# Linear Vectors Spaces

**Edouard Marchais**[a]

[a] *EPITA, 14-16 Rue Voltaire, 94270 Le Kremlin-Bicêtre, France.*

*E-mail:* edouard.marchais@epita.fr

Abstract: In these notes the notion of linear vector spaces is introduced. Complex vectors spaces of finite dimension, as well as operators in these spaces, are discussed in detail using Dirac's notation in a systematic manner.

# Contents

# 1 Definition of a linear vector space

Let us consider a set $S$ of certain abstract objects, represented by the symbol $|\rangle$[1]; in order to distinguish these objects, we provide them with labels.

**Example 1** *For instance, some examples of* $|\rangle$ *are* $|a\rangle$, $|3\rangle$ *and* $|\alpha\beta\rangle$.

Having introduced these objects, we must define « rules of manipulation » or of « composition» as one calls them, of these objects, i.e., their algebra. This is similar to, say, the real numbers. One can introduce the set of real numbers, but unless one also specifies rules of addition and multiplication, one has done little more than distribute names. It is up to us to define the rules of algebra, but we must require that these rules be unambiguous.

The first of these operations to be defined and which in analogy to the case of real numbers we call addition of $|\rangle$, allows us to construct from any two $|\rangle$ a third $|\rangle$, which is called the sum of the first two $|\rangle$. In order to indicate that the particular object $|c\rangle$ is the sum of the particular objects $|a\rangle$ and $|b\rangle$, we write

$$|c\rangle = |a\rangle + |b\rangle \tag{1.1}$$

The second operation to be defined, which we call « multiplication of a $|\rangle$ by a number» , allows us to construct from any complex number and any $|\rangle$ another $|\rangle$. The equation

$$|c\rangle = \alpha \cdot |b\rangle \tag{1.2}$$

will mean that $|c\rangle$ is the product of $|b\rangle$ by the complex number $\alpha$[2]. We have now defined operations of addition and multiplication for a general set of objects $|\rangle$. We will, however, deal only with special sets of objects that have the following properties:

**Properties 1** *We have*

1. *If* $|a\rangle, |b\rangle \in S$, *then* $(|a\rangle + |b\rangle) \in S$
2. *If* $|a\rangle \in S$ *and* $\alpha$ *is a complex number, then* $(\alpha|a\rangle) \in S$
3. *There exists a null element* $|0\rangle \in S$ *such that for any* $|a\rangle \in S$ *one has* $|a\rangle + |0\rangle = |a\rangle$
4. *For any* $|a\rangle \in S$ *there exists an element* $|a'\rangle \in S$ *such that* $|a\rangle + |a'\rangle = |0\rangle$

So far we have said that there exists certain operations, called addition and multiplication, but we have not yet specified the properties of these operations. The following properties 1, 2 and 3 will ensure that addition and multiplication are well-defined operations.

**Properties 2** *For any* $|a\rangle, |b\rangle, |c\rangle \in S$ *and for any complex numbers* $\alpha$ *and* $\beta$ *one has*

1. $|a\rangle + |b\rangle = |b\rangle + |a\rangle$; *(commutative law of addition).*
   $(|a\rangle + |b\rangle) + |c\rangle = |a\rangle + (|b\rangle + |c\rangle)$; *(associative law of addition).*
2. $1 \cdot |a\rangle = |a\rangle$.
3. $\alpha \cdot (\beta \cdot |a\rangle) = (a \cdot \beta) \cdot |a\rangle$; *(associative law of multiplication).*
   $(\alpha + \beta)|a\rangle = \alpha|a\rangle + \beta|a\rangle$; *(distributive law with respect to the addition of complex numbers).*
   $\alpha(|a\rangle + |b\rangle) = \alpha|a\rangle + a|b\rangle$; *(distributive law with respect to the addition of* $|\rangle$*).*

[1] The notation is due to P.A.M Dirac
[2] The dot will often be omitted

A set $S$ of $|\rangle$ that has the properties 1 and **??** is called a *linear vector space*. The elements of this set, $|\rangle$ are called *vectors*.

**Example 2** *Suppose that the set $S$ of objects $|\rangle$ is the set of all complex numbers. Then the list of* **Properties 2** *simply contains the well-known rules of arithmetic for complex numbers. This example justifies the naming of the abstract operations (1.1) and (1.2) as addition and multiplication.*

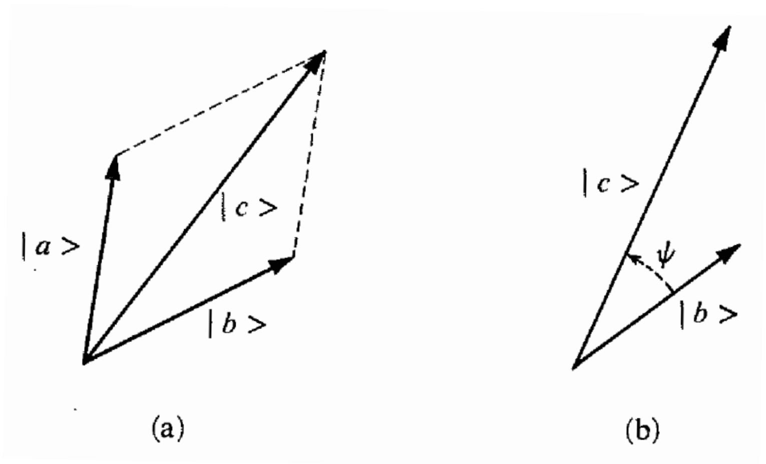


**Figure 1**: Geometrical illustration of (1.1) and (1.2) in the particular case of **Example 3** with (a) $|c\rangle = |a\rangle + |b\rangle$. (b) $|c\rangle = re^{i\psi}|b\rangle$.

