2,..., 8 are fixed on the paper (they are not marked on the square). Now suppose we

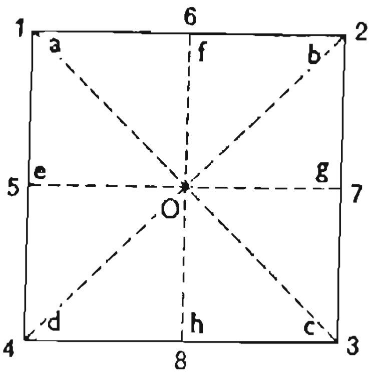


FIGURE 1.1 The axes and the planes of symmetry of a square

rotate the square through a right angle about a line perpendicular to the square and passing through o. But for the labeling a, b,..., h, we would not notice any change in the square. Consider all such symmetry transformations of the square (such as rotating or reflecting it, without bending or stretching) which leave the position of the boundaries of the square unchanged but give a distinct labeling of the marked points a, b, ..., h. Before listing all such transformations, it would be proper to say a few words about the notation we shall be using.

If a rotation through an angle $2\pi/n$ (n a positive integer) about some axis leaves the system invariant, the axis is known as an *n-fold symmetry axis* of the system and the corresponding operation is denoted by C_n. Its integral powers, which will also be symmetry transformations of the system, will be denoted by C_n^k; this represents k successive operations of C_n on the system, or a rotation of $2\pi k/n$ about the axis. A reflection in a plane will be denoted by m or σ with a subscript specifying the plane of reflection. The identity transformation will be denoted by E.

While enumerating all the symmetry transformations of a square, which are listed in Table (1.1), we shall use the shorthand notation 'reflection in a line' to mean 'reflection in a plane perpendicular to the square passing through the line'.

It can be seen that the operations listed in Table (1.1) exhaust the symmetry transformations of a square, i.e., there is no other transformation which leaves the square in the same position and yet gives a distinct labeling for the points $a, b, \ldots, h$. One may think of inversion through the centre o; but it can be readily verified that it is identical to C_4^2.

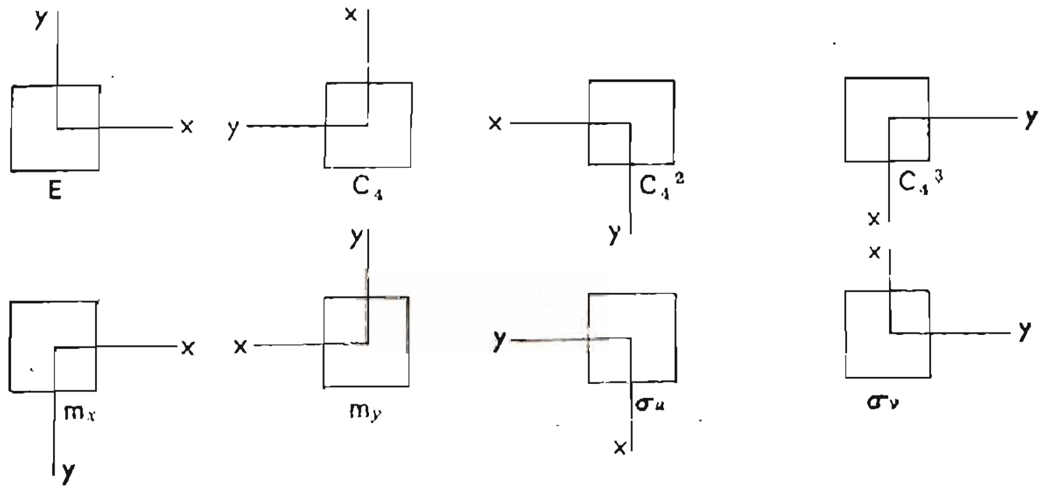
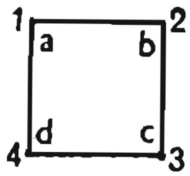
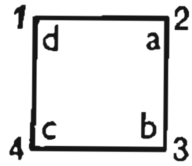
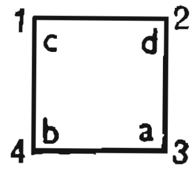
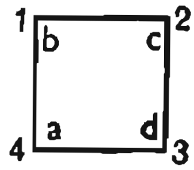
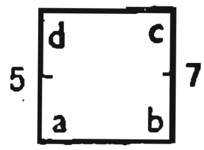
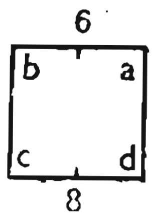
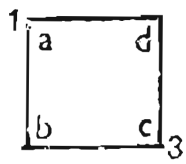
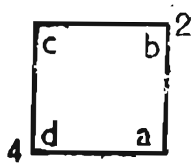


FIGURE 1.2 The equivalance of the transformations of a square with those of a cartesian coordinate system

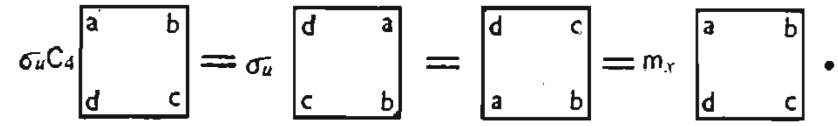
It is interesting to note that these eight transformations correspond to the eight different ways in which we can choose a cartesian coordinate system with axes parallel to the edges of the square. These are shown in Fig. (1.2). We either consider that the coordinate system is held fixed while the square is transformed, which is known as the *active viewpoint*, or that the square is held fixed while the coordinate system is transformed, which is known as the *passive viewpoint*. It should be noted that a transformation in the active viewpoint is

TABLE 1.1 SYMMETRY TRANSFORMATIONS OF A SQUARE

Symbol	*Operation*	*Result*
E	The identity.	
C_4	A clockwise rotation through 90° about an axis normal to the square and passing through o.	
C_4^2	A rotation through 180° about the above axis.	
C_4^3	A clockwise rotation through 270° about the same axis.	
m_x	Reflection in the line 5–7.	
m_y	Reflection in the line 6–8.	
σ_u	Reflection in the line 1–3.	
σ_v	Reflection in the line 2–4.	

equivalent to the inverse transformation in the passive viewpoint. Thus, if in the active viewpoint, we define C_4 as a clockwise rotation of the square, in the passive viewpoint, C_4 would mean an anticlockwise rotation of the coordinate system. This convention will be used throughout this book and is illustrated explicitly in Fig. (1.2).

It can be readily verified that the set of the eight transformations listed in Table (1.1) is a group which is the group of symmetry of a square. Thus, consider the operation of C_4 followed by that of σ_u on the square. This can be found as follows:



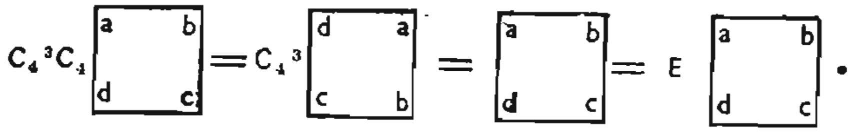
(1.5)

In the operator notation, we can write this as

$$\sigma_u C_4 = m_x, \tag{1.6}$$

meaning thereby that the operations of σ_u C_4 and of m_x on the square or in fact, on any system, give the same result.

The inverse of an operator is that operator which nullifies the effect of the first. Thus, consider the successive operation $C_4^3 C_4$ on the square:



(1.7)

The same result would be obtained if we operate by C_4 and C_4^3 in the reverse order. Thus, by (1.3), C_4 is the inverse of C_4^3 and vice versa. In the operator notation, we may write this as

$$(C_4)^{-1} = C_4^3 \text{ or } C_4 C_4^3 = C_4^3 C_4 = E. \tag{1.8}$$

It is left as an exercise to verify that each of the eight symmetry transformations has an inverse which is just one of these eight transformations.

Finally, the transformations obey the associative law. Hence the set of the symmetry transformations of a square is a group. This symmetry group of a square of order eight is denoted by C_{4v} in crystallography[5].

[5]The crystallographic point groups are dealt with in Chapter 7. If instead of the reflections, we consider rotations through π about the four lines of Fig. (1.1), we have the group D_4 which is also the symmetry group of a square and has the eight elements (E, C_4, C_4^2, C_4^3, C_{57}, C_{68}, C_{13}, C_{24}) where C_{57} denotes a twofold rotation about the line 5–7, etc. See Chapter 7 for more details.

1.2 The Multiplication Table

Let us consider the following operations

$$C_4\, m_x = \sigma_u,\ \sigma_u\, C_4^3 = m_y,$$
$$\sigma_u\, \sigma_v = C_4^2, \text{ and so on.}$$

All such products of the group elements can be represented by a table, known as the *group multiplication table*. It is shown in Table (1.2) for the symmetry group of a square, C_{4v}. Note that in a successive operation such as $ABC\ldots$, the order of operation is *from right to left*. Thus, in the product $C_4 m_x$, m_x is the *first operation* and C_4 the *second operation*. The entry for $C_4 m_x$ would therefore be found in Table (1.2) in the column corresponding to m_x and the row corresponding to C_4.

TABLE 1.2 THE MULTIPLICATION TABLE FOR THE GROUP C_{4v}

SECOND OPERATION	*FIRST OPERATION*							
	E	C_4	C_4^2	C_4^3	m_x	m_y	σ_u	σ_v
E	E	C_4	C_4^2	C_4^3	m_x	m_y	σ_u	σ_v
C_4^3	C_4^3	E	C_4	C_4^2	σ_v	σ_u	m_x	m_y
C_4^2	C_4^2	C_4^3	E	C_4	m_y	m_x	σ_v	σ_u
C_4	C_4	C_4^2	C_4^3	E	σ_u	σ_v	m_y	m_x
m_x	m_x	σ_v	m_y	σ_u	E	C_4^2	C_4^3	C_4
m_y	m_y	σ_u	m_x	σ_v	C_4^2	E	C_4	C_4^3
σ_u	σ_u	m_x	σ_v	m_y	C_4	C_4^3	E	C_4^2
σ_v	σ_v	m_y	σ_u	m_x	C_4^3	C_4	C_4^2	E

The ordering of the rows and the columns in writing down the multiplication table of a group is immaterial. We have chosen a different ordering for the rows and for the columns: the ordering is such that an element in the first column (second operation) is the inverse of the corresponding element in the first row (first operation). If the multiplication table is written in this way, the principal diagonal contains only the identity element E. The advantage of this arrangement will be clear in Section 3.7.

1.2.1 The rearrangement theorem. It will be noticed from the multiplication Table (1.2) that each element of the group occurs once and only once in each column. This is known as the *rearrangement* *theorem.* The arrangement of elements in a row (column) is different from that in every other row (column).

To prove this theorem, we first show that no element can occur more than once in a row or a column. For, suppose an element D occurs twice in a column corresponding to the element A. This means that there exist two elements, say B and C, such that

$$BA=D \text{ and } CA=D.$$

Multiplying from the right by A^{-1}, we get

$$B=DA^{-1},\ C=DA^{-1},$$

showing that $B=C$, which is contrary to the hypothesis that the group elements are distinct. The same line of argument can be used to show that no element can occur more than once in a row.

The second part is now easy to prove: since no element can occur more than once in a row or in a column and since the number of places to be filled in each row or each column is equal to the order of the group, each element must occur once and only once in each row and in each column. This completes the proof.

An important consequence of this theorem is that if f is any function of the group elements, then

$$\sum_{A\in G} f(A)=\sum_{A\in G} f(AB), \tag{1.9}$$

where B is an element of the finite group G and the sum runs over all the group elements.

1.2.2 Generators of a finite group. It is possible to generate all the elements of a group by starting from a certain set of elements which are subject to some relations. Consider the smallest set of elements whose powers and products generate all the elements of the group. The elements of this set are called the *generators of the group.* We shall restrict ourselves here to finite groups only and illustrate by means of two examples.

EXAMPLE 1. We wish to generate a group starting from an element A subject only to the relation $A^n=E$ such that n is the smallest positive integer satisfying this relation.

Since A is an element of the group, all its integral powers must also be in the group. Thus, we generate new elements A^2, $A^3, \ldots,$ of the group and the process stops at $A^n=E$. The higher powers of A do not give us new elements because $A^{n+k}=A^k$. The desired group is thus $(A, A^2, A^3, \ldots, A^{n-1}, A^n=E)$, whose order is n.

EXAMPLE 2. We wish to generate a group from two elements A and B subject only to the relations $A^2=B^3=(AB)^2=E$.

The group must contain the elements E, A, B and B^2, since $A^2=E$ and $B^3=E$. But then it must also contain all the products of A, B and B^2 among themselves. Hence we get two new elements of the group, AB and BA. It can be shown that A and B do not commute, since if they do, then from the relation $(AB)^2=E$, we have

$$E=ABAB=A^2B^2=B^2,$$

which is not true. Therefore AB and BA are distinct elements. We have thus generated the six elements of the group E, A, B, B^2, AB, BA.

It can now be shown that this set is a group, i.e., it is closed under multiplication. Suppose we wish to show that the product $(AB)B=AB^2$ belongs to this set. From the relation $(AB)^2=E$, we have $(AB)^{-1}=AB$ or $B^{-1}A^{-1}=AB$ or $AB=B^{-1}A$ since $A^2=E$. But from $B^3=E$, we have $B^{-1}=B^2$. Hence $AB=B^2A$. Using this, we find that

$$(AB)B=B^2AB=B^2B^2A=BA,$$

which indeed belongs to the set. Similarly, it can be verified that the inverse of each element of the set also belongs to the set. Hence the desired group is (E, A, B, B^2, AB, BA), whose order is six.

The generators of a group are not unique; they can be chosen in a variety of ways. Thus, for example, the group of order six of Example 2 above may be generated by any one of the following sets of generators: (A, B), (A, B^2), (A, AB), (B, AB), etc. See Problem (1.25)

1.3 Conjugate Elements and Classes

Consider a relation such as

$$A^{-1}BA=C, \tag{1.10}$$

where A, B and C are elements of a group. When such a relation exists between two elements B and C, they are said to be *conjugate elements*. The operation is called a *similarity transformation* of B by A. It is clear that

$$ACA^{-1}=B. \tag{1.11}$$

It is not difficult to find such relationships among the elements of the group C_{4v}. Thus,

$$C_4^{-1}m_xC_4=m_y, \tag{1.12}$$

showing that m_x and m_y are conjugate to each other.

It is a simple exercise to show that if B is conjugate to C and B is also conjugate to D, then C and D are conjugate elements; or B, C and D are all conjugate to each other.

