

## Chapter 2

# Mathematical Methods in Quantum Mechanics

**Abstract** The mathematical methods used in quantum mechanics are developed, with emphasis on linear algebra and complex variables. Dirac notation for vectors in Hilbert space is introduced. The representation of coordinates and momenta in quantum mechanics is analyzed and applied to the Heisenberg uncertainty principle.

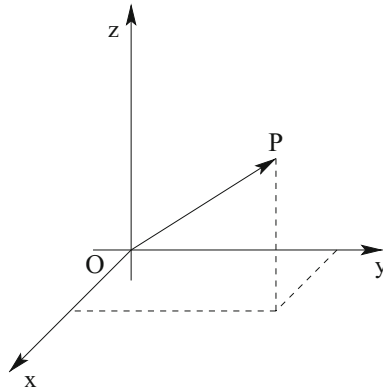
**Keywords** Vectors · Matrices · Hilbert space · Heisenberg uncertainty principle

### 2.1 Vector Analysis

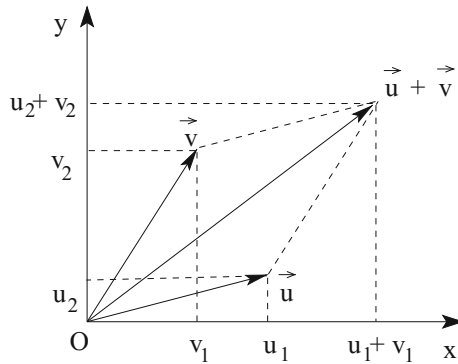
From a geometric point of view, any point  $P$  in the Cartesian  $x$ - $y$  plane can be associated with a *vector*  $\overrightarrow{OP}$ , from the origin  $O$  to the point  $P$ . The corresponding algebraic interpretation of a vector is an ordered pair of real numbers (the coordinates of  $P$ ). We will write either  $\mathbf{v} = (v_1, v_2)$  or  $\vec{v} = (v_1, v_2)$ . Both the boldface and the arrow notation are extensively used in the literature and we will use whichever one looks better in a formula. The origin  $O$  is the vector  $(0, 0)$ . The space of all these vectors is denoted by  $\mathbb{R}^2$ . The superscript 2 reminds us that two coordinates are sufficient to determine  $\mathbf{v}$ . Once we adopt the convention that all our vectors start from the origin, the terms vectors and points are equivalent. The generalization from two to three dimension is straightforward: a vector  $\overrightarrow{OP}$  in three-dimensional space is specified by three real numbers. The origin  $O$  is now  $(0, 0, 0)$ . This space, containing all sets of ordered triples of real numbers, is denoted by  $\mathbb{R}^3$ . A vector in 3-space is shown in Fig. 2.1.

Vectors in classical physics are used to represent forces, velocities, etc. What is the mathematical counterpart of the physical concept of doubling or tripling a force? It is easy to see that this is equivalent to doubling or tripling the coordinates of the endpoint  $P$ . Therefore  $2\mathbf{v}$  has the same direction of  $\mathbf{v}$ , but is twice as long. Its coordinates are  $(2v_1, 2v_2)$ . More generally, for any real number  $a$ , we can define:

$$a\mathbf{v} = (av_1, av_2). \quad (2.1)$$



**Fig. 2.1** A three-dimensional vector  $OP$



**Fig. 2.2** Vectors  $\mathbf{u} = (u_1, u_2)$ ,  $\mathbf{v} = (v_1, v_2)$  and their sum, using the parallelogram rule

From elementary physics, we know also that forces can be added by means of the parallelogram rule (see Fig. 2.2). Given two vectors  $\mathbf{u} = (u_1, u_2)$ ,  $\mathbf{v} = (v_1, v_2)$ , what are the coordinates of the sum  $\mathbf{u} + \mathbf{v}$ ? It is easy to see that the following definition:

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2), \quad (2.2)$$

is in agreement with the parallelogram rule.

We have shown, both from geometric and algebraic points of view, the two fundamental operations of the vector space  $\mathbb{R}^2$ : multiplication by a real number, and summation of two vectors. Following are some properties implied by the fundamental operations of a *real vector space* ( $a, b$  denote real numbers, while  $\mathbf{0} = (0, 0)$  is the null vector):

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}, \quad (2.3)$$

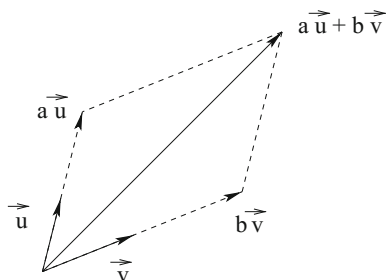
$$\mathbf{u} + \mathbf{0} = \mathbf{u}, \quad (2.4)$$

$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}, \quad (2.5)$$

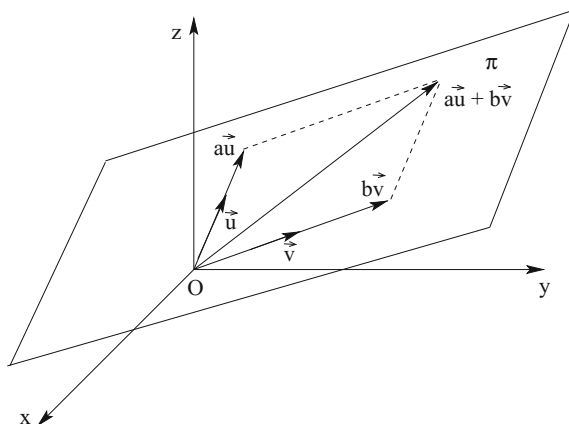
$$(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}. \quad (2.6)$$

Let us now introduce the very important concept of *linear combination* of vectors. Given two vectors  $\mathbf{u}$  and  $\mathbf{v}$  and two real numbers  $a$  and  $b$  the vector

$$\mathbf{w} = a\mathbf{u} + b\mathbf{v} \quad (2.7)$$



**Fig. 2.3** Linear combinations of the fixed two-dimensional vectors  $\mathbf{u}$ ,  $\mathbf{v}$  generate the whole plane  $\mathbb{R}^2$  by running over all the coefficients  $a$ ,  $b$



**Fig. 2.4** Linear combinations of the three-dimensional vectors  $\mathbf{u}$ ,  $\mathbf{v}$  generate the plane  $\pi$ , by varying the coefficients  $a$ ,  $b$

is called a linear combination of  $\mathbf{u}$  and  $\mathbf{v}$ , with coefficients  $a$  and  $b$ . It is evident that, except for the particular case in which  $\mathbf{u}$  and  $\mathbf{v}$  are parallel, two vectors with varying coefficients  $a, b$  can cover the entire plane  $\mathbb{R}^2$  (see Fig. 2.3). Since  $\mathbf{u}$  and  $\mathbf{v}$  generate, with their linear combinations, the entire plane  $\mathbb{R}^2$ , we say that  $\mathbf{u}$  and  $\mathbf{v}$  form a *basis* in  $\mathbb{R}^2$ . Vectors of the form  $\overrightarrow{OS} = \cos \theta \mathbf{u} + \sin \theta \mathbf{v}$  are a particular case of (2.7), representing vectors of unit length (*unit vectors*). If the two vectors  $\mathbf{u}, \mathbf{v}$  live in a space of larger dimension, such as  $\mathbb{R}^3$  (ordinary 3D physical space), the linear combination (2.7) still belongs to the plane containing  $\mathbf{u}$  and  $\mathbf{v}$ , and the set generated by  $\mathbf{u}$  and  $\mathbf{v}$  is the plane  $\pi$ , shown in Fig. 2.4. The construction of Fig. 2.3 is still valid, and the plane  $\pi$  still contains the two basis vectors  $\mathbf{u}$  and  $\mathbf{v}$ . Furthermore when  $a = b = 0$  we get  $\mathbf{w} = (0, 0, 0)$ , the null vector sitting at the origin. Therefore  $\pi$  must contain the origin  $O$ . The plane  $\pi$  is an example of *linear subspace* of  $\mathbb{R}^3$  since it is both a subset of  $\mathbb{R}^3$  and it is itself a linear space (indeed, if two vectors belong to  $\pi$ , their sum also belongs to  $\pi$ , etc.). Linear subspaces will play an important role in QM. The only nontrivial linear subspaces of  $\mathbb{R}^2$  are straight lines through the origin  $O$ .

We have seen that in the fundamental postulate of QM the *distance* of a point  $S$  from the origin  $O$  (the *length* of the vector  $\overrightarrow{OS}$ ) plays an important role. For an arbitrary vector  $\mathbf{v} = (v_1, v_2)$  we know by the Pythagorean theorem that its length (which we denote by  $|\mathbf{v}|$  or  $v$ ) is given by:

$$v = |\mathbf{v}| = \sqrt{v_1^2 + v_2^2} \quad (2.8)$$

If  $\mathbf{v} = (v_1, v_2, v_3)$  is a vector belonging to the space  $\mathbb{R}^3$ , its length being given by a simple generalization of Eq. (2.8):

$$v = |\mathbf{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2} \quad (2.9)$$

The *angle* between two vectors is also important. If  $\phi$  denotes the angle between the vectors  $\mathbf{u} = (u_1, u_2)$  and  $\mathbf{v} = (v_1, v_2)$  of the plane, the following relation holds:

$$u v \cos \phi = u_1 v_1 + u_2 v_2 \quad (2.10)$$

In three-dimensional space, a similar formula holds: denoting by  $\mathbf{u} = (u_1, u_2, u_3)$ ,  $\mathbf{v} = (v_1, v_2, v_3)$  two vectors of  $\mathbb{R}^3$ , and with the same meaning of the angle  $\phi$ , it is possible to prove that:

$$u v \cos \phi = u_1 v_1 + u_2 v_2 + u_3 v_3 \quad (2.11)$$

The reader will note that the expressions (2.10), (2.11) are quite similar. Indeed these expressions are more fundamental than the concept of “angle between two vectors,” which cannot be visualized in dimension higher than three. These expressions define the *scalar product*  $\mathbf{u} \cdot \mathbf{v}$  of two vectors. Thus in  $\mathbb{R}^2$  we have  $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2$ , in

$\mathbb{R}^3$  we have  $\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$ , etc. When writing the scalar product, the two vectors  $\mathbf{u}, \mathbf{v}$  can be represented by the symbols  $||u_1, u_2||, \begin{vmatrix} v_1 \\ v_2 \end{vmatrix}$ , called a *row vector* and a *column vector*, respectively. When a row vector is placed in front of a column vector, you can perform vector multiplication using a “row times column” sum, as follows:

$$||u_1, u_2|| \begin{vmatrix} v_1 \\ v_2 \end{vmatrix} = \mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 \quad (2.12)$$

The scalar product has the following properties:

1. The scalar product of a vector  $\mathbf{v}$  with itself is equal to the square of its length:

$$\mathbf{v} \cdot \mathbf{v} = v_1^2 + v_2^2 = |\mathbf{v}|^2 = v^2. \quad (2.13)$$

2. The commutative property:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}. \quad (2.14)$$

3. The distributive property:

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}. \quad (2.15)$$

4. Multiplying  $\mathbf{u}$  or  $\mathbf{v}$  by a real number  $a$ , gives the same multiple of the scalar product  $\mathbf{u} \cdot \mathbf{v}$ :

$$(a\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (a\mathbf{v}) = a \mathbf{u} \cdot \mathbf{v} \quad (2.16)$$

The scalar product is related to the *projection* of a vector onto a straight line. Consider a vector  $\mathbf{v}$  in the plane  $\mathbb{R}^2$  and a straight line  $r$  through the origin  $O$ . Denote by  $\mathbf{u}$  a unit vector (whose length is equal to 1) directed along  $r$ . The scalar product  $\mathbf{u} \cdot \mathbf{v}$  is equal to  $uv \cos \phi$ , where  $\phi$  is the angle between  $\mathbf{v}, \mathbf{u}$  (for simplicity we assume  $\cos \phi \geq 0$ ). Therefore  $\mathbf{u} \cdot \mathbf{v}$  is equal to the length of the vector  $\mathbf{v}'$  obtained by projecting  $\mathbf{v}$  onto  $r$  (see Fig. 2.5). Clearly then:

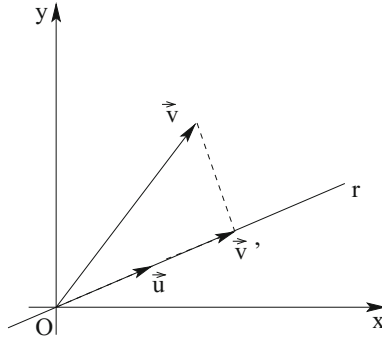
$$\mathbf{v}' = (\mathbf{u} \cdot \mathbf{v})\mathbf{u}. \quad (2.17)$$

Equations (2.13)–(2.17) can readily be generalized to a higher dimensional space. For example, in the four-dimensional space  $\mathbb{R}^4$ , which is the set of ordered *quadruples* of real numbers, the scalar product of two vectors  $\mathbf{u} = (u_1, u_2, u_3, u_4)$ ,  $\mathbf{v} = (v_1, v_2, v_3, v_4)$  is given by:

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3 + u_4v_4 \quad (2.18)$$

Setting  $\mathbf{u} = \mathbf{v}$  we obtain the square of the length of  $\mathbf{u}$ :

$$|\mathbf{u}|^2 = u_1^2 + u_2^2 + u_3^2 + u_4^2. \quad (2.19)$$



**Fig. 2.5** Projection  $\mathbf{v}'$  of vector  $\mathbf{v}$  on the straight line  $r$

Let us return to the simple case of the plane  $\mathbb{R}^2$ , and consider two orthogonal non-null vectors  $\mathbf{u}, \mathbf{v}$ ; from Eq. (2.10) we have:

$$\mathbf{u} \cdot \mathbf{v} = u v \cos \phi = 0 \quad (2.20)$$

since the cosine of a right angle vanishes. The same happens in dimension 3. When the dimension of the vector space is greater than 3, Eq. (2.20) can be regarded as the *definition* of orthogonality of two vectors. Finally, we consider the important concept for the spaces  $\mathbb{R}^2, \mathbb{R}^3, \dots$  of an *orthonormal basis*. Recall that the vectors  $\vec{\Psi}_A = (1, 0)$  and  $\vec{\Psi}_B = (0, 1)$  in our toy Hilbert space also were both of unit length and mutually orthogonal (see Fig. 1.28), therefore:

$$\vec{\Psi}_A \cdot \vec{\Psi}_A = |\vec{\Psi}_A|^2 = 1; \quad \vec{\Psi}_B \cdot \vec{\Psi}_B = |\vec{\Psi}_B|^2 = 1; \quad \vec{\Psi}_A \cdot \vec{\Psi}_B = 0. \quad (2.21)$$

The same property holds for the vectors  $\vec{\Psi}_1$  and  $\vec{\Psi}_2$ , which were also of unit length and orthogonal. Since any vector  $\mathbf{v} = (v_1, v_2)$  of  $\mathbb{R}^2$  can be written as a linear combination of  $\vec{\Psi}_A$  and  $\vec{\Psi}_B$ , for example,  $\mathbf{v} = v_1 \vec{\Psi}_A + v_2 \vec{\Psi}_B$ , we can say that they form a *basis*. In  $\mathbb{R}^3$ , the vectors  $\mathbf{i} = (1, 0, 0)$ ,  $\mathbf{j} = (0, 1, 0)$ ,  $\mathbf{k} = (0, 0, 1)$  form an orthonormal basis, since they are of unit length, pairwise orthogonal, and any vector  $\mathbf{v} = (v_1, v_2, v_3)$  can be written as the linear combination of  $\mathbf{i}, \mathbf{j}, \mathbf{k}$ :

$$\mathbf{v} = v_1 \mathbf{i} + v_2 \mathbf{j} + v_3 \mathbf{k} \quad (2.22)$$

In general, in an  $n$ -dimensional space  $\mathbb{R}^n$  (the set of all  $n$ -tuples of real numbers),  $n$  orthonormal vectors are required to form a *basis*; it is then possible to write *any* vector of the space as a linear combination of basis vectors.