**Example 3** *Suppose that the set $S$ consists of all the arrows lying in a plane, including the « arrow » of zero length. For the rule of addition of $|\rangle$, we take the familiar geometrical rule of the addition of arrows, as illustrated in* **Figure 1(a)**. *We can verify that this rule obeys all the conditions enumerated in* **Properties 2**.

*The multiplication of a $|\rangle$ by the number $z = re^{i\psi}$ ($r$ and $\psi$ being real) will be defined as the elongation of the arrow $r$ times and its subsequent rotation by the angle $\psi$ (as shown in* **Figure 1(b)**). *When $z$ is real, this rule reduces to the conventional one of multiplying arrows by numbers.*

Comparing these properties of the arrows with the **Properties 1**, we see that the set of all arrows constitutes a linear vector space. For example, the addition (as defined above) of two arrows is an arrow, and the multiplication of an arrow by a complex number is another arrow, etc.

Starting from **Properties 1, ??**, one can easily demonstrate that a linear vector space contains only one null vector $|0\rangle$ and that to each vector $|a\rangle$ there corresponds one and only one vector $|a'\rangle$ satisfying $|a\rangle + |a'\rangle = |0\rangle$.

We now verify that any vector $|a\rangle$ multiplied by the number 0 gives $|0\rangle$. We have

$$|a\rangle = 1 \cdot |a\rangle = (0+1)|a\rangle = 0 \cdot |a\rangle + 1 \cdot |a\rangle = 0 \cdot |a\rangle + |a\rangle \qquad \text{hence} \qquad |a\rangle = 0 \cdot |a\rangle + |a\rangle \tag{1.3}$$

Let $|a'\rangle$ be the vector satisfying $|a'\rangle + |a\rangle = |0\rangle$. Then

$$|0\rangle = |a\rangle + |a'\rangle = (0 \cdot |a\rangle + |a\rangle) + |a'\rangle = 0 \cdot |a\rangle + (|a\rangle + |a'\rangle) = 0 \cdot |a\rangle + |0\rangle = 0 \cdot |a\rangle \tag{1.4}$$

or briefly

$$0 \cdot |a\rangle = |0\rangle \quad \text{for any } |a\rangle \in S \tag{1.5}$$

Because of (1.5), no ambiguity will result when, for simplicity, we write briefly 0 instead of $|0\rangle$. It is easy to define subtraction of vectors

$$|a\rangle - |b\rangle = |a\rangle + (-1)|b\rangle \qquad \text{then (of course)} \qquad |a\rangle - |a\rangle = |a\rangle + (-1)|a\rangle = (1-1)|a\rangle = 0 \tag{1.6}$$

## 2 The scalar product

Suppose one has established a rule that associates with any pair of vectors $|b\rangle \in S$ and $|a\rangle \in S$ a certain complex number; we shall denote[3] it by $\langle b|a\rangle$ and call it the *scalar product* of $|b\rangle$ with $|a\rangle$. The properties of the scalar product will be, by definition, the following :

**Properties 3** *We have*

1. $\langle b|a\rangle = \overline{\langle a|b\rangle}$
2. *If* $|d\rangle = \langle c|a\rangle + |b\rangle$ *then* $\langle c|b\rangle = \alpha\langle c|a\rangle + \beta\langle c|b\rangle$
3. $\langle a|a\rangle \geq 0$, *the equality sign appears only when* $|a\rangle = 0$.

Note that because of the property 1, the number $\langle a|a\rangle$ is real. This is an important property, which will enable us to regard $\sqrt{\langle a|a\rangle}$ as the « length » of the vector $|a\rangle$.

From the property 1, we see that in general the scalar product of $|b\rangle$ with $|a\rangle$ is not the same as the scalar product of $|a\rangle$ with $|b\rangle$, since

$$\langle b|a\rangle = \overline{\langle a|b\rangle} \neq \langle a|b\rangle \tag{2.1}$$

Two vectors $|a\rangle$ and $|b\rangle$ are said to be *orthogonal* to each other if their scalar product vanishes

$$\langle a|b\rangle = \langle b|a\rangle = 0 \tag{2.2}$$

The property 3 implies that if a vector $|a\rangle \in S$ is orthogonal to every vector of $S$

$$\langle a|\,\rangle = 0 \qquad \text{hence for all } |\,\rangle \in S \tag{2.3}$$

then $|a\rangle = 0$, since from (2.3) one has in particular $\langle a|a\rangle = 0$.

## 3 Dual vectors and the Cauchy-Schwarz inequality

The form of the property 3 of the preceding section introduces an apparent asymmetry between the vectors $|c\rangle$ and $|d\rangle$ which enter into the scalar product $\langle c|d\rangle$. The meaning of property 2 is that the scalar product $\langle c|d\rangle$ depends linearly upon the vector $|d\rangle$ in the sense that if we set

$$|d\rangle = \alpha|a\rangle + \beta|b\rangle \tag{3.1}$$

then

$$\langle c|d\rangle = \alpha\langle c|a\rangle + \beta\langle c|b\rangle \tag{3.2}$$

[3]* A « closed bracket » expression $\langle\,|\,\rangle$ will, by convention, always denote a number (complex in general) and not a vector.

is a linear function of $\alpha$ and $\beta$. However, if we set $|c\rangle = \alpha|a\rangle + \beta|b\rangle$. then

$$\langle c|d\rangle = \overline{\langle d|c\rangle} = \overline{[\alpha\langle d|a\rangle + \beta\langle d|b\rangle]} = \overline{\alpha}\langle a|d\rangle + \overline{\beta}\langle b|d\rangle \tag{3.3}$$

is no longer a linear function of $\alpha$ and $\beta$, since it depends linearly on $\alpha$ and $\beta$[4].