It immediately follows that we can split a group into sets such that all the elements of a set are conjugate to each other but no two elements belonging to different sets are conjugate to each other. In fact, such sets of elements are called the *conjugacy classes* or simply the *classes* of a group. The identity element E always constitutes a class by itself in any group, since, for any element A of the group, $A^{-1}EA=E$. It is left as an exercise to show that the classes of C_{4v} are

$$(E), (C_4, C_4^3), (C_4^2), (m_x, m_y), (\sigma_u, \sigma_v). \tag{1.13}$$

In case we are dealing with groups of transformations consisting of rotations, reflections and inversion of a physical system, there are some simple rules which allow the determination of the classes of a group without having to perform explicit calculations for all the elements. These are:

(i) Rotations through angles of different magnitudes must belong to different classes. Thus C_4 and C_4^2 of C_{4v} belong to different classes (see Problem 1.17).

(ii) Rotations through an angle in the clockwise and in the anticlockwise sense about an axis belong to a class if and only if there exists a transformation in the group which reverses the direction of the axis or which changes the sense of a cartesian coordinate system (i.e., takes a right-handed system into a left-handed one or vice versa). Thus, C_4 and C_4^3 of C_{4v} belong to the same class because a reflection (such as m_x or σ_u) changes the sense of the coordinate system.

(iii) Rotations through the same angle about two different axes or reflections in two distinct planes belong to the same class if and only if the two axes or the two planes can be brought into each other by some element of the group. Thus, m_x and m_y belong to the same class since the line 5–7 of Fig. (1.1) can be brought into the line 6–8 by the application of C_4; σ_u and m_x do not belong to the same class since there is no operation in C_{4v} which can bring the line 1–3 into the line 5–7.

These simple criteria are very useful in obtaining the classes of the molecular and the crystallogarphic point groups simply by inspection.

1.3.1 Multiplication of classes. We now define the product of two classes as follows. Let $C_i=(A_1, A_2,\ldots,A_m)$ and $C_j=(B_1, B_2,\ldots, B_n)$ be two classes (same or distinct) of a group containing m and n elements, respectively. We define their product as a set containing all the elements obtained by taking the products of each element of C_i with every element of C_j. We keep each element as many times as it occurs in the product. Thus,

$$C_i C_j=(A_1B_1, A_1B_2,\ldots, A_lB_k,\ldots, A_mB_n). \tag{1.14}$$

We can easily show that the set $C_i C_j$ consists of complete classes. It would be enough to show that if an element A_lB_k belongs to the set $C_i C_j$, then any element conjugate to A_lB_k also belongs to the set. Consider an element conjugate to A_lB_k with respect to some element X of the group G:

$$\begin{aligned} X^{-1}(A_lB_k)X &= (X^{-1}A_lX)\,(X^{-1}B_kX) \\ &= A_rB_s, \text{ say,} \end{aligned} \tag{1.15}$$

where, by the definition of a class, A_r belongs to C_i and B_s belongs to C_j. Hence A_rB_s belongs to the set C_iC_j.

We can then express the product of two classes of a group as a sum of complete classes of the group:

$$C_i C_j=\sum_k a_{ijk}\, C_k, \tag{1.16}$$

where a_{ijk} are nonnegative integers giving the number of times the class C_k is contained in the product $C_i C_j$, and the sum is over all the classes of the group.

1.4 Subgroups

A set H is said to be a *subgroup* of a group G if H is itself a group under the same law of composition as that of G and if all the elements of H are also in G.

As an example, consider the four elements (E, C_4, C_4^2, C_4^3) of C_{4v}. It is easy to see that this set satisfies all the axioms defining a group; hence it is a subgroup of C_{4v}. Some more examples of the subgroups of C_{4v} are (E, C_4^2, m_x, m_y), (E, σ_u), etc.

Every group G has two trivial subgroups—the identity element and the group G itself. A subgroup H of G is called a *proper subgroup* if $H\neq G$, i.e., if G has more elements than H.

If we work out the classes of the two subgroups (E, C_4, C_4^2, C_4^3) and (E, C_4^2, m_x, m_y), we find that in both of these groups every element constitutes a class by itself (see Problem 1.12). The elements C_4 and C_4^3 *do not belong to the same class* in the group (E, C_4, C_4^2, C_4^3) because there is no operation in this group which changes the sense of the coordinate system. Similarly, m_x and m_y do not belong to the same class in the group (E, C_4^2, m_x, m_y) because there is no operation in this group which can take the x axis into the y axis. It is therefore important to note that *elements belonging to a class in a larger group may not belong to a class in a smaller subgroup.*

1.4.1 Cyclic groups. If A is an element of a group G, all integral powers of A such as A^2, A^3, ..., must also be in G. If G is a finite group there must exist a finite positive integer n such that

$$A^n = E, \tag{1.17}$$

the identity element. The smallest positive (nonzero) integer satisfying (1.17) is called the *order of the element A*.

The group $(A, A^2, A^3, \ldots, A^n \equiv E)$, which we have already discussed in Example 1 of Section 1.2.2, has the property that each of its elements is some power of one particular element. Such groups are called *cyclic groups*. *A group generated by a single element is a cyclic group.* Clearly, cyclic groups are abelian, while the converse is not necessarily true.

1.4.2 Cosets. Consider a subgroup $H = (H_1 \equiv E, H_2, \ldots, H_h)$ of order h of a group G which is of order g. Let X be any element of G. Construct all the products such as XE, XH_2, etc., and denote the set of these elements by[6]

$$XH = (XE, XH_2, XH_3, \ldots, XH_h). \tag{1.18}$$

Now there arise two cases—X may be in the subgroup H or X may not be in H. If X is a member of H, the set XH must be identical to the group H by the definition of a group. In the set XH, we only have a rearrangement of the elements of H. We may denote this by writing

$$XH = H \quad \text{if} \quad X \in H. \tag{1.19}$$

On the other hand, if X does not belong to H, it can be shown that no element of the set XH belongs to H. This we do by starting from a contrary assumption. Thus, suppose that XH_i for some value of i $(1 \leq i \leq h)$ belongs to H. Now since H is a group, H_i^{-1} also belongs to H. Hence it follows that $(XH_i)\, H_i^{-1} = X$ is in H,

[6]This is the multiplication of a set by an element. We have previously discussed the product of two sets in Section 1.3.1.

contrary to the hypothesis that X is not a member of H. This proves that H and XH have no common element. We say that H and XH are *disjoint sets* and express it, in the set theoretic notation, by saying that the intersection of H and XH is the null set ϕ:

$$H \cap (XH)=\phi. \tag{1.20}$$

The set XH is called the *left coset* of H in G with respect to X. Similarly, we can define the *right coset* of H in G with respect to X as the set of elements

$$HX=(EX,\ H_2X,\ H_3X,\ldots,H_hX), \tag{1.21}$$

which will also be disjoint to H if X is not in H. All the elements of the left coset and the right coset must of course belong to the bigger group G since X as well as H_i belong to G.

1.4.3 A theorem on subgroups. We are almost half-way through to prove an important theorem: *If a group H of order h is a subgroup of a group G of order g, then g is an integral multiple of h.*

To prove this, let $H=(E, H_2,\ H_3, \ldots, H_h)$ be the subgroup of G. As before, form the left coset of H with respect to an element $X \in G$ which does not belong to H. All the elements XH_i $(1 \leq i \leq h)$ belong to G but none of them is a member of H, as already shown above. Thus, we have h new elements of the group G. We have so far generated the following $2h$ members of G:

$$H \cup XH=(E, H_2,\ H_3, \ldots,\ H_h,\ X,\ XH_2, \ldots, XH_h). \tag{1.22}$$

If this does not exhaust the group G, then pick up an element Y from the remaining elements of G such that Y belongs neither to H nor to XH. Again, forming the left coset YH, we see by the previous argument that all the elements YH must belong to G, but that no element of YH can belong to H. That is, the sets H and YH are disjoint. We now prove that the sets YH and XH are also disjoint. For, if an element YH_i were to be identical to an element, say, XH_j $(1 \leq i, j \leq h)$, then we have

$$YH_i=XH_j,$$

or

$$Y=XH_j\ H_i^{-1} \equiv XH_k, \text{ say,} \tag{1.23}$$

with $1 \leq k \leq h$, showing that Y belongs to XH, contrary to the hypothesis. Thus we have a set of h new elements of G, making altogether the $3h$ elements

$$H \cup XH \cup YH$$
$$=(E, H_2,\ \ldots, H_h,\ X,\ XH_2, \ldots, XH_h,\ Y,\ YH_2, \ldots, YH_h). \tag{1.24}$$

If this still does not exhaust the group G, then we pick up one of the remaining elements of G and continue the process. Every time we generate h *new* elements, they must all belong to G and hence the order of G must be an integral multiple of h.

The integer g/h is called the *index* of the subgroup H in G.

If an element A of a finite group G is of order n, we have seen that the set $(A, A^2, \ldots, A^n \equiv E)$ is a subgroup of G. Hence it follows that *the order of every element of a finite group must be an integral divisor of the order of the group.*

1.4.4 Normal subgroups and factor groups. If the left and the right cosets of a subgroup H with respect to *all* the elements $X \in G$ are the same, then H is called a *normal subgroup* or an *invariant subgroup* of G. This condition can be written as

$$XH = HX,$$

or

$$X^{-1}HX = H \text{ for all } X \in G. \tag{1.25}$$

We can also express this condition alternatively by requiring that every element of XH be equal to some element of HX, or

$$XH_i = H_jX,$$

which gives

$$X^{-1}H_jX = H_i. \tag{1.26}$$

But this is just the conjugation relation between the elements H_i and H_j and shows that if an element H_i belongs to a normal subgroup H of G, then all the elements conjugate to H_i also belong to H. This is often expressed by saying that *a normal subgroup consists of complete classes of the bigger group*. The converse also holds, i.e., if a subgroup H consists of complete classes of G, then H is a normal subgroup of G (see Problem 1.26). This may therefore be taken as an alternative definition of a normal subgroup. For example, (E, C_4^2, m_x, m_y) is a normal subgroup of C_{4v} whereas (E, m_x) is not.

We now introduce another important concept, that of a factor group. We shall illustrate this first by an example and then follow with a general discussion.

Consider a normal subgroup of C_{4v}, say $K_1 = (E, C_4^2)$, and form all its distinct cosets with respect to various elements of C_{4v}. There are four such distinct cosets including K_1:

$$K_1 = (E, C_4^2), \quad K_2 = (C_4, C_4^3),$$
$$K_3 = (m_x, m_y), \quad K_4 = (\sigma_u, \sigma_v). \tag{1.27}$$

We can make this set of cosets a group if we define the product of two cosets as follows: The multiplication of two cosets is a set obtained by multiplying each element of the first coset with every element of the other, *repeated elements being taken only once.*[7] In general, the product of two cosets will depend on the order of multiplication. Thus, we consider

$$K_2K_3=(C_4,\ C_4^{\,3})\ (m_x,\ m_y)$$
$$=(\sigma_u,\ \sigma_v,\ \sigma_u,\ \sigma_v)\rightarrow(\sigma_u,\ \sigma_v)=K_4. \tag{1.28}$$

It can then be seen that the set $K\equiv(K_1,\ K_2,\ K_3, K_4)$ is closed under coset multiplication defined above. Similarly, it can be verified that this set also satisfies all the other requirements for being a group. Hence it follows that the set K, where each coset K_i is considered an 'element' on a higher plane of abstraction, is a group under the given law of composition. This group K is called the *factor group* of G with respect to the normal subgroup K_1.

Quite generally, if H is a normal subgroup of G, the set of all the distinct cosets of H in G, together with the coset multiplication defined above, is called the *factor group* or the *quotient group* of G with respect to H and is denoted by

$$K=G/H. \tag{1.29}$$

If g is the order of G and h that of H, then it is easy to see that the order of K is g/h, the index of H in G.

1.5 Direct Product of Groups

The direct product of two groups $H=(H_1\equiv E,\ H_2,\ H_3,\ldots,\ H_h)$ of order h and $K=(K_1\equiv E,\ K_2,\ K_3,\ldots,\ K_k)$ of order k is defined as a group G of order $g=hk$ consisting of elements obtained by taking the products of each element of H with every element of K, provided (i) that H and K have no common element except the identity E and (ii) that each element of H commutes with every element of K. The direct-product group is denoted by

$$G=H\otimes K=(E,\ EK_2,\ EK_3,\ldots,\ EK_k,\ H_2K_2,\ldots,$$
$$H_2K_k,\ldots,\ H_hK_k). \tag{1.30}$$

Clearly, both H and K are normal subgroups of G. The subgroups of C_{4v} afford a simple example of this concept. Thus,

$$(E,\ m_x)\otimes(E,\ m_y)=(E,\ C_4^{\,2}, m_x,\ m_y). \tag{1.31}$$

Taking the direct product of groups provides the simplest method of enlarging a group. This concept finds its immediate use in the study of symmetry of physical systems such as atoms, molecules,

[7]Note that this is different from the class multiplication defined earlier.

crystals, nuclei and elementary particles. To take an example, suppose G is a group of symmetry (of a system) consisting of proper rotations only. Suppose we later discover that the inversion, J, is also a symmetry transformation of the system. The inversion operator J along with the identity E constitutes a group of order 2, (E, J). Since the inversion commutes with all the rotations, we can take the direct product of G with (E, J) to obtain a bigger symmetry group for the system which is now $G\otimes(E, J)$. Although it is not possible in reality to tell whether we have found all the symmetries of a given system, it is naturally desirable to know as many of them as possible. We shall discuss this concept in more detail when we come to the applications of group theory to quantum mechanics in Chapters 5 and 6.

1.6 Isomorphism and Homomorphism

A group multiplication table, such as that shown in Table (1.2) for the group of a square, characterizes the group completely and contains all the information about the analytical structure of the group. All groups having similar multiplication tables have the same structure—they are said to be *isomorphic* to each other.

Mathematically, there is an *isomorphism* between two groups $G=\{E, A, B, C,\ldots\}$ and $G'=\{E', A', B', C',\ldots\}$, both of the same order g, if there exists a one-to-one correspondence between the elements of G and G'. In other words, if the one-to-one correspondence is denoted by $A\leftrightarrow A'$, $B\leftrightarrow B'$, $C\leftrightarrow C'$, etc., then a multiplication such as $AB=C$ in the group G implies that $A'B'=C'$ in the group G'. The multiplication table of G' can thus be obtained from that of G simply by replacing the elements of G by the corresponding elements of G'. It should be noted that the identity element of one group corresponds to the identity element of the other group under isomorphic mapping.