## 2.2 Matrices in Quantum Mechanics

The earliest formulation of QM, developed around 1925 by Heisenberg, Born and Jordan, was called *matrix mechanics*. Classical observables, such as the position  $q$  or momentum  $p$  of a particle, were represented, in this theory, not by simple numbers, but rather by arrays  $Q$  and  $P$  containing an infinite number of rows and columns. The numbers appearing in an array might, for example, be related to the frequencies of radiation observed in a transition between two energy levels of an atom. Indeed the dimension of the array is equal to the number of these levels, and for a complete theory of even a simple system, such as the hydrogen atom, this number is infinite. The theory of infinite matrices is not at all simple, and this is a part of the reason that matrix mechanics is the less popular formulation of QM. Let us consider a simple situation, in which a quantum system has just two levels, which corresponds perfectly to our toy Hilbert space; it will be instructive to see how physical observables are represented in this model. All the formulas we have already encountered will turn out to have counterparts in the general case, almost without modification.

Consider the following  $2 \times 2$  array of real numbers:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad (2.23)$$

which is called a  $2 \times 2$  *matrix*. The numbers  $A_{11}, A_{12}, \dots$  are called *matrix elements*. In the following, unless otherwise specified, we refer to objects such as the array (2.23) as a matrix of dimension  $2 \times 2$ .

We can also think of this matrix as an “operator,” since it determines a transformation among the vectors of  $\mathbb{R}^2$ . Let us see how this happens. Given a vector  $\mathbf{v} = (v_1, v_2)$ , we can produce a new vector  $\mathbf{w} = (w_1, w_2)$  using the formulas:

$$w_1 = A_{11}v_1 + A_{12}v_2, \quad w_2 = A_{21}v_1 + A_{22}v_2. \quad (2.24)$$

A useful mnemonic for Eq. (2.24) is to consider the first row of the matrix as the row vector  $\mathbf{A}_1 = ||A_{11}, A_{12}||$ , and the second row as the row vector  $\mathbf{A}_2 = ||A_{21}, A_{22}||$ ; then Eq. (2.24) can be written using the row times column products:

$$w_1 = ||A_{11}, A_{12}|| \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{A}_1 \cdot \mathbf{v}, \quad w_2 = ||A_{21}, A_{22}|| \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{A}_2 \cdot \mathbf{v}. \quad (2.25)$$

Equation (2.24) can be written symbolically as:

$$\mathbf{w} = A\mathbf{v} \quad (2.26)$$

We say that the vector  $\mathbf{w}$  is the *image* of  $\mathbf{v}$  under the *mapping*  $A$ . Physicists use the term *operator* to denote the mapping  $A$  and, for them, the terms “matrix” and “operator” are used interchangeably (of course physicists are less meticulous than

mathematicians). As an elementary example, let  $A = \begin{pmatrix} 3 & 5 \\ 7 & 2 \end{pmatrix}$  and  $\mathbf{v} = (6, 4)$ . Then  $\mathbf{w} = (3 \times 6 + 5 \times 4, 7 \times 6 + 2 \times 4) = (38, 50)$ .

A simple, but essential, matrix is the following:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2.27)$$

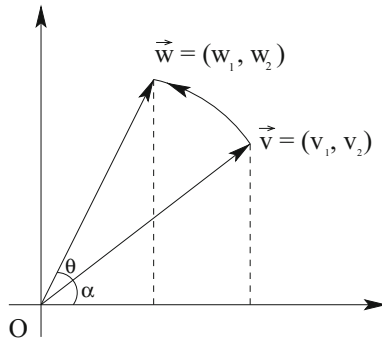
It has the property that it maps any vector into itself:  $I\mathbf{v} = \mathbf{v}$ . It is called the *identity matrix* or simply the *identity*. The matrix with all elements vanishing is called the *null matrix*. We will use the same symbol (in capital letters) for a matrix and the corresponding mapping. An important class of matrices represent rotations. For example, let  $R$  denote the matrix:

$$R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (2.28)$$

Equation (2.24) becomes:

$$\begin{aligned} w_1 &= \cos \theta v_1 - \sin \theta v_2 \\ w_2 &= \sin \theta v_1 + \cos \theta v_2 \end{aligned} \quad (2.29)$$

It is easy to verify that for any vector  $\mathbf{v}$  the vector  $\mathbf{w} = R\mathbf{v}$  is obtained by a counterclockwise rotation through an angle  $\theta$ . Indeed, if  $\alpha$  denotes the angle between  $\mathbf{v}$  and the  $x$  axis of the Cartesian plane, we have:  $v_1 = |\mathbf{v}| \cos \alpha$ ,  $v_2 = |\mathbf{v}| \sin \alpha$ , and substituting in Eq. (2.29) we have:  $w_1 = |\mathbf{v}| \cos(\alpha + \theta)$ ,  $w_2 = |\mathbf{v}| \sin(\alpha + \theta)$  (see Fig. 2.6). It is simple to verify that multiplying the matrix (2.28), which corresponds to a rotation through an angle  $\theta$ , by an analogous matrix with  $\theta$  replaced by  $\alpha$ , gives another rotation matrix with the angle  $\theta + \alpha$ , in agreement with the interpretation of successive applications of the two rotations.



**Fig. 2.6** The vector  $\mathbf{w}$  is obtained from the vector  $\mathbf{v}$  by rotation by an angle  $\theta$



A characteristic property of the mapping defined by Eq. (2.24) is *linearity*. This means that sums of vectors are sent into sums of images (geometrically, parallelograms are sent into parallelograms), linear combinations are sent into linear combinations, and so forth. In formulas, for all vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and all real numbers  $c$ :

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}, \quad A(c\mathbf{v}) = c(A\mathbf{v}). \quad (2.30)$$

These algebraic properties are essential in QM, consistent with the notion that matrices constitute a generalization of real numbers. Mathematicians tell us that, given a suitable definition of addition and multiplication, matrices form a *ring*, as do the real and complex numbers<sup>1</sup>; thus it is not entirely surprising that physical quantities can be represented by matrices.

We define the *sum*  $C$  of two matrices  $A$ ,  $B$ , written  $C = A + B$ , if the matrix elements of  $C$  are sums of the corresponding matrix elements of  $A$ ,  $B$ :

$$C_{ik} = A_{ik} + B_{ik}, \quad (i, k = 1, 2). \quad (2.31)$$

We can multiply a matrix  $A$  by a real number  $c$ :

$$(cA)_{ik} = cA_{ik}, \quad (i, k = 1, 2). \quad (2.32)$$

The simple rules of algebra also apply to matrices, for example:

$$(A + B)\mathbf{v} = A\mathbf{v} + B\mathbf{v}, \quad (cA)\mathbf{v} = c(A\mathbf{v}). \quad (2.33)$$

Let us now define the *product* of two matrices. The idea is that the product of the two successive linear mappings  $A$ ,  $B$ , on a vector, can be done by first applying  $B$ , then  $A$ :

$$(AB)\mathbf{v} = A(B\mathbf{v}) \quad \text{for any vector } \mathbf{v}. \quad (2.34)$$

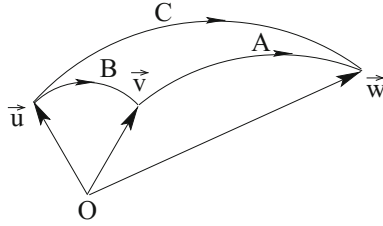
In Fig. 2.7, we see that if the mapping  $B$  sends  $\mathbf{u}$  in  $\mathbf{v}$ , and the mapping  $A$  sends  $\mathbf{v}$  in  $\mathbf{w}$ , then  $C = AB$  sends  $\mathbf{u}$  directly into  $\mathbf{w}$ ; these are pictorial representations of the operations:  $\mathbf{v} = B\mathbf{u}$ ,  $\mathbf{w} = A\mathbf{v}$ , and  $\mathbf{w} = C\mathbf{u}$ . In terms of matrix elements, the corresponding matrix products are given by

$$\begin{aligned} C_{11} &= A_{11}B_{11} + A_{12}B_{21}, & C_{12} &= A_{11}B_{12} + A_{12}B_{22}, \\ C_{21} &= A_{21}B_{11} + A_{22}B_{21}, & C_{22} &= A_{21}B_{12} + A_{22}B_{22}. \end{aligned} \quad (2.35)$$

Again the multiplication rule “rows times columns” applies. Writing  $\mathbf{A}_1 = ||A_{11}, A_{12}||$ ,  $\mathbf{A}_2 = ||A_{21}, A_{22}||$ , two row vectors, and  $\mathbf{B}_1 = \begin{vmatrix} B_{11} \\ B_{21} \end{vmatrix}$ ,  $\mathbf{B}_2 = \begin{vmatrix} B_{12} \\ B_{22} \end{vmatrix}$ , two column vectors, Eq. (2.35) can be written as:

---

<sup>1</sup>A field is a ring in which multiplication is commutative and every nonzero element has a multiplicative inverse. Thus real and complex numbers are also fields, while matrices are just rings.



**Fig. 2.7** The mapping  $C = AB$  is obtained by successive applications of the mappings  $B$ , then  $A$

$$\begin{aligned} C_{11} &= \mathbf{A}_1 \cdot \mathbf{B}_1, & C_{12} &= \mathbf{A}_1 \cdot \mathbf{B}_2, \\ C_{21} &= \mathbf{A}_2 \cdot \mathbf{B}_1, & C_{22} &= \mathbf{A}_2 \cdot \mathbf{B}_2. \end{aligned} \quad (2.36)$$

In general, an  $m \times n$  matrix is a rectangular array of numbers with  $m$  rows and  $n$  columns. For example, if  $m = 2$  and  $n = 3$  we have the matrix  $A$ :

$$A = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{vmatrix}. \quad (2.37)$$

We denote by  $A_{ij}$  the matrix element in the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column. Given a second matrix  $B$ , the matrix product  $AB$  requires that the number  $n$  of columns of  $A$  must match the number of rows of  $B$ ; thus  $B$  must be a  $n \times l$  matrix,  $l$  being arbitrary. In the general case, the matrix elements  $(AB)_{ik}$  are given by:

$$(AB)_{ik} = \sum_{j=1}^n A_{ij} B_{jk}, \quad i = 1, 2, \dots, m, \quad k = 1, 2, \dots, l \quad (2.38)$$

For example, if  $A$  is the matrix (2.37) and  $B$  is the  $3 \times 2$  matrix:

$$B = \begin{vmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{vmatrix}, \quad (2.39)$$

the matrix multiplication row times columns thus gives:

$$AB = \begin{vmatrix} A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31} & A_{11}B_{12} + A_{12}B_{22} + A_{13}B_{32} \\ A_{21}B_{11} + A_{22}B_{21} + A_{23}B_{31} & A_{21}B_{12} + A_{22}B_{22} + A_{23}B_{32} \end{vmatrix}, \quad (2.40)$$

so that  $AB$  is a  $2 \times 2$  square matrix. The geometrical meaning of the “operators”  $A$ ,  $B$  and  $AB$  is the following:  $B$  maps vectors belonging to  $\mathbb{R}^2$  into  $\mathbb{R}^3$ , while  $A$  maps vectors of  $\mathbb{R}^3$  into  $\mathbb{R}^2$ ; therefore  $AB$  maps vectors of  $\mathbb{R}^2$  into vectors of  $\mathbb{R}^2$ .

The application of a square matrix  $m \times m$  to a vector in  $\mathbb{R}^m$  is a particular case of Eq. (2.38); for example, setting  $n = m = 2$ ,  $B_{11} = v_1$ ,  $B_{21} = v_2$ , we obtain:

$$\begin{aligned}(AB)_{11} &= A_{11}v_1 + A_{12}v_2, \\ (AB)_{21} &= A_{21}v_1 + A_{22}v_2.\end{aligned}\tag{2.41}$$

Thus:

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} A_{11}v_1 + A_{12}v_2 \\ A_{21}v_1 + A_{22}v_2 \end{bmatrix}.\tag{2.42}$$

Many of the familiar formulas of elementary algebra still apply; for example, the associative property  $C(AB) = (CA)B$ ; the distributive property  $(A + B)C = AC + BC$ , etc., but a new feature appears: the commutative property does not hold! It is *not* true, in general, that  $AB = BA$ , as in elementary arithmetic. This fact has profound consequences in QM. (It is, in fact, the root of the uncertainty principle.) Let us give an example of two non-commuting matrices:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.\tag{2.43}$$

Their products are then given by:

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.\tag{2.44}$$

Since matrices represent *operations*, it is not unexpected that they sometimes do not commute. In everyday life, we can experience situations in which the order of operations is important: such as writing a letter and sealing it in an envelope. The result in elementary mathematics, that multiplying first by  $a$  and then by  $b$ , gives the same result as multiplying first by  $b$  and then by  $a$ , turns out to be rather exceptional in higher mathematics. The *commutator* of two matrices is defined by

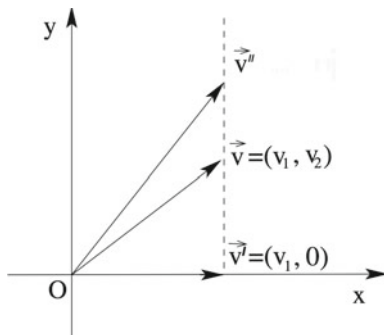
$$[A, B] = AB - BA.\tag{2.45}$$

If the commutator equals 0, then the matrices  $A$  and  $B$  commute:  $AB = BA$ .