To remove this asymetry, it is convenient to introduce, besides the vectors $|\,\rangle$, other vectors belonging to a *different* space and which will be denoted by the symbol $\langle\,|$. We shall assume that there is a one-to-one correspondence between vectors $|\,\rangle$ and vectors $\langle\,|$. A pair of vectors in which each is in correspondence with the other will be called a *pair of dual vectors*, and such pairs will always carry the same identification label. Thus, e.g., $\langle b|$ is the dual vector of $|b\rangle$. We now define the multiplication of vectors $|\,\rangle$ by vectors $\langle\,|$ by requiring the following.

**Properties 4** *We have*

1. *The product of $\langle b|$ with $|a\rangle$ is identified with the scalar product*

$$\langle b|\cdot|a\rangle \equiv \langle b|a\rangle \tag{3.4}$$

2. *The scalar product $\langle c|d\rangle$ depends linearly on $\langle c|$*

$$[\langle a|\alpha + \langle b|\beta]\cdot|d\rangle = \alpha\langle a|d\rangle + \beta\langle b|d\rangle \tag{3.5}$$

From properties 1 and 2 we have

$$\langle c|\cdot[\alpha|a\rangle + \beta|b\rangle] = \alpha\langle c|a\rangle + \beta\langle c|d\rangle \tag{3.6}$$

Thus, the vectors $\langle\,|$ and $|\,\rangle$ play a symmetrical role in the scalar product. Setting

$$\langle c| = \langle a|\overline{\alpha} + \langle b|\overline{\beta} \tag{3.7}$$

we have

$$\langle c|d\rangle = \overline{\alpha}\langle a|d\rangle + \overline{\beta}\langle b|d\rangle \tag{3.8}$$

Comparing (3.8) with (3.7), we see that $\langle a|\overline{\alpha} + \langle b|\overline{\beta}$ is the dual vector of $\alpha|a\rangle + \beta|b\rangle$; hence, the rule for obtaining a dual of a linear combination of vectors $|\,\rangle$ is to replace the vectors by their duals and the coefficients by their complex conjugates. The reason why the scalar product is symmetric with respect to vectors $\langle\,|$ and $|\,\rangle$ is now apparent; we have, so to say, included the complex conjugation in the definition of the vectors $\langle\,|$.

The advantage of considering $\langle\,|\,\rangle$ as a product of $\langle\,|$ with $|\,\rangle$ is that now a simple distributive law of multiplication holds for the vectors $\langle\,|$ as well as for the vectors $|\,\rangle$.

The manner in which we introduced dual vectors, namely, as a device for simplifying the notation, is neither very rigorous nor the most general, although it is quite sufficient for our purpose. The interested reader can find the general definition sketched below.

Let $f$ be a function defined in $S$ by a rule that associates with every vector $|x\rangle \in S$ a complex number $f(|x\rangle)$; such a function is usually called a *functional*. The functional $f$ is linear if

$$f(\alpha|x\rangle + \beta|y\rangle) = \alpha f(|x\rangle) + \beta f(|y\rangle) \tag{3.9}$$

[4] One says that $\langle c|d\rangle$ is linear with respect to $|d\rangle$ but antilinear with respect to $|c\rangle$.

The set of all linear functionals in $S$ forms a linear vector space, for adding two linear functionals and multiplying a linear functional by a number results again in a linear functional. The space of all linear functionals in $S$ is called the *dual space of* $S$.

Suppose that the scalar product has been defined in $S$ and consider all the functionals of the particular type

$$f(|x\rangle) = \langle f|x\rangle \qquad \text{with} \qquad |x\rangle \in S \tag{3.10}$$

Owing to the linearity of the scalar product, these are linear functionals, and it is clear that they form a linear vector space. This space is just the space of vectors $\langle\,|$; the use of vectors $\langle\,|$ corresponds to using the notation

$$f \equiv \langle f| \tag{3.11}$$

which should be understood in the sense that attaching an argument $|x\rangle$ to $f$ is equivalent to « multiplying » $\langle f|$ by $|x\rangle$. The dual vector of $|f\rangle$ is $\langle f|$.

Consider now the vector

$$|c\rangle = |a\rangle - x\langle b|a\rangle|b\rangle \tag{3.12}$$

with real $x$. Since

$$\langle c|c\rangle \geq 0 \tag{3.13}$$

we have

$$x^2\langle b|a\langle a|b\rangle\langle b|b\rangle - 2x\langle b|a\rangle\langle a|b\rangle + \langle a|a\rangle \geq 0 \tag{3.14}$$

Inequality (3.14) implies that the above quadratic equation in $x$ with real coefficients has either a double real root or no real roots. Therefore

$$\langle a|a\rangle\langle b|b\rangle \geq \langle b|a\rangle\langle a|b\rangle = |\langle b|a\rangle|^2 \tag{3.15}$$

or

$$\sqrt{\langle a|a\rangle} \cdot \sqrt{\langle b|b\rangle} \geq |\langle b|a\rangle| \tag{3.16}$$

Inequality (3.16) is known as the Cauchy-Schwarz inequality.

# 4 Real and complex vector spaces

Until now the word « number » has been understood in the sense of a complex number. A real number is, however, a special case of a complex number. It is obvious, therefore, that one can repeat all the considerations of the previous sections, restricting ourselves to real numbers exclusively. The only difference would be that complex conjugation would become a redundant operation and consequently would never appear. For instance, the scalar product would be symmetric, since we would have the relation

$$\langle a|b\rangle = \langle b|a\rangle \tag{4.1}$$

instead of the more general relation

$$\langle a|b\rangle = \overline{\langle b|a\rangle} \tag{4.2}$$

In the light of these remarks, one can speak of « real » and « complex » vector spaces.