As an example, it can be seen that the group $\{1, i, -1, -i\}$ of numbers is isomorphic to the group $\{E, C_4, C_4^2, C_4^3\}$ of rotations under the mapping

$$1\leftrightarrow E,\ i\leftrightarrow C_4, -1\leftrightarrow C_4^2, -i\leftrightarrow C_4^3.$$

Thus, for example, the product $(-1)\ (-i)=i$ in one group corresponds to the product $C_4^2\ C_4^3=C_4$ in the other. We shall come across many other examples of isomorphism later.

Very often we come across a many-to-one correspondence or mapping from one group to another (or one set to another, in general). We say that there is a *homomorphism* from a group G_1 to another G_2 if to each element A in G_1 there corresponds a unique element $\phi(A)$ of G_2 such that $\phi(AB)=\phi(A)\,\phi(B)$. The mapping ϕ must be defined for all elements of G_1. The element $\phi(A)$ of G_2 is called the *image* or *map* of the element A of G_1 under the homomorphism. Although each element A of G_1 is mapped onto a unique element $\phi(A)$ of G_2, several elements of G_1 may be mapped onto the same element in G_2. Thus it may happen that $\phi(A)=\phi(B)$ even if $A\neq B$. If n elements of G_1 are mapped onto each element of G_2, we say that there is an n-to-1 mapping or homomorphism from G_1 to G_2. It is evident that if $n=1$, the mapping reduces to isomorphism.

Let us develop a slightly different notation to make the concepts more clear. Let $G=\{E, A, B, C, \ldots\}$ be a group of order g and let $G'=\{E_1, E_2, \ldots, E_n, A_1, A_2, \ldots, A_n, \ldots\}$ be a group of order ng (note that only one element, say E_1, is the identity of G'). Suppose that it is possible to split the group G' into g sets (E_i), (A_i), etc., each containing n elements such that the elements of G' can be mapped onto the elements of G according to the scheme

$$E_1, E_2, \ldots, E_n \to E;$$
$$A_1, A_2, \ldots, A_n \to A; \text{ etc.} \tag{1.32}$$

Then the group G' is said to be *homomorphic* to G if the mapping is such that the product

$$A_i B_j = C_k, \quad 1 \leqslant k \leqslant n, \tag{1.33}$$

in G' implies $AB = C$ in G, and vice versa, where C is the image in G of the elements $C_1, C_2 \ldots, C_n$ of G'. We say that there is an n-to-1 *homomorphism* or mapping from G' to G.

Again the subgroups of C_{4v} provide a simple example of homomorphism. Thus, the group (E, C_4^2, m_x, m_y) is homomorphic to the group (E, m_x) with the following two-to-one mapping:

$$E,\ C_4^2 \to E;\ m_x,\ m_y \to m_x. \tag{1.34}$$

1.6.1 The set (E_i) is a normal subgroup of G'. It can be shown quite generally that the set (E_i) of G', whose elements $E_1, E_2, \ldots, E_n$ are mapped onto the identity element E of G, is a normal subgroup of G'. To prove this, we first show that the set (E_i) is a group. In the group G, we have $EE=E$; therefore, by the definition of homomorphism, the product of any two elements E_i and E_j of G' must belong to the same set (E_i). Thus, the set (E_i) is closed under multiplication. Now we must show that the identity element, which we denote by E' for a moment, belongs to the set (E_i). Suppose E' belongs to some other set of G', say, $E' \in (A_i)$; then for any element $B_k \in G'$, we must have $E'B_k = B_k$. By homomorphism, we must then have $AB=B$ in G, which is possible only if $A=E$, i.e., only if $E' \in (E_i)$. It is now almost trivial to show that if $E_j \in (E_i)$, then E_j^{-1} also belongs to the set (E_i). Thus we have proved that (E_i) is a group.

To prove the second part, that (E_i) is a normal subgroup of G', we consider its left and right cosets with any other element, say $A_l \in G'$, i.e., we consider $A_i(E_j)$ and $(E_j)A_i$. Because $EA=AE=A$ in G, any product element such as E_jA_l or A_iE_j of G' must belong to the set (A_i). Moreover, the products of A_l with all the n elements E_j of the set (E_j) exhaust the set (A_i). To put it briefly, every element of (A_l) must occur once and only once in the product $A_l\,(E_j)$; the same will clearly be true for $(E_j)\,A_i$. Thus, we have

$$\begin{aligned} A_l\,(E_j) &= (A_i), \\ (E_j)\,A_i &= (A_i), \end{aligned} \tag{1.35}$$

showing that (E_i) is a normal subgroup of G'.

The set (E_i) of G' which is mapped onto E of G is called the *kernel of homomorphism*. The above theorem can therefore be stated briefly by saying that *the kernel of homomorphism from G' to G is a normal subgroup of G'.*

The identity element furnishes a trivial example of homomorphism. There is a homomorphism from any group G onto the group of order one containing only the identity element, which, in turn, is a normal subgroup of any group.

1.7 Permutation Groups

These groups are of considerable importance in the quantum mechanics of identical particles. Consider a system of n identical objects. If we interchange any two or more of these objects, the resulting configuration is indistinguishable from the original one. We can consider each interchange as a transformation of the system and then all such possible transformations form a group under which the system is invariant. Since there are altogether $n!$ permutations on n objects, the group has order $n!$. It is known as the *permutation group* of n objects or the *symmetric group* of degree n and is usually denoted by S_n.

Taking a specific example of three identical objects, we see that there are six possible permutations which may be denoted as:

$$E=\begin{pmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{pmatrix},\quad A=\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix},\quad B=\begin{pmatrix}1 & 2 & 3\\ 3 & 1 & 2\end{pmatrix},$$
$$C=\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix},\quad D=\begin{pmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{pmatrix},\quad F=\begin{pmatrix}1 & 2 & 3\\ 1 & 3 & 2\end{pmatrix}. \tag{1.36}$$

The labels 1, 2 and 3 refer to the positions of the three objects rather than to the objects themselves.[8] The system itself has six possible 'states' which may be denoted by

$$\psi_1=(1\ \ 2\ \ 3),\quad \psi_2=(2\ \ 3\ \ 1),\quad \psi_3=(3\ \ 1\ \ 2),$$
$$\psi_4=(2\ \ 1\ \ 3),\quad \psi_5=(3\ \ 2\ \ 1),\quad \psi_6=(1\ \ 3\ \ 2). \tag{1.37}$$

The six operators of (1.36) then act on any of the above six states and their operations are to be interpreted as follows. The operation of A, for example, on any state ψ_i means that the object in position 2 is to be put in position 1, that in position 3 to be put in position 2, and that in position 1 to be brought to position 3. Thus,

$$A\psi_1=\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix}(1\ 2\ 3)=(2\ 3\ 1)=\psi_2; \tag{1.38a}$$

$$C\psi_2=\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix}(2\ 3\ 1)=(3\ 2\ 1)=\psi_5. \tag{1.38b}$$

It can be readily shown that the set of the six permutations of (1.36) is a group. The successive operation of two permutations on a state can be easily worked out. Thus, operating on (1.38b) from the left, say, by A, we find

$$A(C\psi_2)=\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix}(3\ 2\ 1)=(2\ 1\ 3)=\psi_4. \tag{1.39}$$

But we also have

$$F\psi_2=\begin{pmatrix}1 & 2 & 3\\ 1 & 3 & 2\end{pmatrix}(2\ 3\ 1)=(2\ 1\ 3)=\psi_4. \tag{1.40}$$

Thus, we have

$$AC\psi_2=F\psi_2. \tag{1.41}$$

It will be seen that if we start from any other state, the result is the same, i.e.,

$$AC\psi_i=F\psi_i,\ 1\le i\le 6. \tag{1.42}$$

Therefore, in the operator notation, we can write

$$AC=F. \tag{1.43}$$

It is left as an exercise in Problem (1.19) to work out the multiplication table of S_3.

[8]In quantum mechanics it is futile to try to label identical particles!

Coming back to the general case of n identical objects, we see that each permutation of these objects can be expressed as the successive interchange or transposition of two objects taken at a time. We define a *transposition* (mk) on n identical objects as the operation in which the objects in the positions m and k are to be interchanged leaving all the other objects where they are. It can then be verified that the symmetric group S_n of degree n (n finite) can be generated by the $n-1$ transpositions $(12), (13), \ldots, (1n)$.

As an example, a set of generators of S_3 are the two transpositions (12) and (13). All the elements of S_3 can be written as suitable products of these generators. Thus, $B=(13)(12)$, $F=(13)(12)(13)$, $C=(12)$, etc., where, as per the convention, the order of operation is from right to left.

If a permutation consists of an even number of transpositions, it is called an *even permutation*; if it consists of an odd number of transpositions, it is called an *odd permutation*. Thus, the operators E, A and B of (1.36) are even permutations, while C, D and F are odd permutations.

The product of two even or of two odd permutations is an even permutation, whereas the product of an even permutation with an odd permutation is an odd permutation. It then immediately follows that the set of all even permutations among the group S_n is a subgroup.[9] This is known as the *alternating group* of degree n and is usually denoted by A_n. Its order is clearly $n!/2$. Thus, the alternating group of degree 3 is $A_3=(E, A, B)$, where the elements have been defined in (1.36).

Some more discussion of the permutation group and its classes is given in Section 6.1.3.

1.8 Distinct Groups of a Given Order

We have already mentioned that isomorphic groups have identical analytical structures. A number of isomorphic groups may stand for altogether different physical situations, but it is sufficient to study only one of them mathematically. The elements of a number of isomorphic groups may be matrices or permutations or coordinate transformations; it suffices to study a group which is isomorphic to all of these and its elements need not have any

[9]A similar result does not hold for the set of all odd permutations. Why?

'meaning' and may be treated in the abstract sense. Notice that the whole theory is based on the four fundamental group axioms which are quite independent of any particular interpretation given to the group elements. This part of the theory is therefore called *abstract group theory*. We may 'put in' any interpretation for the group elements demanded by the physical situation at hand and 'take out' the corresponding results.

It is therefore desirable to enumerate the distinct (nonisomorphic) groups of a given order n. It is particularly easy to do so for small values of n. We list below the possible structures of groups of orders upto $n=6$.

(i) $n=1$. There is only one distinct structure: a group having only the identity element E.

(ii) $n=2$. Again, there is only one distinct structure: a group (E, A), where, because the group is of order two, A^2 must equal E. Any group of order 2 must be isomorphic to (E, A). Examples are (E, m_x), (E, σ_u), $(1, -1)$, etc.

(iii) $n=3$. This case also has only one structure: a group generated by an element A of order 3, i.e., $(A, A^2, A^3 \equiv E)$.

(iv) $n=4$. This is the lowest order for which there are two non-isomorphic groups. If we denote the group by (E, A, B, C), then the two possible structures are discussed below.

As discussed at the end of Section 1.4.3, the elements A, B and C can be of order 2 or 4. If any one element, say A, is of order 4, it follows that the remaining three elements must be equal to the powers of A and we get the structure

$$A^2=B, \quad A^3=C, \quad A^4=E. \tag{1.44}$$

This gives us the cyclic group of order 4, $(A, A^2, A^3, A^4 \equiv E)$.

In the second case, when no element is of order 4, it follows that all the elements (excluding the identity) are of order 2; hence

$$A^2=B^2=C^2=E. \tag{1.45}$$

The result of Problem (1.11) then shows that the group must be abelian. Now consider the product AB; the two possibilities are $AB=E$ and $AB=C$. But $AB=E$ implies that B is the inverse of A, whereas, from (1.45), we see that A is its own inverse. In other words, $AB=E$ would imply $B=A$; therefore the only possibility is $AB=C$.

The two nonisomorphic structures are then

(a) a cyclic group of order 4, $(A, A^2, A^3, A^4 \equiv E)$;

(b) a noncyclic abelian group of order 4, (E, A, B, C) with the structure $A^2=B^2=C^2=E$, $AB=C$, $BC=A$, $CA=B$. This is the lowest order noncyclic group.

Any group of order 4 must be isomorphic to one of these two groups.

(v) $n=5$. Only one distinct structure is possible in this case: the cyclic group of order 5, $(A, A^2, A^3, A^4, A^5\equiv E)$.

(vi) $n=6$. There are again two distinct (nonisomorphic) groups. We shall prove only a part of this statement to illustrate the argument involved.

Let us denote the group by (E, A, B, C, D, F). As before, we note that the orders of all the elements except E must be 2, 3 or 6. If the order of any one elements is 6, it follow that we have a cyclic group of order 6, $(A, A^2, A^3, A^4, A^5, A^6\equiv E)$. Therefore, to find the second possible structure we exclude this case.

Now we shall show that not all the elements A, B, C, D and F can be of order 2. For if they are, then by Problem (1.11), the group is abelian. Then consider any two elements, say A and B with $A^2=B^2=E$, and let $AB=BA=C$. It is clear that the set (E, A, B, C) of four elements is a subgroup of order 4. But this is not possible, because it violates the fundamental theorem on subgroups that the order of a subgroup must be an integral divisor of the order of the group. Hence we conclude that at least one element is of order 3.

The remaining part of the proof is left to the reader. The two resulting structures are:

(a) a cyclic group $(A, A^2, A^3, A^4, A^5, A^6\equiv E)$;

(b) a noncyclic group (E, A, B, C, D, F) which is also nonabelian and has the structure $A^3=B^3=E$, $C^2=D^2=F^2=E$, $B=A^2$, $AC=F$, $CA=D$, $BC=D$, etc. This is the lowest order nonabelian group and is isomorphic to S_3.

It is not easy, although possible in principle, to go on in this way to higher values of n. The number of nonisomorphic groups would increase, in general, with increasing n. However, two comments are worthy of note:

(i) For every finite value of n, there is always a cyclic group generated by an element of order n, i. e., $(A, A^2, A^3, \ldots, A^n\equiv E)$.

(ii) If the order n of a group is a prime number, there is only one possible structure, i. e., the cyclic group of order n.

We conclude this chapter with one solved example.

EXAMPLE. Prove that a set of a group G is a system of generators of

G *if and only if* no proper subgroup of G exists which contains all the elements of the set S.

Choose a subset of the group G such that S is a system of generators of G. To begin with, let us assume that there exists a proper subgroup H of G such that $S \subseteq H \subseteq G$. Since H is a group and S is contained in H, the powers and products of the elements of S give elements belonging to the group H alone, not G, which contradicts the assumption that S is a system of generators of G. Hence, *if S is a system of generators of G, there exists no proper subgroup of G which contains S.*

Now, assume that there exists no proper subgroup of G which contains S. Let us generate a group by taking all powers and products of the elements of S. Suppose this gives rise to the group K; evidently, $K \subseteq G$. But, by assumption, G contains no proper subgroup which contains S. Hence it follows that $K=G$, showing that S is a system of generators of G. Thus *if no proper subgroup of G exists which contains S, then S is a system of generators of G.*

The desired result follows immediately on combining the above two results.