The *inverse*  $A^{-1}$  of a matrix  $A$  is defined by the following property:

$$AA^{-1} = A^{-1}A = I.\tag{2.46}$$

where  $I$  denotes the identity matrix;  $A^{-1}$  corresponds to the *inverse* transformation. For example, the inverse of a rotation matrix through an angle  $\theta$  in a counterclockwise sense, is a rotation matrix through the same angle in a clockwise sense; in order to obtain  $R^{-1}$  it suffices to substitute  $-\theta$  in the place of  $\theta$  into Eq. (2.28). For the case of real numbers, the inverse (here, meaning the reciprocal) always exists except for the number zero. For  $2 \times 2$  matrices, the inverse exists unless the following expression



**Fig. 2.8** There are an infinite number of vectors  $\mathbf{v}$  whose projection on the  $x$  axis is equal to  $\mathbf{v}'$

vanishes:  $D = A_{11}A_{22} - A_{12}A_{21}$ . The geometric meaning of  $D$  (which is a  $2 \times 2$  determinant) is the ratio of the area of the parallelogram of two images  $A\mathbf{v}$ ,  $A\mathbf{u}$  to the area of the parallelogram of  $\mathbf{v}$ ,  $\mathbf{u}$ . It is a scale dilatation of the space  $\mathbb{R}^2$  under the action of the operator  $A$ . It is necessary that  $D \neq 0$  for the inverse of  $A$  to exist.

An important class of operators (or matrices) that usually do *not* admit an inverse<sup>2</sup> are *projection operators*. The simplest projection operator can be represented by the matrix:

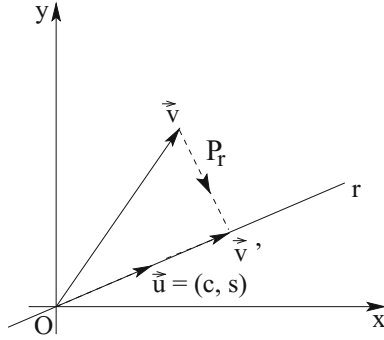
$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (2.47)$$

If  $\mathbf{v} = (v_1, v_2)$  is an arbitrary vector, the image  $P\mathbf{v} = (v_1, 0)$  is the vector  $\mathbf{v}'$  obtained by projecting  $\mathbf{v}$  onto the  $x$  axis, as shown in Fig. 2.8. The reason  $P$  does not admit an inverse is that there exist an infinite number of vectors  $\mathbf{v}''$  whose images  $P\mathbf{v}''$  coincide with  $\mathbf{v}'$ . These vectors  $\mathbf{v}''$  have their free end on a straight line parallel to the  $y$  axis. Mathematicians say that the mapping  $P$  is not *injective*. The inverse  $P^{-1}$  does not exist since it is ill-defined: which vector do we choose?  $\mathbf{v}'$ ,  $\mathbf{v}''$ , ...? Even in the simple space  $\mathbb{R}^2$  there are many projection operators. Given any straight line  $\mathbf{r}$  through the origin, let us denote by  $P_r$  the projection operator onto the line  $\mathbf{r}$ , as shown in Fig. 2.9. The matrix corresponding to  $P_r$  is easily found: Let  $\mathbf{u} = (c, s)$  be a unit vector directed along the line  $\mathbf{r}$ ; since  $|\mathbf{u}| = 1$ ,  $c^2 + s^2 = 1$ . Therefore:

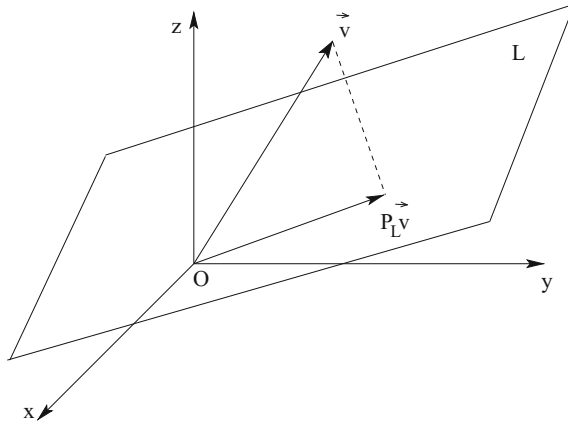
$$P_r = \begin{pmatrix} c^2 & cs \\ cs & s^2 \end{pmatrix}. \quad (2.48)$$

An example of a projection operator in  $\mathbb{R}^3$  is  $P_L$ , defined as follows: given a vector  $\mathbf{v} = (v_1, v_2, v_3)$  belonging to  $\mathbb{R}^3$ , and a plane  $L$  through the origin  $O$ ,  $P_L\mathbf{v}$  is the vector obtained taking the orthogonal projection of  $\mathbf{v}$  on the plane  $L$ , as shown in Fig. 2.10.

<sup>2</sup>Only the identity  $I$  is a projection operator that admits an inverse.



**Fig. 2.9** The action of the projection operator  $P_r$  in two dimensions



**Fig. 2.10** The action of the projection operator  $P_L$  in three dimensions

Some matrices can be associated with physical observables. Apart from some subtle points that we will discuss later, only matrices  $A$  such that  $A_{12} = A_{21}$  are possible candidates. (We are still limiting ourselves to *real* matrices.) These matrices are *symmetric matrices*, for example,

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad (2.49)$$

which is the most general  $2 \times 2$  symmetric matrix. Of course, the matrix  $R$  of Eq. (2.28) is not symmetric, since  $R_{12} = -\sin \theta$ , while  $R_{21} = \sin \theta$ .

We have noted above that matrices can be regarded as generalizations of numbers (which can be considered  $1 \times 1$  matrices). But in some cases, a matrix  $A$  can behave like a number in another way. This happens when it operates on particular vectors, called *eigenvectors*. For these vectors, the application of  $A$  is equivalent to the multiplication by a real number  $\lambda$ . More precisely, we shall say that a vector  $\mathbf{v} = (v_1, v_2)$  (excluding  $(0, 0)$ ) is an *eigenvector* of the matrix  $A$  corresponding to the eigenvalue  $\lambda$ , if the following relation holds:

$$A\mathbf{v} = \lambda\mathbf{v}. \quad (2.50)$$

For example, if  $\lambda = 3$ , the image vector  $A\mathbf{v}$  is three times longer than  $\mathbf{v}$ , if  $\lambda = \frac{1}{2}$ ,  $A\mathbf{v}$  is half of  $\mathbf{v}$ , and so forth. The important point is that the *direction* of the eigenvector remains unchanged. Equation (2.50) for a symmetric matrix  $A$  is equivalent to the two scalar equations:

$$\begin{aligned} A_{11}v_1 + A_{12}v_2 &= \lambda v_1, \\ A_{12}v_1 + A_{22}v_2 &= \lambda v_2. \end{aligned} \quad (2.51)$$

Symmetric matrices have the remarkable property (which is at the root of their utility in QM) of admitting an *orthonormal basis* of eigenvectors. In our toy space, for any symmetric matrix  $A$  there exist two vectors  $\vec{\psi}_1, \vec{\psi}_2$  such that:

$$A\vec{\psi}_1 = \lambda_1\vec{\psi}_1, \quad A\vec{\psi}_2 = \lambda_2\vec{\psi}_2. \quad (2.52)$$

Since these equations do not imply any restriction on their lengths,  $\vec{\psi}_1$  and  $\vec{\psi}_2$  can be chosen with unit lengths. Furthermore it can be shown that if  $\lambda_1$  is different from  $\lambda_2$ ,  $\vec{\psi}_1$  and  $\vec{\psi}_2$  are *orthogonal* to each other.

Matrices, such as the  $2 \times 2$  with  $A_{12} = A_{21} = 0$ , are called *diagonal*. They have the nice property that the eigenvectors are directed along the coordinate axis. We will denote them by  $\vec{\psi}_A, \vec{\psi}_B$ , in agreement with our notation in toy Hilbert space. As an example, let  $Q$  be the matrix:

$$Q = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}. \quad (2.53)$$

We can easily verify that

$$Q\vec{\psi}_A = 3\vec{\psi}_A, \quad Q\vec{\psi}_B = 2\vec{\psi}_B, \quad (2.54)$$

where  $\vec{\psi}_A = (1, 0)$  and  $\vec{\psi}_B = (0, 1)$ . Two diagonal matrices always commute. As an example, consider the matrix:

$$R = \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix} \quad (2.55)$$

and compute the *commutator*  $QR - RQ$ ; one finds a matrix with all elements equal to zero, namely, the null matrix, which we denote by 0. Therefore  $QR - RQ = 0$ . Of course,  $R$  admits as eigenvectors the same eigenvectors as  $Q$ , so that the vectors  $\vec{\Psi}_A$  and  $\vec{\Psi}_B$  are the same; only the eigenvalues of  $Q$  and  $R$  are different. This is a general rule: if two symmetric matrices commute, *they possess a common set of orthonormal eigenvectors*.

Returning to Eq. (2.51) for a symmetric matrix, we can now find the eigenvalues  $\lambda$ . Assuming, for simplicity, that  $v_2$  is not equal to zero, we can divide both equations by  $v_2$ . Denoting by  $r$  the quotient  $v_1/v_2$ , we get:

$$\begin{aligned} A_{11}r + A_{12} &= \lambda r, \\ A_{12}r + A_{22} &= \lambda. \end{aligned} \quad (2.56)$$

Solving for  $r$  in the second equation and substituting in the first, we obtain an equation determining the eigenvalues:

$$(A_{11} - \lambda)(A_{22} - \lambda) - A_{12}^2 = 0, \quad (2.57)$$

which is a simple quadratic equation in the variable  $\lambda$ . In the most general case, it has two solutions  $\lambda_1, \lambda_2$ , and for each solution, the relation  $A_{12}r + A_{22} = \lambda$  determines a possible value of  $r$ , which gives the *direction* of the corresponding eigenvector. (For simplicity, we neglect here the possibility of degeneracy, when more than one eigenvector corresponds to the *same* eigenvalue.) As an example, if all the matrix elements  $A_{ik}$  are equal to 1, the matrix  $A$  is simply:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (2.58)$$

and the eigenvalue Eq. (2.57) reduces to  $(1 - \lambda)^2 = 1$ , so that  $(1 - \lambda) = \pm 1$ , and the eigenvalues have the values 0 and 2. For  $\lambda = 0$  we get  $r = -1$ , thus the eigenvector is  $\vec{\Psi}_1 = (1, -1)$ , and for  $\lambda = 2$  we get  $r = 1$  and the eigenvector  $\vec{\Psi}_2 = (1, 1)$ . Note that  $\vec{\Psi}_1$  and  $\vec{\Psi}_2$  are orthogonal, as they should be. However, their lengths are not equal to 1 (actually to  $\sqrt{2}$ ). *Normalizing* the eigenvectors by dividing  $\vec{\Psi}_1$  and  $\vec{\Psi}_2$  by  $\sqrt{2}$ , we obtain two *orthonormal* eigenvectors (which we still denote by  $\vec{\Psi}_1, \vec{\Psi}_2$ , since they also obey Eq. (2.52) and no confusion need arise):

$$\vec{\Psi}_1 = \left( \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right), \quad \vec{\Psi}_2 = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right). \quad (2.59)$$

Let us compute the commutator  $QA - AQ$  of the matrices (2.53), (2.58). We find, using Eq. (2.35):

$$QA = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 2 & 2 \end{pmatrix}, \quad (2.60)$$

$$AQ = \begin{Bmatrix} 1 & 1 \\ 1 & 1 \end{Bmatrix} \begin{Bmatrix} 3 & 0 \\ 0 & 2 \end{Bmatrix} = \begin{Bmatrix} 3 & 2 \\ 3 & 2 \end{Bmatrix}. \quad (2.61)$$

Evidently,  $QA$  and  $AQ$  have rows and columns interchanged, so the commutator  $QA - AQ$  does not vanish. The noncommutativity of  $Q$ ,  $A$  and the differing eigenvectors of  $Q$  and  $A$  are, in fact, related. Indeed, we can state the following theorem:

**Theorem 2.1** *Two symmetric matrices admit a common basis of orthonormal eigenvectors if and only if they commute.*

### 2.3 Quantum Mechanics in Toy Hilbert Space

In Sect. 1.3, we introduced a “toy Hilbert space,” an extremely simplified representation for a two-state quantum system, whereby quantum states can be represented by unit vectors in the 2D Cartesian plane, with coordinates  $x$ ,  $y$ . Physical observables are correspondingly represented by real symmetric  $2 \times 2$  matrices. The theorem at the end of the last section is relevant to a fundamental interpretative postulate of QM:

**Postulate 1:** To any possible state of a physical system there corresponds a vector  $\vec{OS}$  of length 1. To any physical quantity  $F$  there corresponds a symmetric matrix (also denoted by  $F$ ). The possible results of a measurement of  $F$  on any state are the eigenvalues of the matrix  $F$ . If  $\vec{\psi}$  is a normalized eigenvector of  $F$  corresponding to the eigenvalue  $\lambda$ , so that

$$F\vec{\psi} = \lambda\vec{\psi}, \quad (2.62)$$

then  $|\vec{\psi} \cdot \vec{OS}|^2$  is the probability that the result of a measurement of  $F$  is  $\lambda$ .

To this we add:

**Postulate 2:** After the measurement of  $F$ , if the result is  $\lambda$ , the state vector  $\vec{OS}$  coincides with the eigenvector  $\vec{\psi}$ , thus verifying Eq. (2.62).

In order to explain the motivation for Postulate 2, we quote from Dirac (Dirac 1958):

From physical continuity, if we make a second measurement immediately after the first, the result of the second measurement must be the same as that of the first. Thus after the first measurement has been made, there is no indeterminacy in the result of the second ... This conclusion must still hold if the second measurement is not actually made.

Physical quantities like  $F$  are called *observables* by physicists. They correspond to symmetric matrices (more precisely, Hermitian matrices, see Sect. 2.6). Likewise, the corresponding physical states are represented by *state vectors*.

Suppose now that the matrices  $Q$  and  $A$  correspond to position  $q$  and velocity  $v$ , respectively. The matrix  $Q$  has the eigenvectors  $\vec{\psi}_A$ ,  $\vec{\psi}_B$ , while  $A$  has the eigenvectors  $\vec{\psi}_1$  and  $\vec{\psi}_2$ . In our toy Hilbert space, the possible results of a measurement of  $q$  are 3 and 2 (the eigenvalues of  $Q$ ), while the possible results of a measurement of  $v$  are 0 and 2 (the eigenvalues of  $A$ ). Furthermore, the expression  $|\vec{\psi}_1 \cdot \vec{OS}|^2$  is the square



of the projection of  $\vec{OS}$  on the straight line determined by  $\vec{\psi}_1$ , etc. The pair of “axis”  $\vec{\psi}_1, \vec{\psi}_2$  is “rotated” with respect to the “axes”  $\vec{\psi}_A, \vec{\psi}_B$  by virtue of the fact that the commutator  $QA - AQ$  does not vanish. By Postulate 2, if we first measure  $Q$ , the state vector  $\vec{OS}$  will “jump” to either  $\vec{\psi}_A$  or  $\vec{\psi}_B$ . In either case, a subsequent measurement of  $A$  will be uncertain. Physicists say that the observables  $Q$  and  $A$  are *not compatible*. When physicists realized that the matrices corresponding to very simple observables such as  $q$  and  $p$  (position and momentum of a particle) do not commute, it is not surprising that this possibility was initially regarded with skepticism. Actually, the matrices representing  $q$  and  $p$  are of infinite dimension, but the geometry of our toy Hilbert space is still a valid analogy. As a consequence of the mathematical structure of the theory,  $q$  and  $p$  do not admit common eigenvectors, similar to the situation we found for the matrices  $Q, A$  above. No state vector exists such that we can obtain with certainty (probability 1) a value of  $q$  and a value of  $p$ . The conclusion follows that the observables  $q$  and  $p$  cannot be simultaneously measured. An analogous result applies for any pair of non-commuting observables; and, since symmetric matrices do not, in general, commute, indeterminacy relations are quite commonplace, rather than an exception. Other than position and velocity, some well-known cases of non-commuting observables include two different components of angular momentum, as well as operators representing time and energy.