A simple example of a real vector space is the set of arrows lying in a plane. Here the addition of arrows is to be understood in the usual sense of the parallelogram rule, and the multiplication of an arrow by a number xis to be understood as the elongation of the vector $x$ times. This is not the same as assuming that the arrows belong to a complex vector space, as in Example 3 of section 1, where the multiplication of an arrow by a complex number was defined; it not only elongated the arrow, but also rotated it and brought it onto another arrow. This double operation was possible because a complex number contains two real parameters. The importance of this difference will be properly understood after we have introduced the notion of a dimension of the space.

The reader who feels ill at ease with some of the abstract notions of complex vector spaces introduced in this text is advised to think in terms of arrows in a plane or in space. This should help his intuitive understanding of the more abstract case.

**Example 4** *Consider arrows in a plane. As we mentioned above, they form a real vector space. The scalar product of two vectors $|a\rangle$ and $|b\rangle$ will be defined conventionally as the product of the lengths of the corresponding arrows by the cosine of the angle between them*

$$\langle a|b\rangle\langle b|a\rangle \equiv \vec{a}\cdot\vec{b} = |\vec{a}|\,|\vec{b}|\,\cos\psi_{a,b} \tag{4.3}$$

*The reader can verify that this definition of the scalar product is consistent with C. For example*

$$\langle a|[|b\rangle + |c\rangle] = \vec{a}\cdot(\vec{b}+\vec{c}) = |\vec{a}|\,|\vec{b}+\vec{c}|\,\cos\psi_{a,b+c} \tag{4.4}$$

*On the other hand, using the well-known trigonometrical relations between the sines of angles and the sides of a triangle, after some calculations we get*

$$\begin{aligned}
\langle a|b\rangle + \langle a|c\rangle &= |\vec{a}|\,|\vec{b}+\vec{c}|\left\{\frac{|\vec{b}|}{|\vec{b}+\vec{c}|}\cos\psi_{a,b} + \frac{|\vec{c}|}{|\vec{b}+\vec{c}|}\cos\psi_{a,c}\right\}\\
&= |\vec{a}|\,|\vec{b}+\vec{c}|\left\{\frac{\sin\psi_{a,b+c}}{\sin\psi_{b,c}}\cos\psi_{a,b} + \frac{\sin\psi_{b,b+c}}{\sin\psi_{b,c}}\cos\psi_{a,c}\right\}\\
&= |\vec{a}|\,|\vec{b}+\vec{c}|\,\cos\psi_{a,b+c}
\end{aligned}$$

Thus

$$\langle a|[|b\rangle + |c\rangle] = \langle a|b\rangle + \langle a|c\rangle \tag{4.5}$$

as required by property 2 in section 2. From the definition (4.3) it is evident that the orthogonality of two vectors means that the corresponding arrows are perpendicular.

# 5 Metric spaces

**Definition 1** *A set $R$ is called a metric space if a real, positive number $\rho(a,b)$ is associated with any pair of its elements $a, b \in R$ and if*

1. $\rho(a,b) = \rho(b,a)$
2. $\rho(a,b) = 0$ *only when* $a = b$
3. $\rho(a,b) + \rho(b,c) \geq \rho(a,c)$

The number $\rho(a,b)$ is called the distance between $a$ and $b$. Conditions 1 and 2 simply mean that the distance from $a$ to $b$ is the same as that from $b$ to $a$, and that the distance vanishes only when two elements coincide. The condition 3 is known as the *triangle inequality*.

**Example 5** *Any set of points on a plane is a metric space if $\rho(a,b)$ is identified with the « ordinary » distance between the points $a$ and $b$. The condition 3 is then the familiar statement that the sum of the lengths of two sides of a triangle is not smaller than the length of the third side of this triangle.*

This notion of a distance between elements of a set is now extended to the case where the set constitutes a linear vector space. First, a few comments are in order. The scalar product of $|a\rangle$ with its dual vector $\langle a|$ is, by the very definition of the scalar product, a positive number. $\langle a|a\rangle$ is called the norm, or the length, of the vector $|a\rangle$.

**Example 6** *In the elementary case of arrows in a plane (example of section 4), the norm is simply the length of the arrow*

$$\sqrt{\langle a|a\rangle} = |\vec{a}| \tag{5.1}$$

In elementary vector calculus one considers a vector as an arrow that joins two points of the space; each vector has its origin and its end. In the general theory of linear vector spaces it is also often very helpful to consider vectors as having a common origin and extending out from that origin. Each vector may then be considered as a « radius vector » which defines a point (the « end » of the vector) in the space. It must, however, be kept very clearly in mind that this notion of a point in space is introduced only as a pictorial representation of a vector, and that it never enters in a fundamental way into the theory of linear vector spaces. In fact, in defining a linear vector space (section 1), the idea of a point was not even mentioned.

Let us define the distance between two vectors $|a\rangle$ and $|b\rangle$ (or, if we wish, the distance between the points that they determine) as the norm of the vector $\{|a\rangle - |b\rangle\}$. We do this in analogy to elementary vector calculus, where the distance between two points is defined as the length of the vector joining the ends of the respective radius vectors or, equivalently (according to the rule of vectors addition) the length of the difference between two vectors.