PROBLEMS ON CHAPTER 1

(1.1) Show that the following sets are groups under the given laws of composition and classify them according to their properties:

(i) the set of all rational numbers[10] under addition;

(ii) the set of all nonzero rational numbers under scalar multiplication;

(iii) the set of all complex numbers under addition;

(iv) the set of all nonzero complex numbers under scalar multiplication;

(v) the set of the eight matrices

$$\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix}, \begin{bmatrix}0&1\\-1&0\end{bmatrix}, \begin{bmatrix}-1&0\\0&-1\end{bmatrix}, \begin{bmatrix}0&-1\\1&0\end{bmatrix}, \begin{bmatrix}1&0\\0&-1\end{bmatrix}, \begin{bmatrix}-1&0\\0&1\end{bmatrix}, \begin{bmatrix}0&-1\\-1&0\end{bmatrix}, \begin{bmatrix}0&1\\1&0\end{bmatrix}\right\}$$

under matrix multiplication;

(vi) the set of all unitary matrices of order n under matrix multiplication;

(vii) the set of all even integers under addition;

(viii) the set of all complex numbers of unit magnitude under scalar multiplication.

(1.2) Show that the following sets are not groups under the given laws of composition. Which of the group axioms do they fail to satisfy?

(i) The set of *all* real numbers under multiplication;

[10]A *rational number* is one which can be expressed as the ratio of two integers, p/q. A real number which cannot be expressed as the ratio of two integers (such as $\sqrt{2}$) is called an *irrational number*.

(ii) the set of all nonnegative real numbers under addition;
(iii) the set of all odd integers under (a) multiplication, (b) addition;
(iv) the set $(1, 2, \ldots, p-1)$ of $p-1$ integers under multiplication modulo (p) where p is not a prime number.

(1.3) (a) Do the three matrices

$$E=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad A=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad B=\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

form a group (under matrix multiplication)? Add a minimum number of matrices to this set to make it a group. Find these necessary additional matrices and write down the multiplication table and classes. Is this group isomorphic to (E, C_4, C_4^2, C_4^3) or to (E, C_4^2, m_x, m_y) or to both?

(b) To the group obtained in the above problem, one more matrix is added:

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

Again, add to this set of matrices a minimum number of matrices to make it a group. Show that the resulting group has order eight and that it is isomorphic to C_{4v}. (This fact will be used in Section (3.9))

(1.4) Show that the n n-th roots of unity, i.e., $\exp(i2\pi k/n)$ for $1 \leq k \leq n$, form a cyclic group of order n under scalar multiplication. Show that if m is an integral divisor of n, then the said group has a subgroup of order m.

(1.5) Construct the group multiplication tables for the groups of Example (ix) of Section 1.1 for $k=4$ and 5, and for those of Example (x) for $p=5$ and 7.

(1.6) Write down the multiplication table for the group of the eight matrices of Problem 1.1 (v). Obtain the classes and all the subgroups. Which of them are normal subgroups? Show that this group is isomorphic to the group C_{4v} treated in this chapter by finding a suitable one-to-one correspondence.

(1.7) Generate the matrix group two of whose elements are

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

Show that the group is of order 8 and has 5 classes, but is not isomorphic to C_{4v}. (Hint: Show that the matrix group generated here has six elements of order 4 whereas C_{4v} has only two such elements. The multiplication tables can therefore not be identical.) (This shows that *two groups of the same order having the same number of classes are not necessarily isomorphic.*)

(1.8) Obtain the products of the various classes of the group C_{4v} and express them as sums of classes in accordance with Eq. (1.16).

(1.9) Generate a group from two elements A and B subject only to the relations $A^2=B^k=(AB)^2=E$, where k is a finite integer greater than 1, and find out its order. (Such groups are known as the *dihedral groups* and are denoted by D_k.)

(1.10) What are the generators of the groups C_{4v} and S_3? What are the generators of the matrix group of Problem 1.1 (v)?

(1.11) Show that a group in which each element except the identity is of order 2 is abelian.

(1.12) Show that an element of a group G constitutes a class by itself *if and only if* it commutes with all the elements of G. Hence show that in an abelian group every element is a class.

(1.13) Let H be a subgroup of a group G and let S be an arbitrary subset of G.

(i) Let $C(S\,;\,H)$ be the set of elements of H each of which commutes with every element of S, i.e.,

$$C(S\,;\,H)=(X \in H \mid XA=AX \;\forall\; A \in S).$$

Show that $C(S\,;\,H)$ is a group. (This group is known as the *centralizer* of S in H.)

(ii) Let $N(S\,;\,H)$ be the set of elements of H such that for all $X \in H$, $X^{-1}SX=S$, i.e.,

$$N(S\,;\,H)=(X \in H \mid X^{-1}SX=S).$$

Show that $N(S\,;\,H)$ is a group. (This group is known as the *normalizer* of S in H).

(1.14) Show that the group generated by two commuting elements A and B such that $A^2=B^3=E$ is cyclic. What is its order?

(1.15) Let H be a subgroup of G and let XH be a coset of H which is disjoint to H. Let Y be an element of G belonging neither to H nor to XH. Show that the set YXH need not be disjoint to *both* H and XH. (Hint: Show that if YXH were disjoint to both H and XH, then in the proof of the theorem in Section 1.4.3, we would arrive at the erroneous result that the integer g/h must be an integral power of 2.)

(1.16) Show that every subgroup of index 2 is a normal subgroup.

(1.17) Show that all the elements belonging to a class of a group have the same order. Show, by giving a contrary example, that the converse is not necessarily true.

(1.18) Let C_i be a class of a group and let C_i^* be the set of elements which are the inverses of those of C_i. Show that C_i^* is also a class. (The class C_i^* is usually called the *inverse of the class* C_i.)

(1.19) Construct the multiplication table of the symmetric group S_3 and obtain its classes.

(1.20) Show that the symmetric group S_n of degree n is homomorphic to the symmetric group S_2 of degree 2.

(1.21) Construct the symmetry group of an equilateral triangle (this group is denoted by C_{3v} in crystallography). Write down its multiplication table, classes, subgroups and normal subgroups. Show that C_{3v} is isomorphic to S_3.

(1.22) Construct the alternating group of degree 4, A_4. Write down its multiplication table and obtain its classes.[11]

(1.23) If $G=H \otimes K$, show that

(i) both H and K are normal subgroups of G;

(ii) the factor group G/H is isomorphic to K;

[11]See Falicov (1967), p. 14.

(iii) G is homomorphic to both H and K;

(iv) the number of classes in G is equal to the product of the numbers of classes in H and K.

(1.24) Show that the group C_{4v} is homomorphic to the group $(1, -1)$ under multiplication. Also show that this 4-to-1 homomorphic mapping can be established in three distinct ways.

(1.25) Given that $A^2=B^3=(AB)^2=E$, generate groups starting from the elements (i) (A, AB), (ii) (B^2, BA). Show that in both the cases, you get the same group as that obtained in Example 2 of Section 1.2.2.

(1.26) If a subgroup H of a bigger group G consists of complete classes of G, show that H is a normal subgroup of G, that is, the left and the right cosets of H with respect to any element of G are the same.

(1.27) Consider the symmetric group S_4 of degree 4 with generators (12), (13) and (14). In the notation of the text, this means that $(12)=\begin{pmatrix}1&2&3&4\\2&1&3&4\end{pmatrix}$, etc.

(a) Express the two permutations

$$A=\begin{pmatrix}1&2&3&4\\4&3&2&1\end{pmatrix} \text{ and } B=\begin{pmatrix}1&2&3&4\\3&1&4&2\end{pmatrix}$$

as products of the generators.

(b) What is the order of each of the two elements A and B? Find the number of transpositions in each of these elements.

(c) Obtain both the products AB and BA of these two elements.

(d) Obtain the inverse of each of the two elements.

(1.28) Find the subgroup of the symmetric group S_4 which leaves the polynomial $x_1x_2+x_3+x_4$ invariant. (Such a group is called the group of the given polynomial.)

(1.29) Find the group of the polynomial $x_1x_2+x_3x_4$ and verify that it contains as a subgroup the group obtained in Problem (1.28).

(1.30) Prove that the group of all positive numbers under multiplication is isomorphic to the group of all real numbers under addition. (Hint: The isomorphic mapping is set up by taking logarithms.)

(1.31) Let G denote a cyclic group of order 12 generated by an element A and let H be a subgroup generated by the element A^3. Find all the cosets of H in G and obtain the multiplication table for the factor group G/H.

(1.32) Consider the set of the following six functions:

$$f_1(x)=x, \quad f_2(x)=1-x, \quad f_3(x)=x/(x-1),$$
$$f_4(x)=1/x, \; f_5(x)=1/(1-x), \; f_6(x)=(x-1)/x.$$

Let the operation of composition of two functions be defined as the substitution of a function into another (that is, 'function of a function'). Thus for example,

$$(f_5f_3)(x)=f_5(f_3(x))=f_5(x/(x-1))=1/(1-x/(x-1))$$
$$=1-x\equiv f_2(x),$$

so that $f_5f_3=f_2$, etc. Show that the set is a group under this law of composition. Show that

$$(f_5)^{-1}=f_6, \text{ and } (f_i)^{-1}=f_i \text{ for } i=2, 3, 4.$$

Finally, show that the group is isomorphic to S_3 or C_{3v}.

(1.33) Determine the symmetry groups of a regular pentagon and a regular hexagon. Also find their classes.

Bibliography for Chapter 1

Albert (1956), Chapter 1; Alexandroff (1959); Dixon (1967); Falicov (1967); Hall (1968); Hamermesh (1964); Jansen and Boon (1967); Margenau and Murphy (1966), Chapter 15; Meijèr and Bauer (1962); Schenkman (1965); Tinkham (1964); Wigner (1959).

CHAPTER 2

Hilbert Spaces and Operators

It is an axiom of quantum mechanics that to every physical observable, there corresponds a hermitian operator and that the set of all eigenfunctions of a hermitian operator is a complete set. The Hilbert space of the operator is the set of all linear combinations of the eigenfunctions. Each state of the system is represented by a vector of the Hilbert space on which the operator acts. We then proceed to expand 'any' function as a linear combination of all the eigenfunctions. Sometimes this can be dangerous and misleading unless we know that the function under consideration belongs to the Hilbert space and the conditions under which such an expansion is possible. In this chapter, we shall develop the concepts of Hilbert spaces and operators and prepare the ground for the applications of group theory in quantum mechanics. In most respects, this chapter is independent of the first one. None the less, these two chapters will form the basis of all the remaining chapters.

2.1 Vector Spaces and Hilbert Spaces

In this section, we shall introduce the idea of Hilbert spaces. Some of their important properties will be described in the next section. We are very familiar with the ordinary three-dimensional vector algebra. To a mathematician, however, the familiar three-dimensional space is just a particular example of the generalized concept of a vector space of arbitrary dimensions. This purely abstract concept of n-dimensional spaces (n a finite real positive integer or infinite) indeed becomes essential in many problems in modern physics and mathematics.

Before we begin, it will not be out of place to define in brief a field. Let F be a set of elements $(a, b, c, d, \ldots)$ and suppose that two binary operations are defined for the elements of F: an operation denoted by $+$ (called *addition*) and an operation denoted by . (called *multiplication*). Then F is a *field* if

(i) F is an abelian group under addition, with an identity element denoted by 0 and called *zero*, and

(ii) the set of the nonzero elements of F also is an abelian group under multiplication. The identity element of this group is denoted by 1 and is called the *unity*.

We shall quote only three examples of a field to which we shall frequently refer:

(a) The set of all real numbers, commonly denoted by R;

(b) The set of all complex numbers, commonly denoted by C;

(c) The set of all rational numbers, commonly denoted by Q.

Loosely speaking, the fields are the number systems of mathematics. An example of a finite field is given in Problem 2.12.

The elements of a field are called *scalars*.

We shall now define a vector space and the subsequent subsections will be steps towards defining a Hilbert space.

2.1.1 Vector space. A set L of elements $u, v, w, \ldots$ is called a *vector space*[1] over a field F if the following two conditions are fulfilled:

(a) An operation of addition is defined in L, which we denote by $+$, such that L is an abelian group under addition. The identity element of this group will be denoted by $\mathbf{0}$.

(b) Any scalar of the field F and any element of L can be combined by an operation called *scalar multiplication* to give an element of L such that for every $u, v \in L$ and $a, b \in F$, we have

$$\begin{aligned} a(u+v)&=au+av\in L,\\ (a+b)u&=au+bu\in L,\\ a(bu)&=(a.b)u,\\ 1u&=u,\ 0u=\mathbf{0}. \end{aligned} \tag{2.1}$$

Note here that 0 is an element of the field F, whereas $\mathbf{0}$ is the 'null' element of L.

[1]The names *vector space*, *linear vector space* and *linear space* are all synonimous.

The elements of a vector space are called *vectors*. The 'multiplication' of two elements of a vector space is not necessarily defined.[2]

Henceforth, we shall not distinguish between the two zeros 0 and **0**.

Examples of a vector space are:

(i) The familiar three-dimensional space of position vectors over the field of real numbers. In the sophisticated mathematical language, this should now be described as 'the set of all position vectors together with the operations of ordinary vector addition and multiplication of a scalar by a vector'.

(ii) The set of all n-tuplets of numbers such as $u \equiv (u_1, u_2, u_3, \ldots, u_n)$ over a field to which the scalars u_i belong. Thus, the set of all n-tuplets of complex numbers is a vector space over C; the set of all n-tuplets of real numbers is a vector space over R; the set of all n-tuplets of rational numbers is a vector space over Q. Two elements u and $w \equiv (w_1, w_2, \ldots, w_n)$ of this set are said to be equal if and only if $u_i = w_i$ for all $1 \leq i \leq n$. We denote this by writing $u = w$. The addition of two vectors u and $v \equiv (v_1, v_2, \ldots, v_n)$ of this space and scalar multiplication are defined by

$$\begin{aligned}(u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) &= (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n), \\ c(u_1, u_2, \ldots, u_n) &= (cu_1, cu_2, \ldots, cu_n).\end{aligned} \tag{2.2}$$

Moreover, if $u_i = 0$ for $1 \leq i \leq n$, we say that $u = 0$.

