2 Vector spaces and Hilbert spaces

2.1 What is a vector space?

In a nutshell, the vector spaces used in Physics are sets of mathematical objects which can be combined to each other and multiplied by numbers, pretty much in the same way as vectors representing velocities, forces or accelerations can be added to each other and multiplied by numbers. In this context, the mathematical objects belonging to such sets are called vectors, whether or not they can be represented by arrows in ordinary 3D space, and the numbers by which they are multiplied are called scalars. One talks about a real vector space when these numbers are real (rational or irrational numbers without an imaginary part), and a complex vector space when they are complex numbers. Quantum Mechanics is based on complex vector spaces.

More precisely, a real vector space is a set V in which are defined a vector addition and a multiplication by a scalar subject to the following axioms.

1. The vector addition associates one and only one element of V with each pair $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ of elements of V. This element is called the sum of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ and is denoted by $\boldsymbol{v}_1 + \boldsymbol{v}_2$. (The elements of V are called vectors. One says that V is closed under vector addition, meaning that the sum of two elements of V is always an element of V.)

2. The vector addition is associative. I.e., for any three elements $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ of V, adding the sum of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ to $\boldsymbol{v}_3$ gives the same result as additing $\boldsymbol{v}_1$ to the sum of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$:
$$(\boldsymbol{v}_1 + \boldsymbol{v}_2) + \boldsymbol{v}_3 = \boldsymbol{v}_1 + (\boldsymbol{v}_2 + \boldsymbol{v}_3). \tag{2.1}$$

3. One of the elements of V, $\mathbf{0}$, and only this element, is such that
$$\boldsymbol{v} + \mathbf{0} = \boldsymbol{v} \tag{2.2}$$
for any element $\boldsymbol{v}$ of V. The vector $\mathbf{0}$ which has this property is called the zero vector (or the null vector).

4. Every element $\boldsymbol{v}$ of V has one and only one inverse element in V, namely a vector $-\boldsymbol{v}$ such that
$$\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}. \tag{2.3}$$
(If there is a risk of confusion with other meanings of the word inverse, one can say that the vector $-\boldsymbol{v}$ is the additive inverse of $\boldsymbol{v}$.)

5. The vector addition is commutative. I.e., for any two elements $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ of V,
$$\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{v}_2 + \boldsymbol{v}_1. \tag{2.4}$$

6. The multiplication by a scalar associates one and only one element of V with each real number α and each element $\boldsymbol{v}$ of V. This element is called the product of $\boldsymbol{v}$ by α and is denoted by $\alpha\,\boldsymbol{v}$. (Real numbers are called scalars in this context. One says that V is closed under multiplication by a scalar, meaning that the product of an element of V by a scalar is always an element of V. We stress that the operation we are talking about here gives a vector as a result; it should not be confused with the scalar product or dot product of two vectors, which gives a scalar as result.)

7. This operation is distributive. I.e., for any real numbers α and β and any elements $\boldsymbol{v}$, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ of V,
$$(\alpha + \beta)\,\boldsymbol{v} = \alpha\,\boldsymbol{v} + \beta\,\boldsymbol{v} \tag{2.5}$$
and
$$\alpha\,(\boldsymbol{v}_1 + \boldsymbol{v}_2) = \alpha\,\boldsymbol{v}_1 + \alpha\,\boldsymbol{v}_2. \tag{2.6}$$

8. Multiplication by a scalar is also associative. I.e., for any real numbers α and β and any element $\boldsymbol{v}$ of V,
$$(\alpha\beta)\,\boldsymbol{v} = \alpha\,(\beta\,\boldsymbol{v}). \tag{2.7}$$

9. For every element $\boldsymbol{v}$ of V,
$$1\,\boldsymbol{v} = \boldsymbol{v}. \tag{2.8}$$

The definition of a complex vector space is identical, except that the scalars are taken to be complex numbers, not real numbers.

☞ Vector spaces are also called linear spaces.

☞ These axioms ensure that whatever the elements of a vector space are, calculations involving these elements follow all the rules you routinely use when adding vectors representing positions, forces or velocities and multiplying those by numbers. For example, these axioms imply that for any vector $\boldsymbol{v}$, (1) $0\,\boldsymbol{v} = \mathbf{0}$ and (2) $(-1)\,\boldsymbol{v} = -\boldsymbol{v}$.