We show that the distance so defined satisfies the triangle inequality. Let

$$|3\rangle = |1\rangle + |2\rangle \tag{5.2}$$

We have

$$\langle 3|3\rangle = \langle 1|1\rangle + \langle 2|2\rangle + 2\,\mathsf{Re}\langle 1|2\rangle \le \langle 1|1\rangle + \langle 2|2\rangle + 2\,|\langle 1|2\rangle| \tag{5.3}$$

Using the Cauchy-Schwarz inequality, we get

$$\langle 3|3\rangle \le \langle 1|1\rangle + \langle 2|2\rangle + 2\sqrt{\langle 1|1\rangle\langle 2|2\rangle} = \left(\sqrt{\langle 1|1\rangle} + \sqrt{\langle 2|2\rangle}\right)^2 \tag{5.4}$$

Thus

$$\langle 3|3\rangle \le \sqrt{\langle 1|1\rangle} + \sqrt{\langle 2|2\rangle} \tag{5.5}$$

Putting

$$\begin{aligned}|1\rangle &= |a\rangle - |b\rangle\\ |2\rangle &= |b\rangle - |c\rangle\\ |3\rangle &= |a\rangle - |c\rangle\end{aligned}$$

we recognize in (5.5) the triangle inequality.

It is evident that the norm of $\{|a\rangle - |b\rangle\}$ satisfies conditions 1 and 2. *Therefore, a linear vector space in which there is defined a scalar product is a metric space.*

It should be borne in mind, however, that a linear vector space is not necessarily a metric space. The reader will notice in what follows that there exist properties of linear vector spaces which can be discussed whether or not a scalar product in the space has been defined.

Consider now an infinite sequence of elements of a metric space: $a_{(1)}, a_{(2)}, \ldots, a_{(k)}, \ldots$. Suppose there exists an element of the space such that the distances $\rho\left(a, a_{(k)}\right)$ $(k = 1, 2, \ldots, n, \ldots)$ between the members of the sequence become smaller and smaller as $k$ increases and in the limit as $k \to \infty$ tend to 0

$$\lim_{k\to\infty} \rho\left(a, a_{(k)}\right) = 0 \tag{5.6}$$

We prove that $a$ is unique. In fact, suppose that besides (5.6), one also has

$$\lim_{k\to\infty} \rho\left(b, a_{(k)}\right) = 0 \tag{5.7}$$

Then by virtue of the triangle inequality

$$\rho(a, b) \leq \rho\left(a, a_{(k)}\right) + \rho\left(b, a_{(k)}\right) \tag{5.8}$$

and since both members of the RHS of (5.8) tend to zero as $k \to \infty$, one must have

$$\rho(a, b) = 0 \tag{5.9}$$

This result is based only on the fundamental properties of metric spaces. Therefore, it remains valid in the case of a linear vector space in which there is defined a scalar product. Thus, if a sequence of vectors $|a_{(1)}\rangle, |a_{(2)}\rangle, \ldots, |a_{(k)}\rangle, \ldots$ converges to some vector $|a\rangle$ in the sense that

$$\rho\left(|a\rangle, |a_{(k)}\rangle\right) = \left[\langle a| - \langle a_{(k)}|\right]\left[|a\rangle - |a_{(k)}\rangle\right] \to 0 \tag{5.10}$$

then $|a\rangle$ is unique.

# 6 Linear operators

An example of a function is a rule that associates with a number $x$ another number, say $y$. Let the rule of association be represented by the symbol $f(\,)$. Thus $f(\,)$ associates the number $y = f(x)$ in a particular way with the number $x$. One can also define a function of a *vector argument* $|x\rangle$. In this case, one writes for the function $f(|x\rangle)$. As an example, if one lets

$$f(|x\rangle) \equiv \langle a|x\rangle \tag{6.1}$$

then, (6.1) defines a rule, $f(\,)$, which associates with a vector $|x\rangle$ a number $\langle a|x\rangle$. We can generalize still further the notion of a function and introduce the notion of a *vector function* of a *vector argument*. Thus

$$|f(|x\rangle)\rangle \tag{6.2}$$

defines a rule that associates with the vector $\langle a|x\rangle$ the vector function $|f(|x\rangle)\rangle$. The simplest example of this rule is provided by the multiplication of $|x\rangle$ by a number c.

$$|f(|x\rangle)\rangle \equiv c|x\rangle \tag{6.3}$$

This example suggests a particular notation. We shall say that $|f(|x\rangle)\rangle$ results from the multiplication of $|x\rangle$ by an object called an *operator*. Accordingly, we write $F|x\rangle$ instead of $|f(|x\rangle)\rangle$, where $F$ is an operator. Then $F$ defines a rule that associates with a vector $|x\rangle$ another vector $F|x\rangle$. Of course, when we say that $F|x\rangle$ is a vector, we tacitly assume that $F|x\rangle$ belongs to some linear vector space, provided $|x\rangle$ is such a vector that $F|x\rangle$ is meaningful. In the following discussion, we shall use capital italic letters to denote operators. We shall be interested in a rather special class of operators, the linear ones, which are defined as follows :

**Definition 2** *The operator $A$ is a linear*[5] *operator if*

$$A\{\alpha|a\rangle + \beta|b\rangle\} = \alpha\{A|a\rangle\} + \beta\{A|b\rangle\} \tag{6.4}$$

To simplify the writing, we shall assume that, given an operator $A$, the expression $A|\,\rangle$ is meaningful for any $|\,\rangle \in S$ and, moreover, that $\{A|\,\rangle\} \in S$.

It is well known that a function $f(x)$ may not be defined for all values of its argument $x$. Similarly, the vector function $|f(|x\rangle)\rangle$ of the vector argument $|x\rangle$ may not be defined for all vectors $|x\rangle$. The set of vectors $|x\rangle$, for which $|f(|x\rangle)\rangle = F|x\rangle$ is defined, is called the *domain* of the operator $F$.

In general, the vector $F|x\rangle$ will not belong to $S$, but to some other vector space. The totality of vectors $F|x\rangle$ obtained by letting $F$ operate on all vectors of its domain is called the *range* of $F$.