Example (i) above is clearly a special case of the example at hand—it is the set of all triplets of real numbers.

(iii) The set of all real numbers.

(iv) The set of all complex numbers.

(v) The set of all rational numbers.

In the last three examples above the scalars and the vectors are the same. If a vector space is defined over the field of real numbers, it is called a *real vector space*; a vector space defined over the field of complex numbers is called a *complex vector space*.

2.1.2 Inner product space. A vector space L defined over a field F, where F refers to the field of complex numbers or of real numbers, is further called an *inner product space* if its elements satisfy one more condition:

(c) With every pair of elements $u, v \in L$, there is associated a unique number belonging to the field F—denoted by (u, v) and

[2] If the composition of two elements of a vector space is defined and also belongs to the space (with a few more conditions on the product), we have an *algebra*.

called the *inner product* or the *scalar product* of u and v—for which the following properties hold.

$$\begin{aligned} (u, v) &= (v, u)^*, \\ (au, bv) &= a^*b\,(u, v), \\ (w, au+bv) &= a(w, u)+b(w, v), \end{aligned} \tag{2.3}$$

where the asterisk denotes the complex conjugate.

The linear space of all n-tuplets of complex numbers becomes an inner product space if we define the scalar product of two elements u and v as the complex number given by

$$(u, v) = \sum_{i=1}^{n} u_i^* v_i. \tag{2.4}$$

The ordinary three-dimensional space of position vectors is also an inner product space with the familiar rule for taking the scalar product of two vectors. The vector spaces mentioned as examples after (2.2) are all, in fact, inner product spaces with suitable rules for taking the inner product.

Taking the inner product of an element with itself, we find, from (2.4)

$$(u, u) = \sum_{i=1}^{n} | u_i |^2, \tag{2.5}$$

where $|\ |$ denotes the absolute magnitude of the number enclosed. We introduce the notation

$$||u||^2 \equiv (u, u). \tag{2.6}$$

and the nonnegative square root of this real number, denoted by $||u||$, is called the *norm* of the vector u. Clearly, in the familiar language, this corresponds to the *length* of a vector. It is easy to see that the norm has the following properties:

(i) $||u|| \geqslant 0$, and $||u||=0$ if and only if $u=0$;

(ii) $||u+v|| \leq ||u||+||v||$; this is the usual triangular inequality;

(iii) $||au||=|a|\,||u||$.

Before we go a step further and define a Hilbert space, we must consider what a Cauchy sequence is.

2.1.3 Cauchy sequence. If with each positive integer n we can associate a number c_n (in general, complex), then these numbers $c_1, c_2, c_3, \ldots, c_n, \ldots$ are said to form an infinite sequence or, simply, a *sequence*.

A sequence $c_1, c_2, \ldots, c_n, \ldots$ is said *to converge to* a number c, or *to be convergent with* the limit c, if for every real positive number

ϵ, however small, there exists a positive (finite) integer N such that for every integer $n>N$,

$$|c_n-c|<\epsilon. \tag{2.7}$$

The number c is called the *limit* of the sequence.

A sequence c_1, c_2,... is said to be a *Cauchy sequence* if for every real positive number ϵ, however small, we can find a finite positive integer N such that for any two integers $n>N$ and $m>N$,

$$|c_n-c_m|<\epsilon. \tag{2.8}$$

Examples of convergent, and therefore Cauchy, sequences are:

(i) the sequence of the real numbers whose terms are $c_n=2+5/n$, i.e.,

7, 9/2, 11/3, 13/4, 3, 17/6,..., $(2n+5)/n$,..., with the limit $c=2$;

(ii) 1, 1/2, 1/3,..., $1/n$,..., with the limit $c=0$;

(iii) 1.9, 1.99, 1.999, 1.9999,..., with the limit 2.0;

(iv) the sequence of the complex numbers whose terms are $c_n=(5n+3)/4n+i\,(2n-8)/3n$ with the limit $c=5/4+i\,2/3$.

The following sequences are divergent:

(i) the sequence of numbers whose terms are $c_n=p^n$ for $p>1$,

(ii) the sequence of positive integers, 1, 2, 3, 4,..., n,... .

Although, in the above discussion, we have defined a sequence with reference to numbers (real or complex), it should be clear that we can easily extend the idea to sequences of arbitrary entities provided they are all of the same nature. Thus, we may speak of a sequence of vectors in a two- or a three-dimensional space, a sequence of n-tuplets in their vector space, etc. Of course, in each case we must suitably interpret the quantities $|c_n-c|$ and $|c_n-c_m|$ while studying their convergence. This will be illustrated with reference to a sequence of n-tuplets because all the other examples follow as special cases of this one.

Consider a sequence of elements in the vector space of all n-tuplets (real or complex) whose terms are denoted by $u^{(1)}, u^{(2)},\ldots, u^{(k)},\ldots,$ where

$$u^{(k)}\equiv(u_1^{(k)},\ u_2^{(k)},\ldots,\ u_n^{(k)}). \tag{2.9}$$

We say that this is a Cauchy sequence if for every positive number ϵ there exists a positive integer N such that for any two integers $k>N$ and $m>N$,

$$|u^{(k)}-u^{(m)}|<\epsilon \tag{2.10}$$

in the sense that

$$|u_i^{(k)}-u_i^{(m)}|<\epsilon \text{ for } 1\leq i\leq n.$$

Similarly, the sequence is said to converge to a limit $u \equiv (u_1, u_2, \ldots, u_n)$ if for every real positive number ϵ, we can find a positive integer N such that for all integers $m > N$,

$$| u^{(m)} - u | < \epsilon \tag{2.11}$$

in the sense that

$$| u_i^{(m)} - u_i | < \epsilon \text{ for } 1 \leq i \leq n.$$

2.1.4 Hilbert space. We are now ready to define a Hilbert space. We shall restrict ourselves to the field of real or complex numbers. Consider an inner product space L. If every Cauchy sequence of elements belonging to L has a limit which also belongs to L, the space L is said to be *complete*. A complete inner product space is called a *Hilbert space*.

Examples of Hilbert spaces, as well as contrary examples, are easy to construct. All the inner product spaces discussed above, except the vector space of all n-tuplets of rational numbers (which includes, as a special case for $n=1$, the set of all rational numbers), are also Hilbert spaces. The space of all rational numbers is not complete because we can construct a Cauchy sequence in this space whose limit is an irrational number, which does not belong to this space. For example, the sequence of the successive approximations to the square root of 2, i.e., 1.414, 1.4142, 1.41421, 1.414213, . . ., is a Cauchy sequence whose limit $\sqrt{2}$ does not belong to the set of rational numbers. A similar argument shows that the set of all n-tuplets of rational numbers is not a Hilbert space.

2.2 Coordinate Geometry and Vector Algebra in a New Notation

In what follows, we shall treat Hilbert spaces in general. We shall denote a Hilbert space of n-dimensions (the dimensionality is defined below) by L_n. Although drawing pictures or diagrams for the sake of understanding an argument should not be encouraged in modern pure physics and mathematics, it may be advisable to take some specific examples with $n=2$ or $n=3$ to make the ideas clear. Some important concepts and properties are enumerated below.

(i) In the ordinary three-dimensional space of position vectors, we need a set of three axes, and any point in this space can then be located by means of three coordinates measured along the three axes. Similarly, in an n-dimensional vector space, we would need a set of n 'independent' vectors $r_1, r_2, \ldots, r_n$ to 'span' the whole space.

Two vectors r_i and r_j of L_n are said to be *linearly independent* of each other if one is not a constant multiple of the other, i.e., it is impossible to find a scalar c such that $r_i = cr_j$. In the familiar language, this means that r_i and r_j are not 'parallel' vectors. In general, m vectors of L_n are said to be *a set of linearly independent vectors* if and only if the equation

$$\sum_{i=1}^{m} a_i r_i = 0 \tag{2.12}$$

is satisfied only when all the scalars $a_i = 0$ for $1 \leq i \leq m$. In other words, the m vectors are linearly independent if it is impossible to construct the null element of the space by a linear combination of the vectors with at least one nonzero coefficient. Or again, the set of m vectors is linearly independent if none of them can be expressed as a linear combination of the remaining $m-1$ vectors. A simple test for the linear independence of a set of vectors is to construct the determinant of their scalar products with each other as

$$\Gamma = \begin{vmatrix} (r_1, r_1) & (r_1, r_2) & \dots (r_1, r_m) \\ (r_2, r_1) & (r_2, r_2) & \dots (r_2, r_m) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ (r_m, r_1) & (r_m, r_2) & \dots (r_m, r_m) \end{vmatrix},$$

known as the Gram determinant. If $\Gamma = 0$, it follows that one of the vectors can be expressed as a linear combination of the remaining $m-1$ vectors, so that the vectors are linearly dependent; if $\Gamma \neq 0$, the vectors are linearly independent.

(ii) In an n-dimensional complete vector space, or Hilbert space, L_n, a set of n linearly independent vectors is called a *complete set in L_n*. If the number of vectors chosen is less than n, they are called an *incomplete set in L_n*; clearly they are not enough to span the full space. On the other hand, if more than n vectors are chosen in L_n, they form an *overcomplete* or *redundant set in L_n*. They cannot all be linearly independent and it is possible to find at least two nonvanishing scalars a_i such that

$$\sum_{i=1}^{m} a_i r_i = 0, \; m > n. \tag{2.13}$$

(iii) The dimension of a vector space is the maximum number of linearly independent vectors in the space or the minimum number of vectors required to span the space. In other words, the dimension is the number of linearly independent vectors which are both *necessary* *and sufficient* to span the full space. Thus, in the ordinary three-dimensional space of position vectors, we can find at most three linearly independent vectors; three is also the minimum number of linearly independent vectors required to span the space.

A set of n linearly independent vectors in an n-dimensional vector space is called a *basis*, and the vectors are called the *basis vectors*. Clearly, the choice of the basis vectors is not unique; they can be chosen in an infinite number of ways.

(iv) Any vector u in L_n can now be expanded in terms of a complete set of basis vectors r_i, i.e.,

$$u=\sum_{i=1}^{n} u_i r_i, \tag{2.14}$$

where u_i is the *component* of u along r_i. We say that the space L_n can be *fully spanned by the basis vectors*. This result holds only if $\{r_i\}$ is a complete set. The scalars u_i are also called the *Fourier coefficients* of u and (2.14) is called the *Fourier expansion of u*.

(v) We choose a unit for the norm of the vectors in the space L_n (in the familiar language, a unit for the 'length' of the vectors). A vector of unit norm is called a *unit vector* or *normalized vector*. Rather than choosing the basis vectors r_i of arbitrary norm, we then choose a basis consisting of the unit vectors $e_1, e_2, \ldots, e_n$ in L_n.

(vi) So far, we have not assumed any relationship among the basis vectors except their linear independence. But now, for the sake of convenience and to make our algebra simpler, we will choose a complete set of orthogonal basis vectors, without loss of generality. In the ordinary three-dimensional space, this means that we choose cartesian coordinate axes rather than oblique ones. If e_i are the orthonormal basis vectors, we have

$$(e_i, e_j)=\delta_{ij}, \tag{2.15}$$

where δ_{ij} is the Kronecker delta given by

$$\delta_{ij}=\begin{cases} 1 & \text{if } i=j, \\ 0 & \text{if } i\neq j. \end{cases} \tag{2.16}$$

(vii) The scalar product of two vectors

$$u=\sum_{i=1}^{n} u_i e_i \quad\text{and}\quad v=\sum_{i=1}^{n} v_i e_i \tag{2.17a}$$

is then easily found to be

$$(u, v)=(v, u)^*=\sum_{i=1}^{n} u_i^* v_i. \tag{2.17b}$$

Also $$\| u \|^2 \equiv (u, u) = \sum_{i=1}^{n} |u_i|^2. \tag{2.17c}$$

(viii) A linear transformation in the space L_n can be defined by an operator T such that T acting on a vector $u \in L_n$ gives a vector v, also belonging to L_n. The operation is denoted by

$$Tu = v. \tag{2.18}$$

When this happens, that is, when $Tu \in L_n$ for all $u \in L_n$, the space L_n is said to be *closed* under the action of T.

Note that this is the active view point of transformations discussed in Section 1.1.2.

If the vector Tu is unique for all $u \in L_n$ and if the inverse transformation is also uniquely defined, T is said to be a *one-to-one mapping of the space L_n onto itself.*

We shall be mainly concerned with transformations which preserve the Euclidean properties of the space L_n, such as the norms of the vectors and the scalar product of two vectors. Rotations, reflections and inversion are obvious examples of such transformations.

(ix) In the passive view point, we can define transformations of the basis vectors e_i (keeping everything else fixed) resulting in a new set of basis vectors e_i' as follows:

$$e_i \to e_i' = Te_i = \sum_{j=1}^{n} e_j T_{ji}, \quad 1 \leq i \leq n, \tag{2.19}$$

where T_{ji} is a scalar denoting the component of e_i along e_j. Transformations which take one orthonormal set of basis vectors into another orthonormal set are called *unitary transformations*; the operators associated with them are called *unitary operators*[3]. It can be seen that this definition amounts to preserving the norms and the scalar products of vectors.

(x) Eq. (2.19) is in fact a set of n linear equations which can be written explicitly as

$$\begin{aligned} T(e_1, e_2, \ldots, e_n) &= (e_1', e_2', \ldots, e_n') \\ &= (e_1, e_2, \ldots, e_n) \begin{bmatrix} T_{11} & T_{12} & \ldots & T_{1n} \\ T_{21} & T_{22} & \ldots & T_{2n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ T_{n1} & T_{n2} & \ldots & T_{nn} \end{bmatrix}. \end{aligned} \tag{2.20}$$

[3] If the vectors of the space L_n are real, i.e., if L_n is defined over the field of real numbers, these reduce to *orthogonal transformations* and *orthogonal operators*, respectively.

The square matrix[4] $[T_{ij}] \equiv T$ of order n on the right hand side is called *a representation of the operator T in the basis (e_i).*

(xi) Consider a vector e_i' of (2.19). If we take its scalar product with any of the original basis vectors, say e_k, we get

$$(e_k, e_i') = (e_k, Te_i) = (e_k, \sum_{j=1}^{n} e_j T_{ji}),$$

or

$$(e_k, Te_i) = T_{ki} \tag{2.21}$$

by using (2.15). We call this *the matrix element of the operator T between the basis vectors e_k and e_i.* It means that if the operator T is applied on e_i, the resulting vector has a projection T_{ki} along the vector e_k.