An important quantity in QM is the *average* or *expectation value* of an observable in a given state. Suppose the observable  $F$  is represented by the simple diagonal matrix:

$$F = \begin{vmatrix} F_A & 0 \\ 0 & F_B \end{vmatrix}. \quad (2.63)$$

We consider a completely general state vector  $\vec{OS} = (x, y)$ , requiring only the normalization condition  $x^2 + y^2 = 1$ , such that  $\vec{S}$  lies on a circle of radius 1 centered at the origin. As always, we suppose, that  $\vec{OS}$  represents the state of the system. Let us perform a measurement of  $F$ . We already know that the eigenvalues of  $F$  are the numbers  $F_A, F_B$ , and therefore the probability of finding the value  $F_A$  is  $x^2$ , and the probability of finding the value  $F_B$  is  $y^2$ . Then the *average*  $\bar{F}$  of the results of a measurement of  $F$  can be computed by the standard formula of statistics,  $\bar{F} = \sum_i F_i P_i$ , and we can write:

$$\bar{F} = F_A x^2 + F_B y^2. \quad (2.64)$$

Alternatively, by taking the scalar product of the vectors  $\vec{OS}$  and  $F\vec{OS}$ , we get the same result. In fact, the vector obtained applying the operator  $F$  to  $\vec{OS}$  is simply  $(F_A x, F_B y)$ ; taking the scalar product of this vector with  $\vec{OS} = (x, y)$  we obtain the right-hand side of Eq. (2.64), whereby

$$\bar{F} = \vec{OS} \cdot F\vec{OS}. \quad (2.65)$$

The last formula, which has been obtained in a very particular case, is actually a completely general and very elegant result.

Postulate 1 does not say anything about the time evolution of the state vector. Actually, the motion in the Hilbert space of vector  $\vec{OS}$  is determined by the *time-dependent Schrödinger equation*. In our toy space, the path of point  $S$  is simply the circumference of a circle of radius 1. Clearly, in spaces of higher dimension, this path is more complicated. We must imagine a point  $S$  moving *continuously* (without sudden jumps), maintaining its unit distance from the origin, just like a mass point constrained to the circumference of a circle. In the original formulation of QM, sudden jumps might occur when a measurement is made (see Postulate 2). We will come back to this subtle (and controversial) point later, exemplified by the question: “Are there quantum jumps?”.

In the work we have done thus far, observables  $F$  have been independent of time (as have both the eigenvectors and the eigenvalues), while the vector state  $\vec{OS}$  carries all the dependence on time. This is known as the *Schrödinger picture*. It is not difficult to formulate an alternative interpretation in Hilbert space, which corresponds to the same physical situation, but uses a *fixed* state vector but time-dependent operators. The idea is to rotate the eigenvectors of  $F$  back in such a way that their relative position with  $\vec{OS}$  (which now is fixed) is the same as in the Schrödinger picture. We need first the following result:

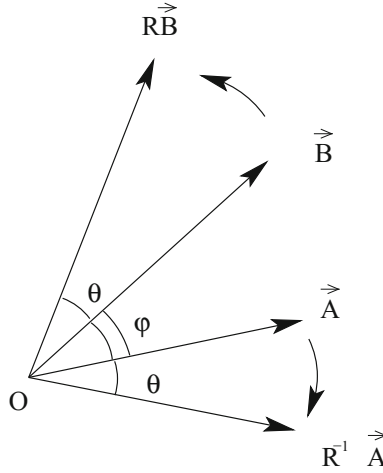
**Lemma 2.1** *If  $R$  is a rotation matrix, and  $\mathbf{A}$ ,  $\mathbf{B}$  are arbitrary vectors, the scalar product of  $\mathbf{A}$  with  $R\mathbf{B}$  is equal to the scalar product of  $\mathbf{B}$  with  $R^{-1}\mathbf{A}$ .*

An algebraic proof is elementary. The matrix of  $R^{-1}$  is obtained from the matrix  $R$  simply by changing the sign of  $\theta$  in Eq. (2.28). The following intuitive argument is perhaps more direct: consider the angle  $\phi$  between the vectors  $\mathbf{A}$ ,  $\mathbf{B}$  (see Fig. 2.11). Rotating  $\mathbf{B}$  in a counterclockwise sense through an angle  $\theta$ , we obtain the vector  $R\mathbf{B}$ , while the angle between  $\mathbf{A}$  and  $R\mathbf{B}$  becomes  $\theta + \phi$ . And if we keep  $\mathbf{B}$  fixed and rotate the vector  $\mathbf{A}$  back in a clockwise sense (by applying  $R^{-1}$ ) through an angle  $\theta$ , the angle between  $\mathbf{B}$  and  $R^{-1}\mathbf{A}$  remains equal to  $\phi + \theta$ . From Eq. (2.10), we see that the scalar product of two vectors depends on the lengths of the vectors and the angle between them. But rotations do not change lengths, and since the angle is  $\theta + \phi$  in both cases, the Lemma is proved.

Let us denote by  $\vec{\Psi}(t)$  the state vector  $\vec{OS}$  as a function of time  $t$ . Suppose, for simplicity, that at  $t = 0$  the state vector coincides with the  $x$  axis,  $\vec{\Psi}(0) = (1, 0)$ , and at time  $t = T$ , the state vector  $\vec{\Psi}(T)$  makes an angle  $\theta(T)$  with the  $x$  axis; in other words, during the time from  $t = 0$  and  $t = T$  the state vector is *rotated* through an angle  $\theta$ . Therefore the mapping from  $\vec{\Psi}(0)$  to  $\vec{\Psi}(T)$  can be obtained by means of the rotation matrix (2.28) and we have:

$$\vec{\Psi}(T) = R(T)\vec{\Psi}(0). \quad (2.66)$$

In the last equation, we have written  $R(T)$  to emphasize the dependence of the rotation  $R$  on time  $T$ . In the more general situation, the analog of Eq. (2.66) provides the



**Fig. 2.11** The scalar products  $\langle \vec{RB} | \vec{A} \rangle$  and  $\langle \vec{B} | \vec{R}^{-1}\vec{A} \rangle$  are equal

solution of the time-dependent Schrödinger equation, once the initial wave function  $\vec{\Psi}(0)$  is specified. Substituting (2.66) in place of  $\vec{OS}$  into Eq. (2.65) we get:

$$\bar{F} = R(T) \vec{\Psi}(0) \cdot F R(T) \vec{\Psi}(0). \quad (2.67)$$

Let us apply Lemma 1 with  $\vec{A} = F R(T) \vec{\Psi}(0)$  and  $\vec{B} = \vec{\Psi}(0)$ . This allows us to move the rotation operator  $R(T)$  to the other side of the scalar product, then replacing  $R$  by  $R^{-1}$ . We obtain:

$$\bar{F} = \vec{\Psi}(0) \cdot R(T)^{-1} F R(T) \vec{\Psi}(0). \quad (2.68)$$

We call the time-dependent operator  $F(T) = R(T)^{-1} F R(T)$  operator  $F$  in the Heisenberg picture, and we write:

$$\bar{F} = \vec{\Psi}(0) \cdot F(T) \vec{\Psi}(0). \quad (2.69)$$

Of course, the state vector in the Heisenberg picture is  $\vec{\Psi}(0)$ . We see from formulas (2.69), (2.65) that the expression for  $\bar{F}$  is the same in the two pictures. However, in the Heisenberg picture, the time dependence has been entirely transferred to the operator representing the observable. This is quite analogous to the picture in classical mechanics, in which we seek the changes in observables with time, as described by equations of motion.

A final topic we want to introduce for our toy Hilbert space is the *density matrix*. For any matrix  $A_{ij}$ , the sum of the elements of the main diagonal is called the *trace* of  $A$ , denoted by  $\text{Tr } A$ . Therefore:

$$\text{Tr } A = A_{11} + A_{22}. \quad (2.70)$$

Given a unit vector  $\vec{OS} = (x, y)$ , the projection operator  $P$  on the line determined by  $\vec{OS}$  is given, using Eq. (2.48) and setting  $c = x$ ,  $s = y$ :

$$P = \begin{vmatrix} x^2 & xy \\ xy & y^2 \end{vmatrix}. \quad (2.71)$$

Given the operator  $F$ , the trace of the product  $PF$  is given by:

$$\text{Tr}(PF) = \text{Tr} \begin{vmatrix} x^2 F_{11} + xy F_{12} & x^2 F_{12} + xy F_{22} \\ xy F_{11} + y^2 F_{12} & xy F_{12} + y^2 F_{22} \end{vmatrix} = x^2 F_{11} + 2xy F_{12} + y^2 F_{22}. \quad (2.72)$$

Let us prove that the last expression constitutes a generalization of Eq. (2.64) when  $F$  is not diagonal. First compute (2.65): the coordinates of the vector  $F\vec{OS}$  are  $(F_{11}x + F_{12}y, F_{12}x + F_{22}y)$ ; then take the scalar product of this vector with  $\vec{OS} = (x, y)$ , giving precisely the expression (2.72). Therefore:

$$\text{Tr}(PF) = \langle \vec{OS} | F \vec{OS} \rangle = \bar{F}. \quad (2.73)$$

Note that if  $F$  equals the identity  $I$ , Eq. (2.27) reduces to

$$\text{Tr}(PI) = \text{Tr } P = x^2 + y^2 = 1. \quad (2.74)$$

Consider now two orthogonal states  $\vec{\psi}_1, \vec{\psi}_2$  and suppose that there is a probability  $p_1$  that the state vector  $\vec{OS}$  of a physical system coincides with  $\vec{\psi}_1$ , and a probability  $p_2$  that it coincides with  $\vec{\psi}_2$ . Note that in the actual case the probabilities  $p_1$  and  $p_2$  are not the fundamental probabilities of QM (which Heaven only knows!). Here,  $p_1$  and  $p_2$  might represent classical probabilities of distinct physical situations, as we encounter, in classical statistical mechanics. In any event, we must have  $p_1 + p_2 = 1$ . In order to obtain the average value of an observable  $F$ , we compute a double average: first we find the two quantum averages  $\langle \vec{\psi}_1 | F \vec{\psi}_1 \rangle, \langle \vec{\psi}_2 | F \vec{\psi}_2 \rangle$ , and then we average these results, making use of the probabilities  $p_1, p_2$ ; at the end of this procedure we get:

$$\bar{F} = p_1 \langle \vec{\psi}_1 | F \vec{\psi}_1 \rangle + p_2 \langle \vec{\psi}_2 | F \vec{\psi}_2 \rangle. \quad (2.75)$$

Now let  $P_1, P_2$  be the projection operators onto the straight lines determined by  $\vec{\psi}_1, \vec{\psi}_2$ . Using Eq. (2.73) we have  $\langle \vec{\psi}_1 | F \vec{\psi}_1 \rangle = \text{Tr}(P_1 F)$ , and  $\langle \vec{\psi}_2 | F \vec{\psi}_2 \rangle = \text{Tr}(P_2 F)$ . Therefore, introducing the *density matrix*  $\rho = p_1 P_1 + p_2 P_2$  and using the fact that taking the trace is a *linear* operation, we can write:

$$\bar{F} = p_1 \text{Tr}(P_1 F) + p_2 \text{Tr}(P_2 F) = \text{Tr}[(p_1 P_1 + p_2 P_2) F] = \text{Tr}(\rho F). \quad (2.76)$$

Equation (2.76) constitutes a generalization of Eq. (2.73); knowledge of the density matrix  $\rho$  allows us to compute averages of any observable  $F$ ; therefore  $\rho$  determines the *state* of the system in way analogous to the state vector  $\vec{OS}$ . When  $p_1 = 1$  and  $p_2 = 0$ , or  $p_1 = 0$  and  $p_2 = 1$ , this reduces to the previous case; we say that the system is in a *pure state*. The more general state defined by  $\rho = p_1 P_1 + p_2 P_2$  is called a *mixed state*. Let us verify that  $p_1, p_2$  are the eigenvalues of  $\rho$ , and  $\vec{\psi}_1, \vec{\psi}_2$ , its eigenvectors. Denoting by  $\vec{0}$  the null vector, we have:

$$P_1 \vec{\psi}_1 = \vec{\psi}_1, \quad P_2 \vec{\psi}_1 = \vec{0}, \quad (2.77)$$

and therefore:

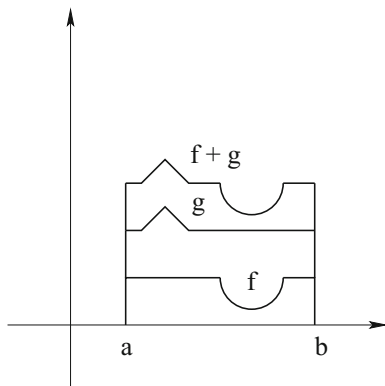
$$\rho \vec{\psi}_1 = (p_1 P_1 + p_2 P_2) \vec{\psi}_1 = p_1 \vec{\psi}_1 + p_2 \vec{0} = p_1 \vec{\psi}_1. \quad (2.78)$$

In the same way we can prove that  $\rho \vec{\psi}_2 = p_2 \vec{\psi}_2$ . Since there is no restriction on the pair of orthogonal vectors  $\vec{\psi}_1, \vec{\psi}_2$ , we see that the most general density matrix is a symmetric matrix whose eigenvalues are positive numbers  $p_1, p_2$  less than or equal to 1, and such that  $p_1 + p_2 = 1$ . Using the relation  $x^2 + y^2 = 1$ , it can be verified that the projection operator  $P$ , given by Eq. (2.71), is *idempotent*, meaning that it obeys the relation  $P^2 = P$ . This condition is, in fact, a defining characteristic of a pure state.

## 2.4 The Hilbert Space of Real Wavefunctions

We have now acquired an understanding of the toy model, but it may still not be entirely clear why wave functions representing “clouds of probability” have anything to do with vectors of the plane  $\mathbb{R}^2$ . The answer of a mathematician might again run: “Both  $\mathbb{R}^2$  and the set of wave functions of a physical system are vector spaces endowed with a scalar product.” However, to show that wave functions do indeed belong in a Hilbert space, we will follow a more elementary, less abstract, line of development: we will exhibit an “analogy” between the vectors of  $\mathbb{R}^2, \mathbb{R}^3, \dots$ , and the set  $C(a, b)$  of continuous wavefunctions defined on an interval  $[a, b]$  of the real axis. Actually, physical wave functions  $\psi(x, y, z)$  are defined on points  $(x, y, z)$  of three-dimensional space; indeed  $|\psi(x, y, z)|^2$  is actually the *probability density* in the clouds drawn in Chap. 1. To simplify the mathematics, we can imagine that our physical system to be one-dimensional, so that the wavefunctions depend on just a single variable  $x$ , on a line segment  $[a, b]$ .