Thus, our assumption means that the domain of any operator that we consider is identical to the space $S$ itself and that the range of the operator is included in $S$.

The first assumption can be abandoned when we discuss the so-called linear differential operators. As for the second condition, it will always be possible to sufficiently enlarge the space $S$ so as to satisfy it.

The operator associated with the function

$$|f(|x\rangle)\rangle = |x\rangle \tag{6.5}$$

is called the *identity*, or *unit, operator*, and will be denoted by $E$.

$$E|\,\rangle = |\,\rangle \qquad \text{for any } |\,\rangle \tag{6.6}$$

# 7 The algebra of linear operators

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

$$A|\,\rangle = B|\,\rangle \qquad \text{for any } |\,\rangle \in S \tag{7.1}$$

[5]Sometimes one also defines antilinear operators. They satisfy $A\{\alpha|a\rangle + \beta|b\rangle\} = \overline{\alpha}\{A|a\rangle\} + \overline{\beta}\{A|b\rangle\}$.

**Definition 3** *We define the addition and multiplication of linear operators as $C = A + B$ and $D = AB$ if for any $|\,\rangle \in S$ we have*

$$C|\,\rangle = (A + B)|\,\rangle = A|\,\rangle + B|\,\rangle \qquad \textit{and} \qquad D|\,\rangle = (A \cdot B)|\,\rangle = A \cdot (B|\,\rangle) \tag{7.2}$$

Using the linearity properties of vector spaces together with the definitions $F$ and $G$, it is easy to show that $A + B$ and $A \cdot B$ are themselves linear operators and that the addition and the multiplication of operators satisfy all the rules of the addition and of the multiplication of numbers, with the exception of the commutative law for multiplication.

The reader may verify the preceding statements by using the methods of the examples below.

**Example 7** *Verification that $A \cdot B$ is itself a linear operator is*

$$(A \cdot B)(\alpha|a\rangle + \beta|b\rangle) = A\{\alpha(B|a\rangle) + \beta(B|b\rangle)\} = \alpha(AB)|a\rangle + \beta(AB)|b\rangle \tag{7.3}$$

**Example 8** *Verification that $C(A + B) = CA + CB$ is*

$$C(A + B)|\,\rangle = C(A|\,\rangle + B|\,\rangle) = CA|\,\rangle + CB|\,\rangle \tag{7.4}$$

We have mentioned that in general

$$AB - BA \neq 0 \tag{7.5}$$

The quantity of the LHS of (7.5) is called the commutator of $A$ and $B$ and is denoted by the symbol $[A, B]$

$$[A, B] = AB - BA \tag{7.6}$$

Operators whose commutator vanishes are called *commuting* operators. Of course any operator commutes with the unit operator, since

$$AE|\,\rangle = A|\,\rangle \qquad \text{and} \qquad EA|\,\rangle = A|\,\rangle \tag{7.7}$$

We shall consider as meaningful the multiplication of operators by numbers, treating the operator equation

$$B = \alpha A = A\alpha \tag{7.8}$$

as equivalent to the vector equation

$$B|\,\rangle = \alpha(A|\,\rangle) \qquad \text{for any } |\,\rangle \tag{7.9}$$

In order to preserve a consistent notation, however, we interpret the vector equation

$$A|\,\rangle = a|\,\rangle \quad \text{for any } |\,\rangle \tag{7.10}$$

as equivalent to the operator equation

$$A = \alpha E \tag{7.11}$$

while the equation $A = \alpha$ is meaningless. Having defined the product of two operators, we can, of course, also define an operator raised to a certain power. For example, by $A^m|\,\rangle$ we mean that

$$A^m|\,\rangle = \underbrace{A \cdot A \cdots A}_{m \text{ factors}}|\,\rangle \tag{7.12}$$

Similarly, one can define functions of operators by their (formal) power series expansions. Thus, for example, $e^A$ formally means

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots \tag{7.13}$$

Given an operator $A$ that acts on vectors $|\,\rangle$, one can define the action of the same operator on vectors $\langle\,|$. The action of $A$ on a vector $\langle\,|$ is defined by requiring[6] that for any $|a\rangle$ and $\langle b|$, one has

$$\{\langle b|A\}|a\rangle = \langle b|\{A|a\rangle\} = \langle b|A|a\rangle \tag{7.14}$$

The preceding definition maintains the symmetry between the vectors $\langle\,|$ and $|\,\rangle$. It should, however, be stressed that $\langle\,|A$ is in general not the dual vector of $A|\,\rangle$, as the following example shows.

**Example 9** *The dual vector of $E|b\rangle = |b\rangle$ is $\langle b|E = \langle b|$, but the dual vector of $(\alpha E)|b\rangle = \alpha|b\rangle$ is $\langle b|(E\overline{\alpha}) = \langle b|\alpha$ and not $\langle b|(E\alpha) = \langle b|\alpha$ ($\alpha$ is an arbitrary complex number).*

# 8 Some special operators

Certain operators with rather special properties play particularly important roles in theory and its applications. We shall consider some of them below.