(xii) The scalar product of any two vectors u and Tv of L_n, where u and v are the vectors of (2.17a), is given by[5]

$$\begin{aligned}(u, Tv) &= (\sum_k u_k e_k, \quad T\sum_i v_i e_i)\\ &= (\sum_k u_k e_k, \quad \sum_{i,j} v_i e_j T_{ji})\\ &= \sum_{i,j,k} u_k^* v_i T_{ji} (e_k, e_j)\\ &= \sum_{i,k} u_k^* v_i T_{ki}.\end{aligned} \tag{2.22}$$

(xiii) Since, by assumption, the transformed basis vectors e_i' are each of unit length and orthogonal to each other, we have

$$(e_i', e_j') = \delta_{ij}. \tag{2.23}$$

It immediately follows that the matrix T has the following properties (see Problem 2.2):

$$\sum_{i=1}^{n} T_{ij}^* T_{ik} = \delta_{jk} \tag{2.24a}$$

$$\sum_{i=1}^{n} T_{ji}^* T_{ki} = \delta_{jk}, \tag{2.24b}$$

$$1 \leq j, k \leq n;$$

$$|\det T| = 1. \tag{2.24c}$$

These are the well-known conditions for a *unitary matrix*. It is

[4] The matrix $T \equiv [T_{ij}]$ should not be confused with the operator T appearing on the left hand side of (2.20). We shall often use the same symbol for an operator and a matrix representing it.

[5] Although u and v are not elements of a complete set of basis vectors and there is no apparent matrix for T here, (u, Tv) is called the 'matrix element' of T between u and v in quantum mechanics.

often said that all the rows (columns) of a unitary matrix are orthogonal to each other and normalized, which is just what Eqs. (2.24) tell. In the matrix notation, (2.24) can be written concisely as

$$T^{\dagger}=T^{-1} \quad \text{or} \quad TT^{\dagger}=T^{\dagger}T=E, \tag{2.25}$$

where E is the unit matrix of order n and $T^{\dagger}$ denotes the hermitian conjugate of T.

(xiv) The scalar product of two vectors in L_n is invariant under a unitary transformation: Let u and v be any two vectors of L_n and T be a unitary operator, then

$$(Tu,\ Tv)=(u,\ v). \tag{2.26}$$

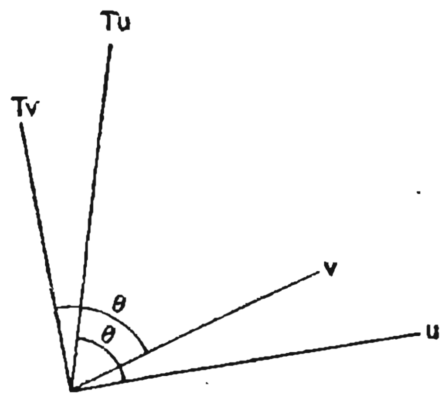


FIGURE 2.1 The scalar product of two vectors is invariant under a unitary transformation

Leaving the proof of (2.26) to Problem (2.3), we show the simple physical interpretation of this result in a two-dimensional space. In Fig (2.1), we have shown the four vectors u, v, Tu and Tv, assuming that T is an anticlockwise rotation through an angle θ about an axis normal to the plane of the paper. The validity of (2.26) for the particular case considered in this figure should be obvious.

(xv) An important operator is the projection operator. This is an operator which, when it operates on a vector $u \in L_n$, gives the projection of u along a given basis vector. It can be written in the form

$$P_i \equiv e_i\ (e_i,\quad), \tag{2.27}$$

where the notation means that the scalar product is to be taken with the vector on which P_i operates. Thus, if u is the vector of (2.17a), then

$$\begin{aligned} P_i u &= e_i\,(e_i,\ u) \\ &= u_i e_i \\ &= \text{the projection of } u \text{ along } e_i. \end{aligned} \tag{2.28}$$

It should be noted that P_i is not a unitary operator.

If we apply the operator P_i once more on the resulting vector $u_i e_i$ of (2.28), clearly, the result is the same vector $u_i e_i$ again, i.e.,

$$P_i (P_i u) = P_i (u_i e_i) = u_i e_i \equiv P_i (u). \tag{2.29}$$

Since this is true for all $u \in L_n$, we can write in the operator notation,

$$P_i^2 = P_i, \tag{2.30}$$

which is an important property of projection operators. In fact, any operator P, acting on a Hilbert space L_n, for which $P^2 = P$, (i.e., $P^2 u = Pu \ \forall \ u \in L_n$) is called a projection operator. It can be readily verified that

$$\sum_{i=1}^{n} P_i = E, \tag{2.31}$$

where E is the identity operator.

(xvi) We now introduce the concept of the direct sum of two or more spaces. Consider a vector space L_n of n dimensions with a coordinate system $(e_1, e_2, \ldots, e_n)$, and a vector space L_m of m dimensions with the basis vectors $(i_1, i_2, \ldots, i_m)$. Provided that the two spaces have no common vector except the null vector, the *direct-sum space* L_t is the vector space defined by the $t = m + n$ basis vectors $(e_1, e_2, \ldots, e_n, i_1, i_2, \ldots, i_m)$. These may be relabeled by the t vectors $(k_1, k_2, \ldots, k_t)$. If L_n and L_m are complete spaces, so is L_t, and any vector u in L_t can be expanded as

$$u = \sum_{i=1}^{t} u_i k_i, \tag{2.32}$$

where u_i are scalars.

As a simple example, consider a two-dimensional vector space (a plane) with the basis vectors (x, y) and a one-dimensional vector space (a line) with the basis vector (z), which does not lie in the plane (x, y). If the null element is common to both the spaces, the direct-sum space is the three-dimensional vector space with the basis vectors (x, y, z).

(xvii) Finally, we consider the direct product (also known as the Kronecker product) of two vector spaces. Consider, again, the two spaces L_n and L_m defined above. The *direct-product space* is a space L_p of dimensions $p = nm$ defined by the p basis vectors $(e_1 i_1, e_1 i_2, \ldots, e_1 i_m, e_2 i_1, \ldots, e_n i_m)$. At the first thought, $e_j i_k$ seems to be a tensor rather than a vector; but it can be seen, without much difficulty, that we can identify it with a vector in the p-dimensional space. If we

make this identification and denote the resulting basis vectors by the new labels $(l_1, l_2, \ldots, l_p)$, then, as before, they form a complete set in L_p if L_n and L_m are complete spaces. Any vector $v \in L_p$ can then be expressed as

$$v = \sum_{j=1}^{p} v_j \, l_j. \tag{2.33}$$

2.3 Function Spaces

Consider the set of all continuous, 'square integrable' functions $f, g, h, \ldots$), each of which is a function of one independent variable x on the interval $[a, b]$. We define the equality of two functions as follows: Two functions f and g are said to be equal on $[a, b]$, denoted by writing $f=g$, if and only if $f(x)=g(x)$ for all values of x on the interval $[a, b]$.

Referring to the definition of vector spaces in Section (2.1.1), we then see that the set of functions considered above is a vector space over a field F if we define the addition of two functions and scalar multiplication by

$$(f+g)(x) = f(x) + g(x), \tag{2.34a}$$

$$(cf)(x) = cf(x). \tag{2.34b}$$

Eq. (2.34a) is called the operation of *pointwise addition* of two functions. If the functions of the set considered are real, we have a vector space over the field of real numbers; if they are complex, we have a vector space over the field of complex numbers. The identity in either case is a function which is identically zero for all values of x on $[a, b]$ and the inverse of a function f is the function $-f$ with the property $(-f)(x) = -f(x)$ (i.e., the value of the function $-f$ at a point x is the negative of the value of f at x).

As a concrete example, consider the set $\{ f_e(x) \}$ of all continuous, square integrable, even, periodic functions of x of period $2l$. We shall allow, in general, complex functions to be included in the set. The sum of two functions of this set is also a continuous, square integrable, even periodic function of period $2l$, and hence belongs to the set. In fact, it is easy to verify that the set is an abelian group under the rule of pointwise addition. Moreover, scalar multiplication by complex numbers as defined in (2.34b) satisfies the conditions (2.1). Hence it follows that the set $\{ f_e(x) \}$ is a vector space, which we shall denote by L_e.

A vector space whose elements are functions is also called a ***function space.***

All the concepts developed in Sections (2.1) and (2.2) can then be applied to function spaces, because, as emphasized in Section 1.8 in connection with groups, the mathematical definition of a vector space is quite independent of the exact nature of its elements. This gives us considerable freedom in handling different vector spaces by the same abstract methods.

Thus, a function space can be made an inner product space if we associate with any two functions a scalar such that the conditions (2.3) are satisfied. This can be easily done if we define the inner product of two functions f and g by

$$(f, g) \equiv \int_a^b f^*(x)\, g(x)\, dx, \tag{2.35}$$

where the integral is over the range $[a, b]$ of x on which the functions of the space are defined. The norm $\|f\|$ of a function f is given by[6]

$$\|f\|^2 \equiv (f, f) = \int_a^b |f(x)|^2\, dx. \tag{2.36}$$

A Cauchy sequence of functions is defined as follows: A sequence $f_1, f_2, \ldots, f_n, \ldots$ of functions of one variable x is said to be a Cauchy sequence *on* $[a, b]$ if for every real positive number ϵ, we can find a positive integer N such that for all integers $n > N$ and $m > N$,

$$\|f_n - f_m\| < \epsilon \tag{2.37}$$

in the sense that

$$\int_a^b |f_n(x) - f_m(x)|^2\, dx < \epsilon.$$

In a similar way (cf. Section 2.1.3), we can define a convergent sequence and its limit. The definition of a Hilbert space of functions follows immediately.

A set of n functions $f_1, f_2, \ldots, f_n$ of a vector space is said to be a set of linearly independent functions *on* $[a, b]$ if and only if the equation

$$\sum_{i=1}^{n} a_i f_i(x) = 0 \tag{2.38}$$

for all x on $[a, b]$ implies that all the scalars $a_i = 0$ for $1 \leq i \leq n$.

Coming back to the vector space L_e of all continuous square integrable even periodic functions of period $2l$, we see that any func-

[6]If the norm of a function is finite, the function is said to be *square integrable.*

tion of this space can be expanded in the well-known Fourier cosine series

$$f(x)=\sum_{n=0}^{\infty} a(n)\,(1/\sqrt{l}\,)\cos(n\pi x/l). \tag{2.39}$$

The infinite set of functions $(1/\sqrt{l})\cos(n\pi x/l)$ for $0\leq n<\infty$ clearly serves as an orthonormal basis in this space, for the functions of this set satisfy the relations

$$\frac{1}{l}\int_{-l}^{l}\cos(n\pi x/l)\cos(m\pi x/l)\,dx=\delta_{mn}. \tag{2.40}$$

Thus the vector space under consideration is denumerably infinite dimensional.

2.3.1 The dual space. For each function f in the space L_e, we have a set of coefficients $a(n)$ for $0\leq n<\infty$ as in (2.39). These can be obtained very easily by Fourier inversion of (2.39), which gives

$$a(n)=\int_{-l}^{l} f(x)\,(1/\sqrt{l})\cos(n\pi x/l)\,dx. \tag{2.41a}$$

These Fourier coefficients are unique, i.e., if we have another function $g\in L_e$ whose Fourier coefficients are

$$b(n)=\int_{-l}^{l} g(x)\,(1/\sqrt{l})\cos(n\pi x/l)dx, \tag{2.41b}$$

then $a(n)=b(n)$ for all $0\leq n<\infty$ if and only if $f=g$ on $[-l, l]$.

Now we may treat a as a function of the discrete variable n. It is easy to see that the function corresponding to $f+g$ would be $a+b$, and that corresponding to $-f$ would be $-a$. In fact, it can be readily verified that the set of functions $(a, b,\ldots)$ is a veetor space which is defined over the same field as the space L_e. This is known as the *dual space* of L_e and its vectors have a one-to-one correspondence with the vectors of L_e. It therefore follows that the dual space is also denumerably infinite dimensional.

It should be clear that this is similar to the space of all n-tuplets where n is now denumerably infinite. The scalar product of two functions in this space is

$$(a,b)=\sum_{n=0}^{\infty} a^*(n)\,b(n). \tag{2.42a}$$

By using Eqs. (2.41) in (2.42a), we find

$$(a, b)=\int_{-l}^{l} f^*(x)g(x)\,dx=(f, g). \tag{2.42b}$$

In the above equation, we have an important property of the Fourier transforms that the scalar product of f and g is the same as that of their transforms a and b.

2.3.2 Direct sum of function spaces. Consider the set $\{f_o(x)\}$ of all continuous square integrable *odd* periodic functions of period $2l$, that is, the set of functions satisfying the relations

$$f_o(x+2l)=f_o(x)$$
$$f_o(-x)=-f_o(x). \tag{2.43}$$

Once again, it can be verified that this set is a vector space[7] which we denote by L_o. Any function $\phi(x)$ of L_o can be expanded in the well-known Fourier sine series

$$\phi(x) = \sum_{n=1}^{\infty} \alpha(n)\,(1/\sqrt{l})\,\sin(n\pi x/l). \tag{2.44}$$

The infinite set of functions $(1/\sqrt{l})\sin(n\pi x/l)$ for $1\leq n<\infty$ can be chosen as the orthonormal basis functions in this space, because

$$\frac{1}{l}\int_{-l}^{l} \sin(n\pi x/l)\,\sin(m\pi x/l)\,dx=\delta_{mn}. \tag{2.45}$$

We can now take the direct sum of the two function spaces L_e and L_o since they have no common element except the function which is identically zero. We then have a space of *all* periodic functions with period $2l$, The Fourier expansion for a function of this space is

$$f(x) = \sum_{n=0}^{\infty} a(n)\;(1/\sqrt{l})\;\cos(n\pi x/l)+\sum_{n=1}^{\infty}\alpha(n)\,(1/\sqrt{l})\,\sin(n\pi x/l). \tag{2.46}$$

The basis functions of this space chosen in (2.46) are clearly orthonormal since, in addition to (2.40) and (2.45), they satisfy

$$\frac{1}{l}\int_{-l}^{l} \cos(n\pi x/l)\,\sin(m\pi x/l)\,dx=0\;\forall\; n,\,m. \tag{2.47}$$

The spaces L_e, L_o and their direct-sum space are all denumerably infinite dimensional. The dual space of L_o is the set all functions $(\alpha, \beta, \ldots)$, each element of which is the Fourier transform of an element of L_o.

It is a fairly easy matter to extend the concepts of this section to functions of more than one variables.