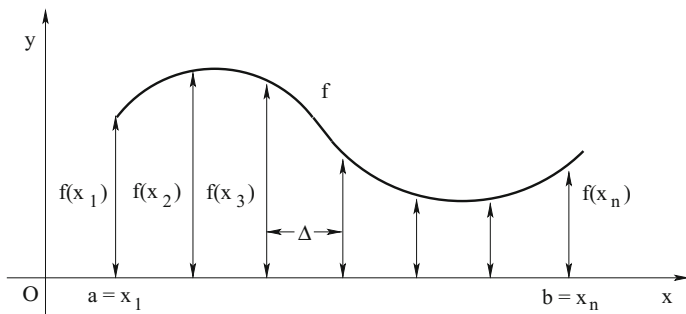
Two of the fundamental operations of a vector space, given in Eqs. (2.1) and (2.2), have obvious analogs for our set of functions: given two continuous functions  $f(x), g(x)$ , their sum is the function  $f(x) + g(x)$ , as shown in Fig. 2.12. Also, just as we can multiply the coordinates of a vector; by a real number  $\lambda$ , to obtain a new vector  $\lambda \mathbf{f}$  in the same direction as the original, we can likewise multiply a function  $f(x)$ , to give the analogous scaled function  $\lambda f(x)$ . The operations of addition and multiplication



**Fig. 2.12** Sum of two functions  $f, g$

by a real number suggests the terminology *linear combination of two functions*  $f(x)$ ,  $g(x)$ , with real coefficients  $a, b$ , namely, the function  $af(x) + bg(x)$ . The set of such linear combinations can be thought of as a *three-dimensional subspace of  $C(a, b)$* , provided  $f(x), g(x)$  do not have the same “direction,” meaning that  $f(x)$  is not simply a multiple of  $g(x)$ . This two-dimensional subspace can be thought of as a plane through the origin. What is the origin? The object analogous to the null vector  $(0, 0)$  of  $\mathbb{R}^2$  is a function which equals zero everywhere:  $f(x) = 0$  for all  $x$ .

A more challenging question is: what constitutes the *coordinates* of a function, which are somehow the analogs of the components of a vector? Later we will give a more rigorous answer to this question; for the moment, we tentatively settle for a more heuristic approach, which will enable us to understand the meaning of the scalar product of two wavefunctions. Let  $f$  be a continuous function defined on  $[a, b]$ ; see Fig. 2.13, where the graph of  $f(x)$  is shown. We choose  $n$  equally spaced points  $x_1, x_2, \dots, x_n$  in the interval (so that  $x_1 = a$  and  $x_n = b$ ), and we compute the values of the functions  $f(x_1), f(x_2), \dots, f(x_n)$ . We can then imagine these numbers to be



**Fig. 2.13** A function  $f(x)$  is approximated by a vector in  $\mathbb{R}^n$ ,  $\mathbf{f} = (f(x_1), f(x_2), \dots, f(x_n))$

the *coordinates* of a vector belonging to  $\mathbb{R}^n$ . For example, let us take  $a = 0$ ,  $b = 3$ ,  $n = 4$  and consider the simple function  $f(x) = x^2$ . Then  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$ ,  $x_4 = 3$ , and  $f(x_1) = 0$ ,  $f(x_2) = 1$ ,  $f(x_3) = 4$ ,  $f(x_4) = 9$ . We have obtained the vector  $(0, 1, 4, 9)$  of  $\mathbb{R}^4$ . You may ask: How do we choose the number  $n$ ? Indeed this number is arbitrary, since given a function on an interval, we can compute it at as many points as we want. This is an inherent weakness in the identification of the values  $f(x_1), f(x_2), \dots, f(x_n)$  as “coordinates of  $f$ .” However, let us boldly proceed nonetheless, and try to find a tentative definition of the *scalar product of two functions* which is analogous to the definitions (2.10), (2.11) of the scalar product of two vectors?

Given two functions  $f$  and  $g$ , both continuous on the interval  $[a, b]$ , let us compute these functions on the equally spaced points  $x_1 = a$ ,  $x_2, \dots, x_n = b$ , as above. We will obtain, in this way, two vectors  $\mathbf{f}, \mathbf{g}$  belonging to  $\mathbb{R}^n$ :

$$\begin{aligned}\mathbf{f} &= (f(x_1), f(x_2), \dots, f(x_n)), \\ \mathbf{g} &= (g(x_1), g(x_2), \dots, g(x_n)),\end{aligned}\tag{2.79}$$

whose scalar product is given by (see the analog in  $\mathbb{R}^4$ , Eq. (2.18)):

$$f(x_1)g(x_1) + f(x_2)g(x_2) + \dots + f(x_n)g(x_n).\tag{2.80}$$

However this expression depends on the number  $n$ , which is arbitrary. Since our knowledge of a function becomes more precise when we know more and more values  $f(x_i)$  (think, for example, of these values as data in an experiment), we can *take the limit* of Eq. (2.80) as  $n$  approaches infinity. Even for the elementary case of two constant functions, say  $f(x) = 2$  and  $g(x) = 3$ , this limit is not finite. On the other hand, there exists an expression which is similar to Eq. (2.80) that admits a finite limit when  $n \rightarrow \infty$ , and will provide us with a more rigorous definition. Denoting by  $\Delta_n$  the distance between two consecutive points, we have  $\Delta_n = x_2 - x_1 = x_3 - x_2, \dots = \frac{(b-a)}{(n-1)}$ . It is easy to see that the limit:

$$\lim_{n \rightarrow \infty} \Delta_n [f(x_1)g(x_1) + f(x_2)g(x_2) + \dots + f(x_n)g(x_n)]\tag{2.81}$$

exists and corresponds to the definition of a Riemann integral:

$$\int_a^b f(x)g(x) dx.\tag{2.82}$$

Thus the expression (2.82) provides us with a consistent definition of *scalar product of two real functions*  $f(x), g(x)$ . To emphasize the analogy with vectors, we can write:

$$\mathbf{f} \cdot \mathbf{g} = \int_a^b f(x)g(x) dx.\tag{2.83}$$

The importance of this definition of scalar product can hardly be overestimated. It allows us to continue using a geometric language in the space of wavefunctions, and suggests an intuitive picture of physical states, even when we are referring, not to physical space, but to an abstract function space. We continue to define the *norm* or *length* of the vector  $\mathbf{f}$  using the expression:

$$|\mathbf{f}| = \sqrt{\mathbf{f} \cdot \mathbf{f}} = \sqrt{\int_a^b dx f(x)^2}. \quad (2.84)$$

As an example, let us evaluate the scalar product of the functions  $x$  and  $1 + x^2$  on the interval  $[0, 1]$ . We must perform the integration:

$$\int_0^1 (1 + x^2)x = \left[ \frac{x^2}{2} + \frac{x^4}{4} \right]_0^1 dx = \frac{3}{4}. \quad (2.85)$$

Table 2.1 shows the analogies between vectors in  $\mathbb{R}^n$  and the corresponding functional relations. On the left of the table, we show expressions involving the vectors  $\mathbf{v}$ ,  $\mathbf{w}$ ,  $\dots$ ; on the right are the analogs, in terms of the functions  $f(x)$ ,  $g(x)$  or  $\mathbf{f}$ ,  $\mathbf{g}$ ,  $\dots$ . Such correspondences will be particularly useful in Dirac's bra/ket formulation of QM.

**Table 2.1** Analogies between vectors and functions

Vectors	Functions
Components of a vector $\mathbf{v} = (v_1, v_2, \dots, v_n)$	Values of a function $f(x_1), f(x_2), \dots, f(x_n)$
Sum of two vectors $\mathbf{v} + \mathbf{w}$	Sum of two functions $f(x) + g(x)$
Linear combination of two vectors $c_1\mathbf{v} + c_2\mathbf{w}$	Linear combination of two functions $c_1f(x) + c_2g(x)$
Scalar product of two vectors $\mathbf{v} \cdot \mathbf{w} = v_1w_1 + v_2w_2 + \dots + v_nw_n$	Scalar product of two functions $\mathbf{f} \cdot \mathbf{g} = \int_a^b f(x)g(x) dx$
Norm of a vector $ \mathbf{v}  = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$	Norm of a function $ \mathbf{f}  = \sqrt{\mathbf{f} \cdot \mathbf{f}} = \sqrt{\int_a^b dx f(x)^2}$
Linearity of the scalar product $\langle c\mathbf{v}   \mathbf{w} \rangle = c \langle \mathbf{v}   \mathbf{w} \rangle$ $\langle \mathbf{v}   \mathbf{w}_1 + \mathbf{w}_2 \rangle = \langle \mathbf{v}   \mathbf{w}_1 \rangle + \langle \mathbf{v}   \mathbf{w}_2 \rangle$	Linearity of the scalar product $\langle c\mathbf{f}   \mathbf{g} \rangle = c \langle \mathbf{f}   \mathbf{g} \rangle$ $\langle \mathbf{f}   \mathbf{g}_1 + \mathbf{g}_2 \rangle = \langle \mathbf{f}   \mathbf{g}_1 \rangle + \langle \mathbf{f}   \mathbf{g}_2 \rangle$
Distance between two vectors $d =  \mathbf{v} - \mathbf{w}  = \sqrt{\langle \mathbf{v} - \mathbf{w}   \mathbf{v} - \mathbf{w} \rangle} = [(v_1 - w_1)^2 + (v_2 - w_2)^2 + \dots + (v_n - w_n)^2]^{\frac{1}{2}}$	Distance between two functions $d =  \mathbf{f} - \mathbf{g}  = \sqrt{\langle \mathbf{f} - \mathbf{g}   \mathbf{f} - \mathbf{g} \rangle} = \left[ \int_a^b dx (f(x) - g(x))^2 \right]^{\frac{1}{2}}$



The proof of the relations:  $\langle c\mathbf{f}|\mathbf{g}\rangle = c\langle\mathbf{f}|\mathbf{g}\rangle$ , and  $\langle\mathbf{f}|\mathbf{g}_1 + \mathbf{g}_2\rangle = \langle\mathbf{f}|\mathbf{g}_1\rangle + \langle\mathbf{f}|\mathbf{g}_2\rangle$  is elementary also in the case of functions, since

$$\int_a^b [cf(x)]g(x) dx = c \int_a^b f(x)g(x) dx, \quad (2.86)$$

and

$$\int_a^b f(x)[g_1(x) + g_2(x)] dx = \int_a^b f(x)g_1(x) dx + \int_a^b f(x)g_2(x) dx. \quad (2.87)$$

The analogy between the distance between vectors and the “distance” between functions deserves a word of comment: if distance  $d$  is very small the coordinates of the vectors  $\mathbf{v}$ ,  $\mathbf{w}$  are almost equal, since the sum of the positive numbers  $(v_1 - w_1)^2$ ,  $(v_2 - w_2)^2$ ,  $\dots$ , cannot be small unless every one of these contributions is small. In an analogous way, for functions, a very small value of  $d$  means, by and large, that the graphs of the functions  $f(x)$ ,  $g(x)$  are very close together. (There might be exceptions, in which the difference of the functions is large in small intervals on the  $x$ -axis.)

Another case to be considered is the existence of functions that do *not* have finite norm. A simple example is the function  $f(x) = \frac{1}{\sqrt{x}}$ , defined on the *open* interval  $(0, 1)$ , that is, the interval excluding the endpoints 0, 1. In fact,  $\int_0^1 f(x)^2 dx = \int_0^1 \frac{1}{x} dx = \infty$  or, better,  $\lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 \frac{1}{x} dx = \infty$ , since the function  $x^{-1}$  becomes very large for small  $x$ . Such behavior is excluded from our formalism, since we have restricted functions to be continuous and well defined in the whole interval  $[a, b]$  ( $\frac{1}{x}$  is not defined for  $x = 0$ ). However, in physics, the interval  $[a, b]$  is often the entire  $x$  axis, so that our integration  $\int_a^b$  becomes  $\int_{-\infty}^{+\infty}$ . Therefore, even some very simple functions such as  $x^2$ ,  $x^4$ , etc., must be excluded since their integrals diverge to  $\infty$ . However, since a wave function  $f(x)$  gives the probability amplitude of finding a particle at point  $x$ , it is reasonable to assume that this amplitude goes to zero when  $x$  becomes very large (for example, an electron bound to an atom has practically zero probability of being found on the Moon). Coming back to the purely mathematical aspects of the theory (while leaving aside certain mathematical subtleties), we will define as a *Hilbert space*, denoted by  $L^2(a, b)$ , the set of functions  $f(x)$  such that

$$\int_a^b [f(x)]^2 dx < \infty, \quad (2.88)$$

meaning that the integral must be *finite*. Therefore,  $\frac{1}{x}$ , for example, does *not* belong to  $L^2(0, 1)$ . In many physical applications, we will have  $a = -\infty$ ,  $b = +\infty$ , with the corresponding Hilbert space denoted by  $L^2(-\infty, +\infty)$ .

We usually assume that functions belonging to the Hilbert space correspond to vectors of *finite* length. The scalar product must then also be finite. As an example, the wave function:

**Table 2.2** Vector products and integrals of functions

Two vectors $\mathbf{v}, \mathbf{w}$ are orthogonal if $\mathbf{v} \cdot \mathbf{w} = 0$ , so that $v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = 0$	Two functions $f, g$ orthogonal if $\mathbf{f} \cdot \mathbf{g} = 0$ , so that $\int_a^b f(x)g(x) dx = 0$
A vector $\mathbf{v}$ is normalized if $ \mathbf{v} ^2 = v_1^2 + v_2^2 + \cdots + v_n^2 = 1$	A function $f(x)$ is normalized if $ \mathbf{f} ^2 = \int_a^b f(x)^2 dx = 1$
A basis of $n$ orthonormal vectors in $\mathbb{R}^n$ is a set of $n$ vectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(n)}$ such that $\mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$	A basis of orthonormal functions in $L^2$ is a sequence of $(\infty)$ functions $f_1(x), f_2(x), \dots, f_n(x) \dots$ such that $\mathbf{f}_i \cdot \mathbf{f}_j = \int_a^b f_i(x)f_j(x) dx = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$
Expansion of a vector in an orthonormal basis: $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}^{(i)}$ $v_i = \mathbf{e}^{(i)} \cdot \mathbf{v}$	Expansion of a function in an orthonormal basis: $f(x) = \sum_{i=1}^{\infty} c_i f_i(x)$ $c_i = \mathbf{f}_i \cdot \mathbf{f} = \int_a^b f_i(x)f(x)dx$

$$\frac{1}{\sqrt[4]{\pi}} e^{-(x-a)^2/2} \quad (2.89)$$

has norm equal to one and represents a “cloud” of probability localized around the point  $x = a$ , decreasing rapidly as  $|x - a|$  becomes large. To complete our analogy between vectors and functions, there is no problem in extending the concept of orthogonality to functions; by the definition of scalar product, we can say that two functions  $f(x), g(x)$  are *orthogonal* in  $L^2(a, b)$  if  $\langle \mathbf{f} | \mathbf{g} \rangle = \int_a^b f(x)g(x) dx = 0$ . With this definition in mind, we define an *orthonormal set of functions*  $f_1(x), f_2(x), \dots, f_n(x)$ , such that all functions have “length” 1 and are orthogonal to one other, as in the simple example of the vectors  $(1, 0), (0, 1)$  in the plane  $\mathbb{R}^2$  or the vectors  $\mathbf{i} = (1, 0, 0), \mathbf{j} = (0, 1, 0)$  and  $\mathbf{k} = (0, 0, 1)$  in the space  $\mathbb{R}^3$ . We know (see Eq. (2.22)) that any vector  $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$  can be written as the linear combination  $\mathbf{v} = v_1 \mathbf{i} + v_2 \mathbf{j} + v_3 \mathbf{k}$ . Furthermore, the components  $v_1, v_2, v_3$  satisfy the relations:

$$\mathbf{v} \cdot \mathbf{i} = v_1, \quad \mathbf{v} \cdot \mathbf{j} = v_2, \quad \mathbf{v} \cdot \mathbf{k} = v_3, \quad (2.90)$$

which can be generalized to any orthonormal basis of  $n$  vectors in  $\mathbb{R}^n$ . In other words, the scalar product of any vector  $\mathbf{v}$  with the  $i^{\text{th}}$  basis vector, gives the magnitude of the  $i^{\text{th}}$  “coordinate” of  $\mathbf{v}$  with respect to the  $i^{\text{th}}$  “axis.” This result suggests other analogous definitions and formulas, which we summarize in Table 2.2.