The operator $X$ satisfying $XA = E$ is called the *left inverse* of $A$ and will be denoted by $A_l^{-1}$. Thus,

$$A_l^{-1}A = E \tag{8.1}$$

Similarly, the *right inverse* operator of $A$ is defined by the equation

$$AA_r^{-1} = E \tag{8.2}$$

It is worth mentioning that, in general, $AA_r^{-1} \neq E$ and $A_r^{-1}A \neq E$. Also, $A_l^{-1}$ or $A_r^{-1}$, or both, may not be unique and even may not exist at all. One has, however, the following important theorem :

**Theorem 1** *If, for a given $A$, both operators $A_l^{-1}$ and $A_r^{-1}$ exist, they then are unique and*

$$A_l^{-1} = A_r^{-1} \tag{8.3}$$

*If $A_l^{-1}$ is unique, then*

$$AA_l^{-1} = E \tag{8.4}$$

*and $A_r^{-1}$ is also a unique right inverse of $A$. Similarly, if $A_r^{-1}$ is unique, then*

$$AA_r^{-1} = E \tag{8.5}$$

*and $A_r^{-1}$ is a unique left inverse of $A$.*

**Proof 1** *Multiplying (8.1) from the right by $A_r^{-1}$ and (8.2) from the left by $A_l^{-1}$, we get*

$$A_l^{-1}AA_r^{-1} = A_r^{-1} \qquad \text{and} \qquad A_l^{-1}AA_r^{-1} = A_l^{-1} \tag{8.6}$$

[6] We shall use the convention of operating on $\langle\,|$ from the right.

*Hence*

$$A_l^{-1} = A_r^{-1} \tag{8.7}$$

*The proof holds for any pair of operators $A_l^{-1}$ and $A_r^{-1}$ and (8.6) ensures that there exists only one such pair. Multiplying (8.1) from the left by $A$, we have*

$$AA_l^{-1}A = A \tag{8.8}$$

*Thus, adding (8.1) and (8.8), we get*

$$AA_l^{-1}A + A_l^{-1}A = A + E \tag{8.9}$$

*or*

$$(AA_l^{-1} + A_l^{-1} - E)A = E \tag{8.10}$$

*Assuming that $A_l^{-1}$ is unique, we obtain*

$$AA_l^{-1} + A_l^{-1} - E = A_l^{-1} \tag{8.11}$$

*and therefore*

$$AA_l^{-1} = E \tag{8.12}$$

*Hence, $A_l^{-1}$ is also a right inverse of $A$, and from the first part of the theorem it follows that it is a unique right inverse of $A$. Similarly, one proves an analogous result for $A_r^{-1}$.*

When both $A_l^{-1}$ and $A_r^{-1}$ exist, then the unique operator $A^{-1}$ defined by the equation

$$A^{-1} = A_l^{-1} = A_r^{-1} \tag{8.13}$$

is called the operator *inverse* to $A$. Using the rules of operator multiplication one easily obtains[7]

$$(AB)^{-1} = B^{-1}A^{-1} \tag{8.14}$$

provided $B^{-1}$ and $A^{-1}$ exist, since then we have

$$(AB)^{-1}AB = B^{-1}(A^{-1}A)B = B^{-1}EB = B^{-1}B = E \tag{8.15}$$

$$AB(AB)^{-1} = A(BB^{-1})A^{-1} = AEA^{-1} = AA^{-1} = E \tag{8.16}$$

Suppose now that the scalar product is defined in $S$. Then the operator $A$ satisfying

$$\langle a|X|b\rangle = \overline{\langle b|A|a\rangle} \tag{8.17}$$

for any $|a\rangle, |b\rangle \in S$ is called the *adjoint* operator of $A$ and is denoted by $A^\dagger$. Hence

$$\langle a|A^\dagger|b\rangle = \langle b|A|a\rangle \tag{8.18}$$

for any $|a\rangle, |b\rangle \in S$. By inspection of (7.14) we see that $\langle\,|A^\dagger$ is a dual vector of $A|\,\rangle$. From (8.18) we find

$$\langle b|(A^\dagger)^\dagger|a\rangle = \langle b|A|a\rangle \tag{8.19}$$

[7]The analogous relation holds for right and left inverses of product of operators.

for any $|a\rangle$ and $|b\rangle$. Thus

$$(A^\dagger)^\dagger = A \tag{8.20}$$

Since, for any $|a\rangle$, $|b\rangle$, $\langle\,|B^\dagger$ and $B|b\rangle$ and $\langle a|A^\dagger$ and $A|a\rangle$ are pairs of dual vectors, one has and therefore

$$\langle b|B^\dagger A^\dagger|a\rangle = \left[\langle b|B^\dagger\right]\left[A^\dagger|a\rangle\right] = \overline{\left[\langle b|A\right]\left[B|a\rangle\right]} = \overline{\langle b|AB|a\rangle} = \langle b|(AB)^\dagger|a\rangle \tag{8.21}$$

and therefore

$$A^\dagger B^\dagger = (AB)^\dagger \tag{8.22}$$

We leave to the reader the verification that

$$(A+B)^\dagger = A^\dagger + B^\dagger \tag{8.23}$$

An operator H that is equal to its *adjoint*, i.e., which obeys the relation

$$H = H^\dagger \tag{8.24}$$

is called *Hermitian*. An operator $U$ that satisfies the condition

$$U^\dagger = U^{-1} \tag{8.25}$$

is called *unitary*. Unitary operators have the remarkable property that their action on a vector preserves the length of that vector. In fact, the length of $|a\rangle$ is $\sqrt{\langle a|a\rangle}$ which is the same as the length of $U|a\rangle$, for we have

$$\{\langle a|U^\dagger\}\{U|a\rangle\} = \langle a|U^{-1}U|a\rangle = \langle a|a\rangle \tag{8.26}$$

A particular notation turns out to be quite useful. A symbol of the type of $|a\rangle\langle b|$ has all the properties of a linear operator; multiplied from the right by a $|\,\rangle$, it gives another $|\,\rangle$; multiplied from the left by a $\langle\,|$ it gives a $\langle\,|$. The linearity of $|a\rangle\langle b|$ results from the linear properties of the scalar product. One also has

$$\{|a\rangle\langle b|\}^\dagger = |b\rangle\langle a| \tag{8.27}$$

since

$$\langle x|\{|b\rangle\langle a|\}|y\rangle = \langle x|b\rangle\langle a|y\rangle = \overline{\langle y|a\rangle\langle b|a\rangle} = \overline{\langle y|\{|a\rangle\langle b|\}|a\rangle} \tag{8.28}$$

for arbitrary $|a\rangle$ and $|y\rangle$.