[7] The function which is identically zero for all values of x is even as well as odd in x. It is therefore common to, and is the 'zero' element of, both the spaces $\{f_e(x)\}$ and $\{f_o(x)\}$.

2.4 Operators

In this section, we shall use the symbols $\phi_n(x)$ for the orthonormal basis functions of a Hilbert space L of functions,[8] which may be finite or infinite dimensional.

An operator T is said to be defined on the space L if the action of T on any function $f \in L$ results in a function which also belongs to L. Thus,

$$Tf(x)=g(x) \text{ where } g \in L. \tag{2.48}$$

To know the action of an operator on any function of L, it is enough to know its effect on the basis functions of L. Thus, when an operator T acts on a basis function $\phi_n(x)$, the result is some function of L, say $\phi_n'(x)$, which can be expanded in a linear combination of the original basis functions:

$$T\,\phi_n(x)=\phi_n'(x)=\sum_m \phi_m(x)T_{mn},\ n, m=1, 2, \ldots . \tag{2.49}$$

This represents a system of linear equations, one for each value of n. Written out in an expanded form, this becomes

$$\begin{aligned}(\phi_1', \phi_2', \ldots, \phi_n', \ldots)&=T(\phi_1, \phi_2, \ldots, \phi_n, \ldots)\\ &=(\phi_1, \phi_2, \ldots, \phi_n \ldots)\begin{bmatrix} T_{11} & T_{12} \ldots T_{1n} \ldots \\ T_{21} & T_{22} \ldots T_{2n} \ldots \\ \cdot & \\ \cdot & \\ T_{n1} & T_{n2} \ldots T_{nn} \ldots \\ \cdot & \\ \cdot & \end{bmatrix}\end{aligned} \tag{2.50}$$

The matrix $[T_{ij}]$ is the representation of the operator T with the basis $\{\phi_n\}$. It can be seen in analogy with (2.21) that a matrix element of T is given by

$$\begin{aligned}T_{mn}&=(\phi_m, \phi_n')=(\phi_m, T\phi_n)\\ &=\mathcal{S}\phi_m^*(x)\ T\phi_n(x),\end{aligned} \tag{2.51}$$

where $\mathcal{S}$ denotes summation over the discrete variables and integration over the continuous variables of the set x on which ϕ's depend (see footnote 8).

If we introduce the following notation for row vectors

$$\begin{aligned}\Phi&\equiv(\phi_1, \phi_2, \ldots, \phi_n, \ldots),\\ \Phi'&\equiv(\phi_1', \phi_2', \ldots, \phi_n', \ldots),\end{aligned} \tag{2.52}$$

[8] Here, x stands for the set of variables on which the functions of L may depend.

then (2.49) can be simply written in the matrix notation as

$$\Phi' = \Phi\, T. \tag{2.53}$$

2.4.1 Special operators. We shall consider some special operators in this subsection. An operator T is said to be a *linear operator* if for every f and g in L,

$$T\,(cf+dg) = cTf + dTg, \tag{2.54}$$

where c and d are any scalars of the field over which L is defined. On the other hand, T is called an *antilinear operator* if

$$T\,(cf+dg) = c^*Tf + d^*Tg \;\; \forall\; f, g \in L. \tag{2.55}$$

An obvious example of such an operator is the operator for complex conjugation. If we denote it by K, it is defined by

$$Kf = f^*,\;\; K(cf) = c^*\, Kf = c^*\, f^*. \tag{2.56}$$

If two operators A and B satisfy the relation

$$(f,\, Ag) = (Bf,\, g) \;\; \forall\; f, g \in L, \tag{2.57}$$

A is said to be the *hermitian conjugate* of B, and vice versa, which is expressed by writing

$$A = B^\dagger, \quad A^\dagger = B. \tag{2.58}$$

Let

$$f = \sum_n a_n\phi_n,\;\; g = \sum_n b_n\phi_n. \tag{2.59}$$

Then, on using the orthogonality of ϕ_n, (2.57) becomes

$$\sum_{n,m} a_n{}^* b_m A_{nm} = \sum_{n,m} a_n{}^* b_m B_{mn}{}^*. \tag{2.60}$$

Since this must be true for all f and g in L, i.e., for all scalars a_n and b_n, it follows that

$$A_{nm} = B_{mn}{}^*. \tag{2.61}$$

If the scalars of the space L are real numbers, (2.58) and (2.61) reduce to

$$A = \widetilde{B},\; \widetilde{A} = B,\; A_{nm} = B_{mn}, \tag{2.62}$$

and A is said to be the *transpose* of B, and vice versa.

If an operator T is its own hermitian conjugate (adjoint), it is said to be *hermitian* or *self-adjoint*. From (2.57), we see that T is hermitian if and only if

$$(f,\, Tg) = (Tf,\, g) \;\; \forall\; f, g \in L. \tag{2.63}$$

With (2.59), this reduces to

$$T_{nm} = T_{mn}{}^*. \tag{2.64}$$

This is just the definition of a hermitian matrix—that is, a matrix which equals its own hermitian conjugate—and is written as

$$T=T^{\dagger}=(\widetilde{T})^{*}=\widetilde{T}^{*}. \tag{2.65}$$

Thus a hermitian operator is represented by a hermitian matrix in a linear vector space.

T is said to be a unitary operator if

$$TT^{\dagger}=T^{\dagger}T=E, \tag{2.66}$$

where E is the identity operator. It can be readily seen that if T is unitary, then

$$(Tf,\ Tg)=(f,\ g)\ \forall\ f,\ g\in L. \tag{2.67}$$

If the scalars of the space are real numbers, (2.66) reduces to

$$T\widetilde{T}=\widetilde{T}T=E, \tag{2.68}$$

in which case T is said to be an *orthogonal operator.*

2.4.2 The eigenvalue problem. We have already discussed the operation of an operator T on a basis function, which is

$$T\phi_n=\sum_m \phi_m T_{mn}. \tag{2.49}$$

The choice of the set of basis functions $\{\phi_n\}$ is not unique, and, as such, we would like to choose that set of orthonormal basis functions $\{\psi_n\}$ in L which simplifies Eq. (2.49) as much as possible. Clearly, the simplest nontrivial case arises when the only nonvanishing term on the right-hand side is the n-th term, in which case we have

$$T\psi_n=T_{nn}\psi_n\equiv t_n\psi_n, \tag{2.69}$$

which defines the scalars t_n. A nonzero vector ψ_n satisfying (2.69) is called an *eigenvector* or an *eigenfunction* of T corresponding to the *eigenvalue* t_n. The problem of obtaining the eigenvalues and the eigenfunctions of an operator (acting on a Hilbert space) is usually referred to as the *eigenvalue problem*, and (2.69) is often called the *eigenvalue equation*.

The eigenvalues need not all be distinct, that is, two or more eigenvectors may correspond to the same eigenvalue; in this case, such eigenvectors are said to be *degenerate*. The *multiplicity* of an eigenvalue is defined as the number of linearly independent eigenvectors which have the same eigenvalue under consideration.

It is proper to ask whether each operator has eigenvalues and eigenvectors. If the vector space L is defined over the field of real numbers, every operator acting on L does not necessarily possess

eigenvalues and eigenvectors. Thus, consider the operation of **a** rotation **through 90°** on **a two-dimensional vector** space of (real) position vectors. This operator has no eigenvectors since there is no nonzero vector in this space which transforms into a real multiple of itself.

However, if L is a vector space over the field of complex numbers, every operator on L has eigenvectors. If we count each eigenvalue as many times as it occurs, then the number of eigenvalues is precisely equal to the dimension of the space L.

The set of the eigenvalues of an operator is called its *spectrum*.

2.4.3 Diagonalization. We see from (2.69) that if we choose the set $\{\psi_n\}$ as the basis in the space L, rather than the original set $\{\phi_n\}$, then the matrix representing the operator T is diagonal, i.e.,

$$T_d = \begin{bmatrix} t_1 & & & & \\ & t_2 & & 0 & \\ & & \cdot & & \\ & 0 & & \cdot & t_n \\ & & & & \cdot \\ & & & & & \cdot \end{bmatrix} \tag{2.70}$$

The eigenvalues t_n are the solutions of the N-th order equation

$$\det(T - tE) = 0. \tag{2.71}$$

As we have said, N may be infinite, as is indeed the case in most physical problems. We are then faced with the problem of solving an infinite determinant. However, we are usually interested only in a few lowest eigenvalues in the spectrum of the operator and we can suitably reduce the determinant to a new determinant of a finite order N with small error if the subspace is properly chosen.

Once the eigenvalues are determined in this way, the eigenfunctions can be easily obtained. For this, we express an eigenfunction ψ_n corresponding to the eigenvalue t_n as a linear combination of the original basis functions:

$$\psi_n = \sum_{m=1}^{N} \phi_m U_{mn}. \tag{2.72}$$

If both the sets $\{\psi_n\}$ and $\{\phi_n\}$ are orthonormal, U will be a unitary matrix. Let us express ψ_n in the row vector notation as $\psi_n = (U_{1n}, U_{2n}, \ldots, U_{Nn})$. The eigenvalue Eq. (2.69) then becomes

$$T\psi_n \equiv (U_{1n}, U_{2n}, \ldots, U_{Nn}) \begin{bmatrix} T_{11} & T_{21} & \ldots & T_{N1} \\ T_{12} & T_{22} & \ldots & T_{N2} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ T_{1N} & T_{2N} & \ldots & T_{NN} \end{bmatrix}$$

$$= t_n (U_{1n}, U_{2n}, \ldots, U_{Nn}). \tag{2.73a}$$

where we have used (2.69) in the last step. Note that the matrix of transformation which appears in (2.73a) is the transpose of that appearing in (2.49). This is because in (2.49), T acts on the basis vectors ϕ_n (the passive viewpoint), while in (2.73a), it acts on vectors of the space leaving the basis vectors unchanged (the active viewpoint).

Writing the m-th column of (2.73a), we have

$$\sum_{k=1}^{N} U_{kn} T_{mk} = t_n U_{mn}, \tag{2.73b}$$

where $1 \leq n \leq N$. This is a system of N linear equations for the N unknowns U_{mn} ($1 \leq m \leq N$, fixed n). However, these equations are not all independent due to the condition (2.71). If the eigenvalue t_n is k-fold degenerate, it can be shown that the matrix $(T - t_n E)$ has rank $N-k$ and hence only $N-k$ equations from (2.73) are independent. This means that we can determine at most $N-k$ components U_{mn} (fixed n). The general method is then to fix arbitrarily, say, the first k components and to obtain the remaining $N-k$ components in terms of them.[9] Thus there is a considerable arbitrariness which results from the fact that any linear combination of the degenerate eigenfunctions is also an eigenfunction with the same eigenvalue. We may conveniently choose any k orthonormal functions in this k-dimensional subspace of the full space.

Having obtained in this way a set of N orthonormal eigenfunctions, we can show that the representation of T with the basis $\{\psi_n\}$ is a diagonal matrix. We write Eqs. (2.49) and (2.72) in the matrix notation as

$$T\Phi = \Phi[T], \tag{2.74a}$$

$$\Psi = \Phi U, \tag{2.74b}$$

where Φ and Ψ stand for the row vectors

$$\Phi = (\phi_1, \phi_2, \ldots, \phi_N),$$

$$\Psi = (\psi_1, \psi_2, \ldots, \psi_N),$$

[9] Joshi (1984), Section 8; Kreyszig (1972), Section 6.9.

and we have distinguished between the operator T and the matrix $[T]$. From (2.72), it is clear that the n-th column of the matrix U just contains the components of the eigenfunction ψ_n, i.e.,

$$U=\begin{bmatrix} U_{11} & U_{12} & \dots & U_{1n} & \dots & U_{N1} \\ U_{21} & U_{22} & \dots & U_{2n} & \dots & U_{2N} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ U_{N1} & U_{N2} & \dots & U_{Nn} & \dots & U_{NN} \end{bmatrix}. \tag{2.75}$$

Multiplying (2.74a) from the right by U, we get

$$T\,\Phi\, U=\Phi\, U\, U^{-1}\,[T]\, U.$$

or

$$T\Psi=\Psi\;(U^{-1}\,[T]\;U). \tag{2.76}$$

Thus, the matrix $U^{-1}\,[T]\,U$ is the representation of the operator T with the basis $\{\psi_n\}$. Now it can be readily verified that, by the construction of U as in (2.75), we have

$$U^{-1}\,[T]\;U=T_d.$$

This can be seen by taking the (l, n) element of the left-hand side of the above equation, which gives

$$\sum_{m,k}[U^{-1}]_{lm}\;T_{mk}\;U_{kn}=\sum_m[U^{-1}]_{lm}\;U_{mn}\;t_n\;[\text{by (2.73b)}]$$

$$=t_n\;\delta_{ln},$$

which is just the (l, n) element of T_d. Eq. (2.76) then finally gives us

$$T\Psi=\Psi\;T_d. \tag{2.77}$$

which is the desired result. This process is called the *diagonalization* of an operator.[10]

2.4.4 The spectral Theory of operators. We shall restrict ourselves to the case when the Hilbert space of the operator T is finite dimensional. Moreover, we shall consider T to be a hermitian operator or a unitary oparator.[11]

Let L_n be the n-dimensional $(0 < n < \infty)$ Hilbert space of T. We assume that L_n is defined over the field of complex numbers, so that T has exactly n eigenvalues. Let $t_1, t_2, \dots, t_m$ be the *distinct*

[10]See also Joshi (1984), pp. 95-97

[11]The discussion of this subsection is, in fact, valid for a more general class of operators known as normal operators. An operator T is *normal* if it commutes with its own hermitian conjugate, that is, if $TT^{\dagger}=T^{\dagger}T$. Hermitian and unitary operators are clearly normal operators.

eigenvalues of T, so that $m \leq n$. If the eigenvalue t_i is k_i-fold degenerate, there are k_i linearly independent eigenvectors of T in L_n which have the same eigenvalue t_i. These eigenvectors constitute the basis for a k_i-dimensional subspace M_i of L_n; M_i is called the *eigenspace* of T corresponding to the eigenvalue t_i. Any vector of M_i is an eigenvector of T with the eigenvalue t_i.

We thus have the eigenspaces $M_1, M_2, \ldots, M_i, \ldots, M_m$, corresponding to the eigenvalues $t_1, t_2, \ldots, t_i, \ldots, t_m$, respectively. If T is a hermitian or a unitary operator, then these subspaces are pairwise orthogonal;[12] two spaces are said to be *orthogonal* if every vector of one space is orthogonal to every vector of the other. In our case, this is denoted by writing $M_i \perp M_j$ if $i \neq j$.