According to the last row of the table, for the case of vectors in  $\mathbb{R}^n$ , the equality  $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}^{(i)}$  has an obvious meaning. It implies that  $n$  orthonormal vectors do form a *basis* on which we can expand any vector; we know that in order to have such a basis we need two vectors in  $\mathbb{R}^2$  or three vectors in  $\mathbb{R}^3$ , etc. What happens in the Hilbert space  $L^2(a, b)$ ? There must then exist sequences of an infinite number of orthonormal functions. For example, if we take  $a = 0$  and  $b = 2\pi$ , the following functions:

$$\begin{aligned}
f_1(x) &= \frac{1}{\sqrt{2\pi}}, & f_2(x) &= \frac{\cos x}{\sqrt{\pi}}, & f_3(x) &= \frac{\sin x}{\sqrt{\pi}}, \\
f_4(x) &= \frac{\cos 2x}{\sqrt{\pi}}, & f_5(x) &= \frac{\sin 2x}{\sqrt{\pi}}, & \dots,
\end{aligned} \tag{2.91}$$

do form an orthonormal system in  $L^2(0, 2\pi)$ . The function  $f(x)$  is now represented by an infinite sum over these basis functions (this might be recognized as a Fourier series):

$$f(x) = \sum_{i=1}^{\infty} c_i f_i(x). \tag{2.92}$$

There remain questions of convergence and such, but we will not worry about these. If, indeed, Eq. (2.92) does hold for some orthonormal system of functions, such as the sequence (2.91), this set of functions is called *complete* and thereby provides a *basis* for expanding any admissible function in the Hilbert space. Given a function  $f(x)$ , the coefficients  $c_i$  (for  $i = 1, 2, 3 \dots$ ) are the best candidates to be designated “coordinates” of the vector  $\mathbf{f}$ ; this interpretation has a more rigorous foundation than the one we have introduced earlier, when we cited the values  $f(x_1), f(x_2), \dots, f(x_n)$ . In Dirac’s formalism, the two interpretation can be unified in an elegant (but not entirely rigorous) way, which is beyond the scope of our coverage. The limit implied by the infinite summation appearing in Eq. (2.92) must be understood in the following sense: the Hilbert space *distance*  $|\mathbf{f} - \mathbf{f}_n|$  between the function  $f_n(x) = \sum_{i=1}^n c_i f_i(x)$  and the function  $f(x)$  tends to zero when  $n \rightarrow \infty$ .

Now that we have a better understanding of what the Hilbert space is, we can further extend our analogy between vectors and functions, and ask: what are the “matrices” or better the *linear operators* relevant to QM which act in the Hilbert space  $L^2(-\infty, +\infty)$ , analogous to the way  $2 \times 2$  matrices act on vectors in the plane? Let us give two key examples of operators defined in this Hilbert space (leaving aside mathematical subtleties, that are treated, for example, in Fano 1971):

1. The operator that multiplies any function  $\psi(x)$  by the variable  $x$ . This operator is the famous  $q$  operator of QM, which represents the position of a particle. The result of the application of  $q$  to  $\psi$  is the function  $x\psi(x)$ , as follows:

$$(q\psi)(x) = x\psi(x), \tag{2.93}$$

where  $(q\psi)(x)$  is the image function  $q\psi$  computed at the point  $x$ . For example,  $q$  maps  $x^n$  into  $x^{n+1}$ ,  $\sin x$  into  $x \sin x$ , etc.

2. The operator that takes the derivative of a function  $\psi(x)$ . Denoting this operator by  $\frac{d}{dx}$ , we write:

$$\frac{d}{dx}\psi(x) = \frac{d\psi(x)}{dx}. \tag{2.94}$$

This means that  $\frac{d}{dx}$  maps  $x^n$  into  $nx^{n-1}$ ,  $\sin x$  into  $\cos x$ , etc.

Clearly, the operators  $q$  and  $\frac{d}{dx}$  are *linear*. They satisfy the analog of Eq. (2.30). Thus for  $\frac{d}{dx}$ , since the derivative of the sum of two functions is the sum of the derivatives, we have, for example,

$$\frac{d}{dx}(f + g) = \frac{d}{dx}f + \frac{d}{dx}g. \quad (2.95)$$

The operators  $q$  and  $\frac{d}{dx}$  are of primary importance in QM, since the first represents the *position* of a particle, and the second is proportional to its momentum. An important fact about these two operators, is that they *do not commute*. Let us denote by  $D$  the operator  $\frac{d}{dx}$ . For “any” function  $\psi(x)$ <sup>3</sup> we have:

$$\begin{aligned} q D\psi &= q \frac{d}{dx}\psi = x \frac{d\psi}{dx}, \\ D q\psi &= \frac{d}{dx} q\psi = \frac{d}{dx}(x\psi) = \psi + x \frac{d\psi}{dx}. \end{aligned} \quad (2.96)$$

Subtracting the two equations, we obtain, for “any”  $\psi$ ,  $Dq\psi - qD\psi = \psi$ , or:

$$Dq - qD = I. \quad (2.97)$$

Powers of the operators  $q$  and  $D$  are easy to compute. For example, the functions  $q^2\psi$  and  $D^2\psi$  are:

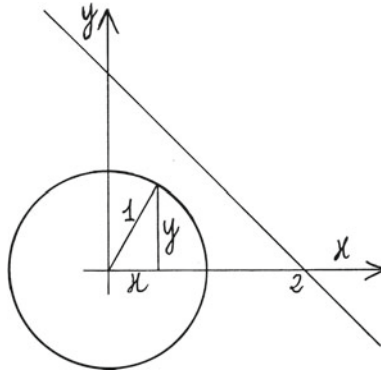
$$\begin{aligned} q^2\psi &= x(x\psi) = x^2\psi, \\ D^2\psi &= \frac{d}{dx} \frac{d}{dx}\psi = \frac{d^2}{dx^2}\psi. \end{aligned} \quad (2.98)$$

## 2.5 Complex Variables

To extend our repertoire of mathematical proficiency, this section will review some aspects of complex numbers and complex functions. (Our apologies to readers already well versed in this subject.) Mathematicians define an algebraic structure called a *field* as a set of (usually) numbers, along with two operations, which can be identified with addition and multiplication (subtraction and division are implicitly included), and satisfies the associative and distributive laws. The most commonly encountered fields are the real numbers, the rational numbers and, the subject of this section, the complex numbers. Complex analysis turns out to be mandatory for understanding the full mathematical structure of quantum mechanics. It is not strictly necessary for classical mechanics or electrodynamics, although complex variables can provide very useful enhancements to their mathematical formulation.

---

<sup>3</sup>The quotation marks refer to some mathematical conditions that the function  $\psi(x)$  must fulfill: in essence,  $\psi$  must be differentiable almost everywhere and  $D\psi$  must remain in the Hilbert space.



**Fig. 2.14** The circle  $\Gamma$  and the straight line  $\Sigma$  do not intersect; therefore their equations admit no simultaneous solution in the field of real numbers

To begin, consider a circle  $\Gamma$  in the Cartesian plane and a straight line  $\Sigma$  lying outside  $\Gamma$  (see Fig. 2.14), for example, the circle with center at the origin and radius 1, represented by the equation

$$x^2 + y^2 = 1, \quad (2.99)$$

and the straight line represented by the equation:

$$x + y = 2. \quad (2.100)$$

The simultaneous equations for the circle  $\Gamma$  and the line  $\Sigma$ , Eqs. (2.99) and (2.100), therefore do *not* have any real simultaneous solutions. Let us nevertheless solve (2.100) for  $y$ , to get  $y = 2 - x$ , and substitute this into (2.99). We obtain  $x^2 + (2 - x)^2 = 1$ ; therefore:

$$2x^2 - 4x + 3 = 0. \quad (2.101)$$

The solutions of the quadratic equation  $ax^2 + bx + c = 0$  are:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (2.102)$$

In our case  $a = 2$ ,  $b = -4$ , and  $c = 3$ . Thus

$$x = \frac{4 \pm \sqrt{16 - 24}}{4} = \frac{2 \pm \sqrt{-2}}{2}. \quad (2.103)$$

The square root of a negative number appears. This appears contradictory: since the square of a real number is always positive (for example  $(+2) \times (+2) = +4$  and  $(-2) \times (-2) = +4$ ), the argument of a square root should always be a positive

number. For many centuries, it was believed that roots of negative numbers have no meaning, consistent with the nonexistence of points common to a circle and a nonintersecting straight line. But in 1572, Rafael Bombelli in his book *L'Algebra*, gave meaning to the expression  $\sqrt{-1}$ .

We will denote  $\sqrt{-1}$  by the usual symbol  $i$  and call it the *imaginary unit*. Thus, by definition  $i^2 = -1$ , and (2.103) becomes  $x = \frac{2 \pm \sqrt{2}i}{2}$ . We are now dealing with a new kind of numbers, which we call *complex numbers*. If  $z = a + ib$ , with  $a$  and  $b$  real;  $a$  is called the *real part* and  $b$  the *imaginary part* of  $z$ . (In the above case  $a = 1$  and  $b = \pm \frac{\sqrt{2}}{2}$ ). Complex numbers have the following properties:

- (1)  $a + ib = c + id$  if and only if  $a = c$  and  $b = d$ .
- (2)  $(a + ib) + (c + id) = (a + c) + i(b + d)$ .
- (3)  $(a + ib)(c + id) = ac - bd + i(ad + bc)$ .

The multiplication law (3) is consistent with the usual properties of real numbers with the addition of a new rule:  $i^2 = -1$ . Furthermore, Items (1) and (2) suggest a representation of the complex number  $x + iy$  by the vector  $(x, y)$  in the Cartesian plane  $R^2$ . The  $x$ -axis now serves as the *real axis*, while the  $y$ -axis is the *imaginary axis*, since it consists of points of type  $(0, y)$ . The  $x$ - $y$  plane is now called the *complex plane* or an *Argand diagram*. Figure 2.15 shows the vector corresponding to the complex number  $z = a + ib$ , while Fig. 2.16 shows the vector corresponding to the complex number  $-z = -a - ib$ . Item (2) above implies that the parallelogram rule applies to the sum of two complex numbers (see Fig. 2.17).

A suggestive property is the following: if we multiply the complex number  $a + ib$  times the imaginary unit  $i$ , the corresponding vector is rotated by  $\frac{\pi}{2}$  (90 degrees). Indeed the complex number  $i(a + ib) = -b + ia$  corresponds to the vector  $(-b, a)$  which is rotated by  $\frac{\pi}{2}$  with respect to  $(a, b)$  (see Fig. 2.18). The “vectors”  $i, i^2 = -1, i^3 = -i, i^4 = +1$  are related by successive rotations by  $\frac{\pi}{2}$  (see Fig. 2.19). The *complex conjugate*  $z^*$  (alternatively written  $\bar{z}$  in many texts) of the number  $z = a + ib$

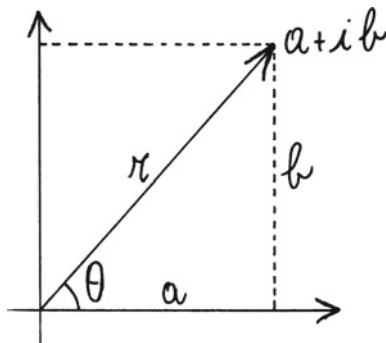
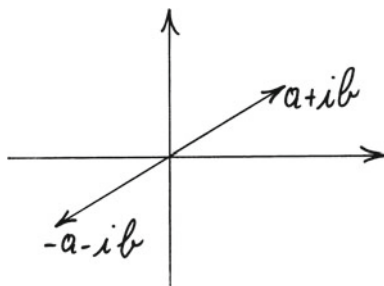
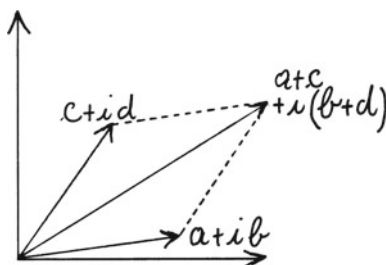


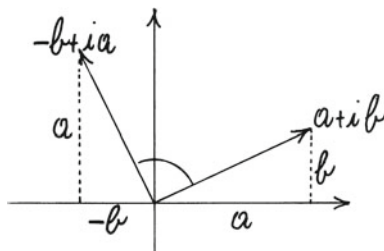
Fig. 2.15 Identification of the complex number  $a + ib$  with the point  $(a, b)$  in the Cartesian plane



**Fig. 2.16** Identification of the complex number  $-z = -a - ib$  with the point  $(-a, -b)$  of the Cartesian plane



**Fig. 2.17** The parallelogram's rule holds for the sum of two complex numbers  $a + ib$  and  $c + id$

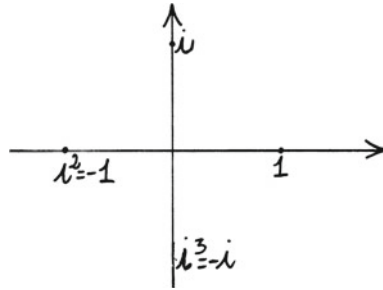


**Fig. 2.18** The complex number  $i(a + ib) = -b + ia$  corresponds to the vector  $(-b, a)$  which is rotated by 90 degrees with respect to  $(a, b)$