Note that since quantities of the type $\langle\,|\,\rangle$ are pure numbers, they can be placed either to the left or to the right of vectors $|\,\rangle$ or $\langle\,|$.

We need to assume

$$|a\rangle\{\langle b| + \langle c|\} = |a\rangle\langle b| + |a\rangle\langle c| \tag{8.29}$$

From (8.27) and (8.29) it follows then that

$$\{|b\rangle + |c\rangle\}\langle a| = |b\rangle\langle a| + |c\rangle\langle a| \tag{8.30}$$

since

$$[|b\rangle + |c\rangle]\langle a| = \{|a\rangle[\langle b| + \langle c|]\}^\dagger = [|a\rangle\langle b| + |a\rangle\langle c|]^\dagger = [|a\rangle\langle b|]^\dagger + [|a\rangle\langle c|]^\dagger = |b\rangle\langle a| + |c\rangle\langle a| \quad (8.31)$$

Consider the set $S_e$ of all vectors that can be obtained by multiplying the vector $|c\rangle$ of unit length, $\langle c|c\rangle = 1$, by a complex number. Evidently this set constitutes a linear vector space, and moreover $|a\rangle \in S_e$ implies $|a\rangle \in S$ so that $S_e \subset S$. A space that is a subset of a larger space is called a *subspace* of this space.

The operator $P_e = |e\rangle\langle e|$ has the property that if any $|\,\rangle$ is multiplied by it, one gets a vector proportional to $|e\rangle$, and therefore belonging to $S_e$.

$$P_e|\,\rangle = \langle e|\,\rangle|e\rangle \in S_e \quad (8.32)$$

since $\langle e|\,\rangle$ is simply a complex number. Also

$$P_e|\,\rangle = |\,\rangle \qquad \text{for any } |\,\rangle \in S_e \quad (8.33)$$

We say that $P_e$ projects $|\,\rangle$ on the subspace $S_e$. $P_e$ is a very particular example of a *projection operator*.

A linear operator $P$ is called a projection operator if it is Hermitian and if

$$P^2 = P \quad (8.34)$$

If $P$ had an inverse, then by multiplying both sides of (8.34) by $P^{-1}$, one would have

$$P = E \quad (8.35)$$

Therefore, the only projection operator that has an inverse is the identity operator ($E$ is obviously a projection operator).

If $P_1$ and $P_2$ are two projection operators, then $P_1 + P_2$ is also a projection operator if and only if

$$P_1P_2 = P_2P_1 = 0 \quad (8.36)$$

To see this, we note that $P_1 + P_2$ is a projection operator if

$$(P_1 + P_2)^2 = P_1 + P_2 \quad (8.37)$$

i.e., since $P_1^2 = P_1$ and $P_2^2 = P_2$, if

$$P_1P_2 + P_2P_1 = 0 \quad (8.38)$$

Multiplying (8.38) by $P_1$ from either the right or left, we have

$$P_1P_2P_1 + P_2P_1 = 0 \qquad \text{and} \qquad P_1P_2 + P_1P_2P_1 = 0 \quad (8.39)$$

Hence

$$P_1P_2 - P_2P_1 = 0 \quad (8.40)$$

Equations (8.38) and (8.40) yield (8.36) Conversely, if (8.36) is satisfied, then so is (8.37). Operators that satisfy the conditions if (8.36) are called *orthogonal operators*. More generally if the $P_i$ ($i = 1, 2, \ldots, N$) are a set of $N$ orthogonal projection operators satisfying

$$P_iP_j = \begin{cases} P_i & i = j \\ 0 & i \neq j \end{cases} \quad (8.41)$$

then $P = \sum_{n=1}^{N} P_i$ is also a projection operator.

**Example 10** *Take a real space of vectors represented by arrows in a plane, as discussed in example 3, section 1. Let* $|e_1\rangle$ *and* $|e_2\rangle$ *be two orthogonal (i.e., perpendicular) unit vectors*

$$\langle e_1|e_1\rangle = \langle e_2|e_2\rangle = 1 \qquad \textit{and} \qquad \langle e_1|e_2\rangle = \langle e_2|e_1\rangle = 0 \tag{8.42}$$

*Then*

$$P_2 = |e_1\rangle\langle e_1| \qquad \textit{and} \qquad P_2 = |e_2\rangle\langle e_2| \tag{8.43}$$

*are projection operators, since*

$$P_1 = |e_1\rangle\langle e_1|e_1\rangle\langle e_1| = |e_1\rangle\langle e_1| = P_1 \tag{8.44}$$

*and similarly for* $P_2$*. Furthermore,* $P_1$ *and* $P_2$ *are orthogonal projection operators, since for any vector* $|a\rangle$*, we have (on account of (8.42))*

$$P_1P_2|a\rangle = |e_1\rangle\langle e_1|e_2\rangle = 0 \tag{8.45}$$

*and similarly*

$$P_2P_1|a\rangle = 0 \tag{8.46}$$

*Applying* $P_1$ *to an arbitrary vector* $|a\rangle$*, we have*

$$P_1|a\rangle = |e_1\rangle\langle e_1|a\rangle = |e_1\rangle\, |\vec{a}|\, \cos\psi_{a,e_1} \tag{8.47}$$

*Thus,* $P_1|a\rangle$ *is a vector directed along* $|e_1\rangle$ *and has a length reduced, as compared to* $|\vec{a}|$*, by the usual cosine factor. Analogously,* $P_2$ *projects an arbitrary vector along the direction of* $|e_2\rangle$*.*