Any vector $u \in L_n$ can now be expressed uniquely in the form

$$u = u_1 + u_2 + \ldots + u_m, \tag{2.78}$$

where u_i is in M_i. The u_i's are therefore pairwise orthogonal. The operation of T on u then gives

$$\begin{aligned} Tu &= Tu_1 + Tu_2 + \ldots + Tu_m \\ &= t_1 u_1 + t_2 u_2 + \ldots + t_m u_m. \end{aligned} \tag{2.79}$$

This then determines uniquely the action of T on any vector of the Hilbert space L_n. To express the above result in a more convenient form, we define the m projection operators P_i on the eigenspaces M_i, such that the action of P_i on u gives the projection of u on M_i, or

$$P_i u = u_i. \tag{2.80}$$

Eq. (2.79) then becomes

$$Tu = t_1 P_1 u + t_2 P_2 u + \ldots + t_m P_m u \;\forall\; u \in L_n,$$

so that we can write

$$T = t_1 P_1 + t_2 P_2 + \ldots + t_m P_m. \tag{2.81}$$

This expression is known as the *spectral resolution* of T. For every hermitian or unitary operator acting on a finite-dimensional Hilbert space, the spectral resolution exists and is unique.

The concepts developed in this section are closely related to, and find useful applications in, the eigenvalue problem in physics, because in quantum mechanics, we are concerned with the eigenvalues and the eigenfunctions of hermitian operators.

[12]In this subsection, we shall state the important results of the spectral theory without proofs. For proofs, the reader is referred to Simmons (1963).

2.5 Direct Sum and Direct Product of Matrices

We now digress a little in this section and consider two important operations with matrices which are not normally treated in elementary books on matrix algebra. These are the direct sum and the direct product (also known as the *outer product* or the *Kronecker product*) of matrices.

2.5.1 Direct sum of matrices. The *direct sum of two square matrices* $A\equiv[A_{ij}]$ of order m and $B\equiv[B_{ij}]$ of order n is a square matrix C of order $m+n$ defined by

$$C=A\oplus B=\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}=\left[\begin{array}{cccc:cccc} A_{11} & \cdots & A_{1m} & & & & & \\ \cdot & & \cdot & & & \mathbf{0}_1 & & \\ \cdot & & \cdot & & & & & \\ A_{m1} & \cdots & A_{mm} & & & & & \\ \hdashline & & & & B_{11} & \cdots & B_{1n} & \\ & \mathbf{0}_2 & & & \cdot & & \cdot & \\ & & & & \cdot & & \cdot & \\ & & & & B_{n1} & \cdots & B_{nn} & \end{array}\right], \tag{2.82}$$

where $\mathbf{0}_1$ and $\mathbf{0}_2$ are null matrices of order $m\times n$ and $n\times m$, respectively. Here the symbol $\oplus$ stands for the direct sum. This idea can be easily extended to more than two matrices. For example, the direct sum of

$$A=a,\quad B=\begin{bmatrix} b & c \\ d & e \end{bmatrix},\quad \text{and } C=\begin{bmatrix} f & g & h \\ i & j & k \\ l & m & n \end{bmatrix}$$

is a matrix of order six given by

$$D=A\oplus B\oplus C=\left[\begin{array}{c:cc:ccc} a & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & b & c & 0 & 0 & 0 \\ 0 & d & e & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & f & g & h \\ 0 & 0 & 0 & i & j & k \\ 0 & 0 & 0 & l & m & n \end{array}\right]. \tag{2.83}$$

Such a matrix, which has nonvanishing elements in square blocks along the main diagonal and zeros elsewhere, is said to be in the *block-diagonalized form*. It has the important properties:

$$\det D = (\det A)(\det B)(\det C), \tag{2.84a}$$

$$\text{trace } D = \text{trace } A + \text{trace } B + \text{trace } C, \tag{2.84b}$$

$$D^{-1} = A^{-1} \oplus B^{-1} \oplus C^{-1}, \tag{2.84c}$$

which should be clear from (2.83). Also, if A_1 and A_2 are square matrices of the same order, say n, and B_1 and B_2 are square matrices of the same order, say m, then[13]

$$(A_1 \oplus B_1)(A_2 \oplus B_2) = (A_1 A_2) \oplus (B_1 B_2). \tag{2.84d}$$

2.5.2 Direct product of matrices. The *direct product* of two matrices $A \equiv [A_{lm}]$ of order $L \times M$ and $B \equiv [B]_{pq}$ of order $P \times Q$ is a matrix C of order $I \times J$ where $I = LP$ and $J = MQ$. It can be written as

$$C = A \otimes B = \begin{bmatrix} A_{11}B & A_{12}B & \dots & A_{1M}B \\ A_{21}B & A_{22}B & \dots & A_{2M}B \\ \vdots & & & \\ A_{L1}B & A_{L2}B & \dots & A_{LM}B \end{bmatrix}, \tag{2.85}$$

where an 'element' $A_{lm}B$ stands for a matrix of order $P \times Q$ given by

$$A_{lm}B = \begin{bmatrix} A_{lm}B_{11} & A_{lm}B_{12} & \dots & A_{lm}B_{1Q} \\ A_{lm}B_{21} & A_{lm}B_{22} & \dots & A_{lm}B_{2Q} \\ \vdots & & & \\ A_{lm}B_{P1} & A_{lm}B_{P2} & \dots & A_{lm}B_{PQ} \end{bmatrix}. \tag{2.86}$$

To obtain an element of C in terms of the elements of A and B, we use the notation $C \equiv [C_{lp,\,mq}]$ where a row of C is denoted by a dual symbol (lp) and a column of C by a dual symbol (mq), such that

$$C_{lp,\,mq} = A_{lm}\, B_{pq}. \tag{2.87}$$

We may relabel the rows and the columns of C by two new indices i and j $(1 \le i \le I,\ 1 \le j \le J)$ so that

$$C \equiv [C_{ij}] = [C_{lp,\,mq}]. \tag{2.88}$$

This rather complicated notation can be made clear by an example. The direct product of

$$A = \begin{array}{c c} & \begin{array}{ccc} (1) & (2) & (3) \end{array} \\ \begin{array}{c} (1) \\ (2) \end{array} & \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \end{array}, \quad B = \begin{array}{c c} & \begin{array}{cc} (1) & (2) \end{array} \\ \begin{array}{c} (1) \\ (2) \\ (3) \end{array} & \begin{bmatrix} h & r \\ k & s \\ l & t \end{bmatrix} \end{array}$$

[13]For proofs of various results mentioned in this and the following subsections, see Joshi (1984), Section 13.

is the 6×6 matrix

$$C=A\otimes B=\begin{array}{r} \\ (11)\\ (12)\\ (13)\\ (21)\\ (22)\\ (23)\end{array}\begin{array}{c}\begin{array}{cccccc}(11)&(12)&(21)&(22)&(31)&(32)\end{array}\\ \left[\begin{array}{cccccc} ah & ar & bh & br & ch & cr\\ ak & as & bk & bs & ck & cs\\ al & at & bl & bt & cl & ct\\ dh & dr & eh & er & fh & fr\\ dk & ds & ek & es & fk & fs\\ dl & dt & el & et & fl & ft \end{array}\right]\end{array}. \qquad (2.89)$$

Note that *the rows and the columns of the matrix C are labeled by different schemes*. Thus, while the third row of C is labeled as the (13) row, the third column is labeled as the (21) column. An element of C is, for example,

$$C_{21,\,31}=fh=A_{23}B_{11},$$

which is consistent with (2.87). We now relabel the rows and the columns by identifying each dual symbol with one number, separately for the rows and for the columns. We then have the matrix $[C_{ij}]\equiv[C_{lp,\,mq}]$ with $(lp)\rightarrow i$, $(mq)\rightarrow j$ and $1\leq i, j\leq 6$. Thus, in the above example, $C_{21,\,31}\equiv C_{45}$.

In the general case, the identification of the dual symbol with the single running index can be made by letting $i=(l-1)\,P+p$ and $j=(m-1)\,Q+q$; thus,

$$C_{lp,\,mq}\equiv C_{ij}=C_{(l-1)P+p,\,(m-1)\,Q+q}.$$

The concept can once again be extended to the direct product of more than two matrices. There is no restriction on the order of the matrices whose direct product is to be taken.

If A_1, A_2, B_1 and B_2 are any matrices whose dimensions are such that the ordinary matrix products A_1A_2 and B_1B_2 are defined, then the direct product has the important property

$$(A_1\otimes B_1)\,(A_2\otimes B_2)=(A_1A_2)\otimes(B_1B_2). \qquad (2.90a)$$

Further, if F is the direct product of a number of square matrices $A, B, C, \ldots$, that is, $F = A\otimes B\otimes C\otimes\ldots$, then

$$\text{trace } F=(\text{trace } A)\,(\text{trace } B)\,(\text{trace } C)\ldots. \qquad (2.90b)$$

The operation of the direct product of matrices is associative, so that

$$A\otimes(B\otimes C)=(A\otimes B)\otimes C\equiv A\otimes B\otimes C. \qquad (2.91)$$

The operation is also distributive with respect to matrix addition. Thus,

$$A\otimes(C+D)=A\otimes C+A\otimes D. \qquad (2.92)$$

Moreover, from (2.90a), we have

$$(AB)\otimes(AB)\otimes(AB)=(AB)\otimes((A\otimes A)(B\otimes B))$$
$$=(A\otimes A\otimes A)(B\otimes B\otimes B). \tag{2.93}$$

Generalizing the above equation, we have

$$(AB)^{[k]}=(A)^{[k]}(B)^{[k]}, \tag{2.94}$$

where

$$A^{[k]}=A\otimes A\otimes A\otimes \ldots \otimes A \text{ (k times)}. \tag{2.95}$$

Finally, if A and B are square matrices with eigenvalues and eigenvectors λ_i, x_i and μ_j, y_j, respectively, the eigenvalues of $A\otimes B$ are $\lambda_i\mu_j$ and its eigenvectors are $x_i\otimes y_j$. That is, if $Ax_i=\lambda_i x_i$ and $By_j=\mu_j y_j$, then

$$(A\otimes B)(x_i\otimes y_j)=\lambda_i\mu_j(x_i\otimes y_j). \tag{2.96}$$

The proof follows directly from (2.90a).

We shall find these concepts very useful in the next chapter.

PROBLEMS ON CHAPTER 2

(2.1) Show that the following sets are vector spaces. Also indicate how you would choose a basis in each space. What is the dimension of each space? Which is the field over which each vector space is defined?

(i) The set of all vectors denoting the possible velocities of a free particle in classical mechanics.

(ii) The set of all vectors denoting the possible wave vectors of a free particle in classical or quantum mechanics (note that this is usually referred to as the **k**-space).

(iii) The set of all continuous square integrable solutions of an n-th order ordinary linear homogeneous differential equation.

(iv) The set of all continuous square integrable functions which depend on a set of variables.

(v) The set of all real square matrices of order n.

(vi) The set of all complex square matrices of order n.

(2.2) Prove Eq. (2.24).

(2.3) Prove Eq. (2.26). [Hint: Use (2.24).]

(2.4) State whether the following statements are true or false and explain your answer:

(i) If all the vectors of a set are pairwise orthogonal, it necessarily follows that it is an orthogonal set.

(ii) If all the vectors of a set are pairwise independent of each other, it necessarily follows that it is a set of linearly independent vectors.

(2.5) Consider the projection operators P defined in (2.28). Show that $P_iP_j=0$ if $i\neq j$. (This is expressed by saying that the *projection operators are pairwise orthogonal*.)

(2.6) Show that the eigenvalues of a hermitian operator are real and that those of a unitary operator have absolute magnitude equal to unity.

(2.7) Show that the functions $P_0(x)=1$ and $P_1(x)=x$ are orthogonal on the interval $-1\leq x\leq 1$. Find scalars a' and b' such that $P_2(x)=1+a'x+b'x^2$ is orthogonal to both $P_0(x)$ and $P_1(x)$ on the same interval. In this way, generate polynomials $P_n(x)=1+ax+bx^2+\ldots+gx^n$ such that $P_n(x)$ is orthogonal to each $P_m(x)$, $0\leq m\leq n-1$, on the interval $[-1, 1]$. [Note that these are the Legendre polynomials, apart from constant factors.]

(2.8) Obtain the eigenvalues and the eigenvectors of the following matrices:

(i) $\begin{bmatrix} 1/2 & 0 & -3\sqrt{3}/2 \\ 0 & 1 & 0 \\ -3\sqrt{3}/2 & 0 & -5/2 \end{bmatrix}$, (ii) $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$.

(2.9) Obtain the direct sum and the direct product of the following matrices:

(i) $\begin{bmatrix} 2 & 5 & 9 \\ 1 & 4 & 7 \\ 3 & 3 & 3 \end{bmatrix}$ and $\begin{bmatrix} 6 & 4 \\ 2 & 7 \end{bmatrix}$.

(ii) $\begin{bmatrix} 10 & 3 & -5 \\ -9 & 2 & 5 \\ 0 & 5 & -1 \end{bmatrix}$ and $\begin{bmatrix} 3 & 9 & 0 \\ 5 & -7 & 8 \\ 4 & 2 & -2 \end{bmatrix}$.

(2.10) Obtain the direct product of the two matrices:

$\begin{bmatrix} -2 & 3 & 4 \\ 8 & 7 & -6 \end{bmatrix}$ and $\begin{bmatrix} 9 \\ 6 \\ 3 \end{bmatrix}$.

(2.11) In Problem (2.9) verify Eqs. (2.84a), (2.84b), (2.84c) and (2.90b).

(2.12) Let p be a prime number and consider the set of the p integers $(0, 1, 2, \ldots, p-1)$. Show that this set is a field with addition mod (p) and multiplication mod (p) as the two binary operations. (A finite field is called a *Galois field.*)

(2.13) If $T(A)$ is the matrix representing an operator T in the vector space L_a and $T(B)$ that representing T in the vector space L_b, show that the matrix representing T in the vector space $L_a \otimes L_b$ is $T(A)\otimes T(B)$.

Bibliography for Chapter 2

Albert (1956), Chapter 3; Courant and Hilbert (1966), Chapter 1; Halmos (1958); Helmberg (1969); Jackson (1962); Joshi (1984); Margenau and Murphy (1966), Chapter 10; Meijer and Bauer (1962), Chapter 1; Meschkowski (1968); von Neumann (1955); Newing and Cunningham (1967); Schmeidler (1965); Shilov (1965); Simmons (1963), Chapters 10 and 11; Trigg (1964); Van der Waerden (1949).