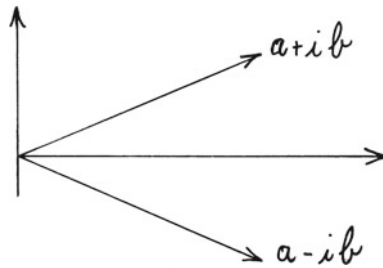
is defined as  $z^* = a - ib$ . Notice that  $a + ib$  and  $a - ib$  are symmetric with respect to reflection in the real axis (see Fig. 2.20). Clearly,  $(a - ib)^* = a + ib$ , so that  $(z^*)^* = z$ . If we multiply  $a + ib$  times its complex conjugate  $a - ib$ , we obtain the square of the length of the vector  $(a, b)$ ; indeed,

$$(a + ib)(a - ib) = a^2 - iab + iab + b^2 = a^2 + b^2. \quad (2.104)$$

The *modulus*  $r$  of the complex number  $a + ib$  is defined as its length,  $r = \sqrt{a^2 + b^2}$ . If  $\theta$  denotes the angle measured counterclockwise from the real axis to  $(a, b)$ , we



**Fig. 2.19** The complex numbers  $i, i^2, i^3, i^4$ , correspond, respectively, to the points  $(0, 1), (-1, 0), (0, -1), (1, 0)$ . This is a simple example of a cyclic group (designated  $\mathbb{Z}_4$ )



**Fig. 2.20** The complex conjugate  $a - ib$  of  $a + ib$  can be obtained by reflection through the real axis

have (see Fig. 2.15):

$$a = r \cos \theta, \quad b = r \sin \theta. \quad (2.105)$$

Thus

$$z = a + ib = r(\cos \theta + i \sin \theta). \quad (2.106)$$

The angle  $\theta$  is called *argument* of  $z$ . The real numbers  $r$  and  $\theta$  uniquely determine the complex number  $z$ . For example, if  $r = 1$  and  $\theta = \frac{\pi}{4}$  (45 degrees),  $z = \frac{1}{\sqrt{2}}(1 + i)$ . In general, complex numbers of modulus 1 are represented by points on the unit circle (with center at the origin and radius 1). For  $z = a + ib$ , the angle  $\theta$  is given by:

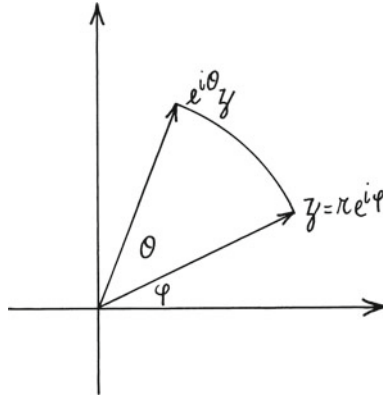
$$\theta = \arctan \frac{b}{a}, \quad (2.107)$$

with  $\theta$  determined up to multiples of  $2\pi$  (360 degrees). For example, the pair  $r, \theta$  and the pair  $r, \theta + 2\pi$  correspond to the *same* complex number.

The following very useful and beautiful formula can be used in place of (2.106):

$$z = a + ib = r e^{i\theta}, \quad (2.108)$$





**Fig. 2.21**  $e^{i\theta} z$  can be obtained from  $z$  by rotation by the angle  $\theta$

where  $e$  is Euler's constant,  $e = 2.7182818 \dots$  (the base of natural logarithms). The usual algebraic properties of the exponential function:  $e^0 = 1$ ,  $e^a e^b = e^{a+b}$ ,  $e^{-a} = 1/e^a$ , etc., still hold even when the exponent is imaginary or complex. However, the corresponding geometric representation is entirely different from the real case. A complex number  $e^{i\theta}$  can alternatively be regarded as an *operator*. Given the complex number  $z = r e^{i\phi}$ ,  $e^{i\theta} z = z e^{i\theta} e^{i\phi} = r e^{i(\theta+\phi)}$  is a number with the same modulus  $r$  but a new argument  $\phi + \theta$  (see Fig. 2.21). The factor  $e^{i\theta}$  “rotates”  $z$  by an angle  $\theta$ , similar to the way an orthogonal matrix, such as Eq. (2.28) rotates a two component vector.

## 2.6 Complex Vector Spaces and Dirac Notation

In order to make our development as simple as possible up to now, we have considered only *real* vector spaces, totally avoiding *complex numbers*. However, complex quantities turn out to be mandatory for complete understanding of the fundamental equations of QM, in particular, the Schrödinger equation itself. We have already noted that in QM, the operator  $D$  is proportional to the momentum  $p = mv$  of a particle. In fact, the proportionality factor is  $-i\hbar$ , where  $i = \sqrt{-1}$ , the *imaginary unit*. We will also, in this section, be introducing *Dirac's bra/ket notation*, invented in 1939 by P.A.M. Dirac (1958), one of the founding fathers of quantum mechanics. This is now a standard notation for describing quantum states, using angle brackets and vertical bars to represent abstract vectors and linear operations. The notation has also become popular in other mathematical applications.

We denote by  $\mathbb{C}^n$  the set of the ordered  $n$ -tuples  $(z_1, z_2, \dots, z_n)$  of *complex numbers*. An element  $z = (z_1, z_2, \dots, z_n)$  of such a set is now what we designate a *vector*. Thus, for  $n = 2$ , a vector  $(z_1, z_2) = (x_1 + iy_1, x_2 + iy_2)$  of  $\mathbb{C}^2$  is determined by 4 *real numbers*. In the following, unless explicitly stated, we will consider the

complex space  $\mathbb{C}^2$ . In Dirac notation, a vector  $z = (z_1, z_2)$  will be denoted by  $|z\rangle$ , in place of  $\mathbf{z}$ . The *null* vector  $|0\rangle$  is the vector with all components equal to zero. Therefore, in  $\mathbb{C}^n$ ,  $|0\rangle = (0, 0, \dots, 0)$ . Dirac denoted vectors, such as  $|\psi\rangle$ , representing quantum states, as “kets.” Adjoint vectors (associated with the complex conjugate of a wavefunction), such as  $\langle\phi|$ , were called “bras.” The product of a bra and a ket is a *bracket*, representing a scalar product  $\langle\phi|\psi\rangle$ . This connects with the notation we have already introduced for scalar products.

In  $\mathbb{C}^2$  space, all the familiar linear properties still apply. The linear combination  $z = a|u\rangle + b|v\rangle$  is, in general, constructed with *complex*  $a, b$ . We can visualize the sum  $|z\rangle = |u\rangle + |v\rangle$  as in Fig. 2.2, but to determine  $|z\rangle$  now requires 4 real numbers (although our physical space remains  $\mathbb{R}^3$ , not  $\mathbb{R}^4$ ). For example, if  $u_1 = 1 + i$ ,  $u_2 = 2 + 3i$ ,  $v_1 = 1 - 2i$ ,  $v_2 = 3 - 2i$ , then  $|u\rangle + |v\rangle = (2 - i, 5 + i)$ . The scalar product used in quantum mechanics is a generalization of the form of Eqs. (2.11) and (2.18), used in real spaces. Instead, an *Hermitian* scalar product is defined in  $\mathbb{C}^n$ , as follows:

**Definition 2.1** Given two vectors  $|z\rangle = (z_1, z_2, \dots, z_n)$ ,  $|w\rangle = (w_1, w_2, \dots, w_n)$ , the *Hermitian scalar product*  $\langle z|w\rangle$  (sometimes written  $\langle z|w\rangle_H$ ) is defined by:

$$\langle z|w\rangle = z_1^* w_1 + z_2^* w_2 + \dots \quad (2.109)$$

Clearly, if  $|z\rangle, |w\rangle$  have only real components,  $\langle z|w\rangle$  reduces to  $\mathbf{z} \cdot \mathbf{w}$ . The Hermitian scalar product is not necessarily symmetrical:  $\langle z|w\rangle$  is not, in general, equal to  $\langle w|z\rangle$ . Instead:

$$\langle v|u\rangle = \langle u|v\rangle^*. \quad (2.110)$$

The *norm* or length  $|u|$  of a vector  $|u\rangle$  is defined by a formula analogous to Eqs. (2.8) and (2.9):

$$|u| = +\sqrt{\langle u|u\rangle}. \quad (2.111)$$

It is still, of course, a nonnegative real number. It is also easy to verify that  $|u| = 0$ , if and only if  $|u\rangle = |0\rangle$ . The Hermitian scalar product  $\langle u|v\rangle$  is linear with respect to  $|v\rangle$ , but *antilinear* with respect to  $\langle u|$ , thus,

$$\langle u|cv\rangle = c\langle u|v\rangle, \quad \langle cu|v\rangle = c^* \langle u|v\rangle, \quad (2.112)$$

and

$$\langle u|v + w\rangle = \langle u|v\rangle + \langle u|w\rangle. \quad (2.113)$$

The notion of *orthonormal basis* can be extended. The vectors  $\mathbf{u}, \mathbf{v}$  form an orthonormal basis in  $\mathbb{C}^2$  if they are of unit length and orthogonal:

$$\langle u|u\rangle = \langle v|v\rangle = 1, \quad \langle u|v\rangle = 0. \quad (2.114)$$

The definition of the scalar product, as integrals over complex-valued functions, must be generalized from the form of Eq. (2.83). Using Dirac notation and extending the integration over the whole real axis, we now write

$$\langle f|g\rangle = \int_{-\infty}^{+\infty} f(x)^* g(x) dx. \quad (2.115)$$

If  $|f|$  and  $|g|$  are finite, the scalar product  $\langle f|g\rangle$  is also finite, as implied by the Cauchy–Schwarz inequality:

$$|f| |g| \geq |\langle f|g\rangle|. \quad (2.116)$$

In our toy Hilbert space  $\mathbb{R}^2$ , the Cauchy–Schwarz inequality follows simply from  $|\cos \phi| \leq 1$  in Eq. (2.10).

The generalization of matrices to the complex field is straightforward. All the matrix formulas in Sect. 2.2 remain valid, with the real numbers replaced by complex numbers. A very important operation for complex  $n \times n$  matrices, leading to the definition of the adjoint matrix, can be defined as follows:

**Definition 2.2** Given an operator  $A$ , represented in an orthonormal basis by the matrix  $A_{ij}$ , the operator whose matrix is obtained by taking the complex conjugate and interchanging rows and columns is called the *adjoint* of  $A$ , denoted by  $A^\dagger$ . Thus:

$$A_{ij}^\dagger = A_{ji}^* \quad (i, j = 1, 2, \dots, n). \quad (2.117)$$

The adjoint has the following properties:

$$\begin{aligned} (AB)^\dagger &= B^\dagger A^\dagger, \\ (A^\dagger)^\dagger &= A, \\ (A + B)^\dagger &= A^\dagger + B^\dagger, \\ (cA)^\dagger &= c^* A^\dagger, \end{aligned} \quad (2.118)$$

where  $c$  is a complex number. The *identity* is the operator  $I$  such that  $I|x\rangle = |x\rangle$  for any  $|x\rangle$ . The matrix representing the identity has all elements along the main diagonal equal to 1, and 0 everywhere else (the same as for real matrices).

The adjoint operation allows us to move an operator from one side of a scalar product to the other, by virtue of the following Theorem:

**Theorem 2.2** For every pair of vectors  $|x\rangle, |y\rangle$  and every operator  $A$  the following relations hold:

$$\langle y|Ax\rangle = \langle A^\dagger y|x\rangle, \quad \langle Ay|x\rangle = \langle y|A^\dagger x\rangle. \quad (2.119)$$

The proof of the first equality follows from the sequence of operations:

$$\begin{aligned} \langle A^\dagger y | x \rangle &= \langle x | A^\dagger y \rangle^* = \left( \sum_k x_k^* (A^\dagger y)_k \right)^* = \left( \sum_k x_k^* \sum_i A_{ki}^\dagger y_i \right)^* = \\ &= \left( \sum_k \sum_i x_k^* A_{ki}^* y_i \right)^* = \sum_{i,k} x_k A_{ik} y_i^* = \langle y | Ax \rangle. \end{aligned} \quad (2.120)$$

The second equality follows from the first, since  $(A^\dagger)^\dagger = A$ . The relations (2.119) are easy to verify explicitly for the case  $n = 2$ .

Recall that rotations leave the lengths of real vectors invariant. What are the corresponding linear operators that leave the lengths of complex vectors in  $\mathbb{C}^2$  or  $\mathbb{C}^n$  invariant? The answer is *unitary* operators.

**Definition 2.3** An operator  $U$  is called unitary if

$$U U^\dagger = U^\dagger U = I, \quad (2.121)$$

where, as usual,  $I$  denotes the identity.

A unitary operation leaves invariant the scalar product of two vectors  $|x\rangle$ ,  $|y\rangle$ , since, using (2.119), we have:

$$\langle Ux | Uy \rangle = \langle x | U^\dagger U y \rangle = \langle x | I y \rangle = \langle x | y \rangle. \quad (2.122)$$

Thus the norm (or length) of a vector is left invariant by a unitary transformation. The correct generalization of ordinary rotations of vectors (which are generated by orthogonal matrices) are thus unitary operations on complex vectors. Our tentative postulate was that dynamical variables are represented by symmetric matrices; more generally, these should be *Hermitian* matrices. A matrix  $H_{ij}$  ( $i, j = 1, 2, \dots, n$ ) is *Hermitian* (or *self-adjoint*) if  $H = H^\dagger$ , so that its elements are related by

$$H_{ij} = H_{ji}^* \quad (i, j = 1, 2, \dots, n). \quad (2.123)$$

For example, if  $n = 2$ ,  $H_{11}$  and  $H_{22}$  are real, and  $H_{21} = H_{12}^*$ . Thus, the matrix

$$\begin{pmatrix} 3 & 1+i \\ 1-i & 4 \end{pmatrix} \quad (2.124)$$

is Hermitian. If  $H$  is both Hermitian and real, it is again simply a symmetric matrix.

Hermitian matrices admit an orthonormal basis of eigenvectors, just like symmetric matrices in the real case. Furthermore, the eigenvalues of an Hermitian matrix  $H$  are *real*; indeed, taking the scalar product of both sides of the equation

$$H|x\rangle = \lambda|x\rangle \quad (2.125)$$

by the eigenvector  $|x\rangle$  we obtain:

$$\langle x|H|x\rangle = \lambda \langle x|x\rangle, \quad \lambda = \frac{\langle x|H|x\rangle}{\langle x|x\rangle}. \quad (2.126)$$

and by (2.110), we have:

$$\langle x|H|x\rangle^* = \langle x|H^\dagger|x\rangle = \langle x|H|x\rangle. \quad (2.127)$$

Thus, both numerator and denominator of (2.126) are real, therefore the eigenvalue  $\lambda$  must be real. The matrix (2.124) has the eigenvectors  $|u\rangle = \left(\frac{1+i}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$ ,  $|v\rangle = \left(\frac{1+i}{2}, 1\right)$ , thus

$$H|u\rangle = 2|u\rangle, \quad H|v\rangle = 5|v\rangle, \quad \langle u|v\rangle = 0. \quad (2.128)$$

Two Hermitian matrices  $A$ ,  $B$  admit a common basis of eigenvectors, if and only if they *commute*, that is, if  $[A, B] = 0$  or  $AB = BA$ . We omit the simple proof of these results for Hermitian matrices. We note that the fundamental interpretative postulate of QM (Postulate 1) is generalized with the simply substitution of “Hermitian” for “symmetric.”

Observables are represented in QM by Hermitian matrices (more precisely, linear Hermitian operators), and states of a physical system by real or complex vectors. Of course, the case of two component vectors or spinors is the simplest (see Chap. 4). However, as noted above, the Hilbert space of realistic physical systems is usually infinite dimensional.

## 2.7 Coordinates and Momenta in Quantum Mechanics

We have seen in Chap. 1 that de Broglie’s formula associates a “matter wave” with the rectilinear motion of a particle, with a wavelength  $\lambda = h/mv = h/p$ , where  $m$  is the mass,  $v$  the velocity and  $p$  the momentum of the particle. Accordingly, let us consider a very general instance of wave motion propagating in the  $x$ -direction. At a given instant of time, a periodic wave with wavelength  $\lambda$  might be represented by a function of the form

$$\psi(x) = f\left(\frac{2\pi x}{\lambda}\right), \quad (2.129)$$

where  $f(\theta)$  is most often a sinusoidal function such as  $\sin \theta$ ,  $\cos \theta$ ,  $e^{\pm i\theta}$ , or some linear combination of these. Each of these is a periodic function, its value repeating every time its argument increases by  $2\pi$ . This happens when  $x$  increases by one wavelength  $\lambda$ . The most useful form will turn out to be the complex exponential, which is related to the sine and cosine by Euler’s formulae  $e^{i\theta} = \cos \theta + i \sin \theta$ . We

consider the wavefunction

$$\psi(x) = e^{i2\pi x/\lambda}, \quad (2.130)$$

apart from an arbitrary multiplicative constant. The wavelength  $\lambda$  of this complex-valued wavefunction can be replaced by  $h/p$ , where  $p$  is the particle momentum, in accordance with the de Broglie formula. Thus,

$$\psi(x) = e^{i2\pi px/h} = e^{ipx/\hbar}, \quad (-\infty < x < \infty), \quad (2.131)$$

where  $\hbar \equiv h/2\pi$ . Since Planck's constant occurs in most formulas with the denominator  $2\pi$ , this symbol, pronounced “aitch-bar,” was introduced by Dirac in 1930.

Now that we have a mathematical representation of a matter wave, we should next try to find a “wave equation,” a differential equation which the wavefunction satisfies. As a first step let us apply the operator  $D = \frac{d}{dx}$  to Eq. (2.131). We find

$$\frac{d}{dx}\psi(x) = \frac{ip}{\hbar}\psi(x), \quad (2.132)$$

which can be rearranged to

$$-i\hbar \frac{d}{dx}\psi(x) = p\psi(x). \quad (2.133)$$

This can be recognized as an *eigenvalue equation* (see Eqs. 2.50 and 2.125) for the  $x$ -component of momentum  $p_x$ :

$$p_x\psi(x) = p\psi(x), \quad (2.134)$$

with the momentum operator evidently given by

$$p_x = -i\hbar \frac{d}{dx}. \quad (2.135)$$

This, incidentally, confirms our earlier speculation that the operator  $D = \frac{d}{dx}$  is proportional to the velocity  $v$  (hence the momentum  $p$ ) of a particle. In Dirac notation, the eigenvalue equation can be written

$$p_x|\psi\rangle = p|\psi\rangle. \quad (2.136)$$

Evidently, an eigenvalue  $p = mv$  for a free particle, can be *any* real number:  $-\infty < p < \infty$ . This is a *continuous spectrum* of eigenvalues, in contrast to the energy levels of a bound atom or molecule, which was a distinguishing feature in the early development of quantum theory. Actually, highly excited states of atoms or molecules, in which ionization or dissociation has occurred, also show a continuum of energy eigenvalues. The momentum eigenfunctions  $\psi(x)$  are complex-valued

(except when  $p = 0$ ). If  $\psi_p(x) = e^{ipx/\hbar}$  and  $\psi_{p'}(x) = e^{ip'x/\hbar}$  represent eigenstates with *different* eigenvalues,  $p$  and  $p'$ , respectively, then the corresponding eigenfunctions are orthogonal. This can be shown by an intuitive (although not entirely mathematically rigorous) argument:

$$\begin{aligned} \langle p|p' \rangle &= \int_{-\infty}^{+\infty} \psi_p(x)^* \psi_{p'}(x) dx = \int_{-\infty}^{+\infty} e^{i(p'-p)x/\hbar} dx = \\ &= \int_{-\infty}^{+\infty} \left( \cos[(p'-p)x/\hbar] + i \sin[(p'-p)x/\hbar] \right) dx = 0 \quad (p' \neq p). \end{aligned} \quad (2.137)$$

In the last line, the infinite number of positive and negative contributions to the sine or the cosine integrals cancel each other out to give a result of zero.

The Hilbert space  $\mathcal{L}^2(-\infty, +\infty)$  appropriate for QM is a set of *complex valued* functions  $\psi(x)$  such that the following integral is finite:

$$\int_{-\infty}^{+\infty} \psi(x)^* \psi(x) dx = \int_{-\infty}^{+\infty} |\psi(x)|^2 dx < \infty. \quad (2.138)$$

An apparent disaster occurs when we try to evaluate Eq.(2.138) using a momentum eigenfunction (2.131). With  $\psi(x) = e^{ipx/\hbar}$ , the complex conjugate is  $\psi(x)^* = e^{-ipx/\hbar}$ . Thus  $\psi(x)^* \psi(x) = |\psi(x)|^2 = 1$  and  $\int_{-\infty}^{+\infty} 1 dx = \infty$ , violating the condition for a valid Hilbert space. There are several ways that we can talk our way out of this difficulty.

(1) We might limit our consideration to quantum systems with *bound states*, for which wavefunctions conforming to (2.138) can always be found. This could be done, for example, by replacing the infinite domain  $-\infty < x < \infty$  by a finite interval  $-a \leq x \leq a$ . This excludes the free particle, Eq.(2.131), despite the fact that this system has been so fundamental in deriving some essential results in QM.

(2) We recognize that for a system in a momentum eigenstate, the probability density function  $|\psi(x)|^2 = 1$  for *all* values of  $x$ ,  $-\infty < x < \infty$  (even beyond the bounds of the known Universe!). This is in accord with the uncertainty principle, since a precisely known momentum  $p$  implies a completely indefinite position  $x$ . More realistically, a free particle should be described by a *wavepacket*, which is a superposition of momentum eigenstates  $\psi_p(x)$  of the form

$$\psi(x) = \int \phi(p) \psi_p(x) dp. \quad (2.139)$$

Then the integral (2.138) converges, provided that  $\int |\phi(p)|^2 dp$  is finite.

(3) Hilbert space is redefined to accommodate continuous spectra and divergent integrals. Dirac himself was aware that “the bra and ket vectors that we now use form a more general space than a Hilbert space.” A modern extension, known as

*rigged Hilbert space*<sup>4</sup> has the desired structure (the term “rigged” here implies “well-equipped and ready for action”). A “conventional” Hilbert space can accommodate a denumerably infinite number of basis vectors, labeled, for example, by  $n = 1, 2, 3, \dots$ . But, in a rigged Hilbert space, the number of basis vectors can be *nondenumerably infinite*, labeled, perhaps by indices with a continuum of allowed values, such as  $\nu$ , with  $-\infty < \nu < \infty$ .

Fortunately, we can carry on, using naive conventional Hilbert space, knowing that the mathematicians (however reluctantly) have us well covered regarding any inconsistencies or complications. As an illustrative example, consider the scalar product of eigenstates. For the discrete spectrum, we have

$$\langle m|n\rangle = \int_{-\infty}^{+\infty} \psi_m(x)^* \psi_n(x) dx = \delta_{m,n}, \quad (2.140)$$

where  $\delta_{m,n}$  is the Kronecker delta, equal to 1 for  $m = n$  and 0 for  $m \neq n$ . For eigenstates belonging to a continuous spectrum, we can write

$$\langle \mu|\nu\rangle = \int_{-\infty}^{+\infty} \psi_\mu(x)^* \psi_\nu(x) dx = \delta(\mu - \nu). \quad (2.141)$$

We have already seen that  $\langle \mu|\nu\rangle = 0$  for  $\mu \neq \nu$  (Eq.(2.137)). We have also found that  $\langle \nu|\nu\rangle = \infty$  for  $\mu = \nu$ . But Dirac here introduced a special kind of infinity, as represented by the *delta function*,  $\delta(\mu - \nu)$ .

The Dirac delta function was intended as the continuum analog of the Kronecker delta. It is, however, *not* a true function in the mathematical sense, but rather a *generalized function or distribution*. The delta function was regarded with much disdain by mathematicians, until a rigorous theory was proposed by Laurent Schwartz in 1950.<sup>5</sup> It is defined as a hypothetical “function” such that  $\delta(x - a) = 0$ , if  $x \neq a$ , and  $\delta(x - a) = \infty$ , if  $x = a$ . However, the infinite value is somewhat special: an integral over that singular point is presumed to equal 1. Intuitively, the delta function can be pictured as the limit of a distribution with integrated area 1, of infinitesimal width but very large height, centered around the point  $x = a$ . The delta function is presumed to satisfy the integral relations:  $\int_{-\infty}^{+\infty} \delta(x - a) dx = 1$  and  $\int_{-\infty}^{+\infty} f(x) \delta(x - a) dx = f(a)$ . To physicists, the delta function is invaluable for representing idealized objects such as point masses and point charges and for constructing Green’s functions.

The eigenvalues and eigenvectors of the position operator  $q$  are rather tricky, but, fortunately, rarely needed. A particle *localized* around a point  $x = a$  might be represented by a delta function:  $\psi(x) = \delta(x - a)$ . The relation  $x\delta(x - a) = a\delta(x - a)$  can then be interpreted as an eigenvalue equation  $x\psi(x) = a\psi(x)$ . As a consequence

<sup>4</sup>R de la Madrid (2005), *The role of the rigged Hilbert space in Quantum Mechanics*, Eur J Phys 26:287–312.

<sup>5</sup>Schwartz L (1950–51) *Théorie des distributions*, Hermann, Paris. See also: Lighthill MJ (1958) *An Introduction to Fourier Analysis and Generalised Functions*. Cambridge University Press.



of the uncertainty principle, a particle with an exactly known value of  $x$ , has a totally undetermined value of  $p_x$ , the possible values being spread over  $-\infty < p < \infty$ .

The commutators of components of position and momentum are of central importance in the formalism of QM. Let us first evaluate  $[x, p_x] = x p_x - p_x x$ , where  $p_x = -i\hbar \frac{d}{dx}$ . These operators have meaning only when applied to a function of  $x$ , say  $\phi(x)$ . Now,

$$[x, p_x]\phi(x) = x(-i\hbar) \frac{d}{dx} \phi(x) - (-i\hbar) \frac{d}{dx} [x\phi(x)] = i\hbar \frac{dx}{dx} \phi(x) = i\hbar \phi(x), \quad (2.142)$$

in which we have taken the derivative of a product  $x\psi(x)$  and simplified by cancellation. Since the function  $\phi(x)$  is arbitrary, we can abstract the operator relation

$$[x, p_x] = i\hbar. \quad (2.143)$$

Obvious analogs of Eq. (2.135) for the  $y$ - and  $z$ -components of momentum are

$$p_y = -i\hbar \frac{d}{dy}, \quad p_z = -i\hbar \frac{d}{dz}. \quad (2.144)$$

It is then simple to derive the analogous commutation relations:

$$[y, p_y] = i\hbar, \quad [z, p_z] = i\hbar. \quad (2.145)$$

Since  $x$  commutes with  $p_z$ , etc., and the different components of position commute with one another, as do the different components of momentum, we can collect the entire set of commutation relations in the following compact form:

$$[q_i, q_j] = [p_i, p_j] = 0, \quad [q_i, p_j] = i\hbar \delta_{i,j}, \quad (2.146)$$

where  $i, j = 1, 2, 3$  and  $q_1 = x$ ,  $p_1 = p_x$ , etc.

## 2.8 Heisenberg Uncertainty Principle

The Heisenberg uncertainty principle is a fundamental consequence of the noncommutativity of position and momentum operators. Let us consider the one-dimensional case with  $x$  and  $p_x$ , which we write simply as  $p$ . The average, mean or *expectation variable* of a dynamical variable  $A$  in a quantum state  $|\Psi\rangle$  is written

$$\bar{A} = \langle \Psi | A | \Psi \rangle. \quad (2.147)$$

The mean square deviation from the mean is then given by

$$(\Delta A)^2 = \langle \Psi | (A - \bar{A})^2 | \Psi \rangle, \quad (2.148)$$

where  $\Delta A$  is the *root mean square*, which is designated as the *uncertainty* in  $A$ . Now define two functions

$$f = (x - \bar{x})\Psi, \quad \text{and} \quad g = i(p - \bar{p})\Psi. \quad (2.149)$$

We then find

$$\langle f | f \rangle = \langle (x - \bar{x})\Psi | (x - \bar{x})\Psi \rangle = \langle \Psi | (x - \bar{x})^2 | \Psi \rangle = (\Delta x)^2, \quad (2.150)$$

and, analogously,

$$\langle g | g \rangle = \langle i(p - \bar{p})\Psi | i(p - \bar{p})\Psi \rangle = \langle \Psi | (p - \bar{p})^2 | \Psi \rangle = (\Delta p)^2. \quad (2.151)$$

Next, let us evaluate:

$$\langle f | g \rangle + \langle g | f \rangle = \langle \Psi | (x - \bar{x})i(p - \bar{p}) - i(p - \bar{p})(x - \bar{x}) | \Psi \rangle. \quad (2.152)$$

After some cancelation, we find

$$\langle f | g \rangle + \langle g | f \rangle = i \langle \Psi | [x, p] | \Psi \rangle = -\hbar, \quad (2.153)$$

recalling the commutation relation  $[x, p] = i\hbar$ . From the Cauchy–Schwarz inequality, Eq. (2.116),

$$\sqrt{\langle f | f \rangle} \sqrt{\langle g | g \rangle} \geq |\langle f | g \rangle|. \quad (2.154)$$

Finally, we arrive at the Heisenberg uncertainty principle:

$$\Delta x \Delta p \geq \frac{\hbar}{2}. \quad (2.155)$$

This implies that exact values of a position variable and its conjugate momentum cannot be simultaneously known. Thus the trajectories of Bohr orbits are illusory: quantum behavior is not deterministic with regard to classical variables.

Twenty-First Century Quantum Mechanics: Hilbert Space  
to Quantum Computers

Mathematical Methods and Conceptual Foundations

Fano, G.; Blinder, S.M.

2017, XVI, 271 p. 142 illus., 12 illus. in color., Hardcover

ISBN: 978-3-319-58731-8