Proof: (1) Let $\boldsymbol{w} = 0\,\boldsymbol{v}$. Setting $\alpha = \beta = 0$ in Eq. (2.5) gives $0\,\boldsymbol{v} = 0\,\boldsymbol{v} + 0\,\boldsymbol{v}$, i.e., $\boldsymbol{w} = \boldsymbol{w} + \boldsymbol{w}$. Adding the vector $-\boldsymbol{w}$ to each side of this equation gives $\boldsymbol{w} + (-\boldsymbol{w}) = (\boldsymbol{w} + \boldsymbol{w}) + (-\boldsymbol{w})$. By virtue of Eqs. (2.1), (2.2) and (2.3), this last equation simplifies to $\mathbf{0} = \boldsymbol{w}$, which shows that indeed $0\,\boldsymbol{v} = \mathbf{0}$ for any vector $\boldsymbol{v}$. (2) Setting $\alpha = 1$ and $\beta = -1$ in Eq. (2.5) gives $0\,\boldsymbol{v} = 1\,\boldsymbol{v} + (-1)\,\boldsymbol{v}$, i.e., $\mathbf{0} = \boldsymbol{v} + (-1)\,\boldsymbol{v}$. Hence $(-1)\,\boldsymbol{v}$ the inverse of an additive inverse As, by axiom 4, $-\boldsymbol{v}$ is the only vector which fulfills this equation, $(-1)\,\boldsymbol{v}$ is necessarily $-\boldsymbol{v}$. □

☞ A mathematical concept related to vector spaces, and also very important in Theoretical Physics, is that of groups. In fact, the first five axioms in the list above mean that the elements of a vector space form an Abelian group under vector addition (i.e., a group for which the operation associating the elements of the set is commutative). Further information about this topic and examples of applications in Physics can be found, e.g., in the Mathematics textbook by Riley, Hobson and Bence (group theory as such is outside the scope of this course).

☞ One can also define more general vector spaces in which the scalars multiplying the vectors are not specifically real or complex numbers. Instead, these scalars are taken to be the elements of what mathematicians call a field (not to be confused with a vector field), namely a set of numbers or other mathematical objects endowed with two operations following the same rules as the ordinary addition and multiplication between real numbers.

Notation

We will normally represent 3D geometric vectors (e.g., position vectors, velocity vectors, angular momentum vectors, etc.) by boldface upright letters (e.g., $\mathbf{v}$), and elements of other vector spaces (e.g., vectors representing quantum states) by normal fonts or some other symbols (e.g., χ_+ or $|\psi\rangle$). Also, we will generally use the symbol 0 to represent the zero vector, unless it would be desirable to emphasize the different between this vector and the number 0.

Examples of vector spaces

- 3D geometric vectors

  By 3D geometric vectors we mean the "arrow vectors" you have often used to describe physical quantities which have both a magnitude and a direction. Suppose that $\mathbf{v}_1$ and $\mathbf{v}_2$ are two such vectors (e.g., two different forces acting on a same particle). You are familiar with the fact that $\mathbf{v}_1$ can be summed to $\mathbf{v}_2$ and that the result is also a geometric vector (the vector given by the parallelogram rule). One can also multiply a geometric vector by a real number: by definition of this operation, the product of a vector $\mathbf{v}$ by a real number α is a vector of the same direction as $\mathbf{v}$ (if

$\alpha > 0$) or the opposite direction (if $\alpha < 0$) and whose length is α times the length of $\mathbf{v}$. Geometric vectors form a real vector space under these two operations.

As an exercise, check that these two operations have all the properties required by the definition of a vector space. The zero vector here is the vector whose length is zero (and whose direction is therefore undefined). Moreover, any geometric vector $\mathbf{v}$ has an additive inverse, $-\mathbf{v}$, which can be obtained by multiplying $\mathbf{v}$ by -1. (If $\mathbf{v}$ is not the zero vector, then $-\mathbf{v}$ is a vector of same length as $\mathbf{v}$ but of opposite direction.)

- Two-component column vectors of complex numbers

These vectors are single-column arrays of two complex numbers, for example the vectors χ_+ and χ_- mentioned in Section 1.1 in regards to the Stern-Gerlach experiment. As we will see later in the course, two-component complex column vectors can be used to represent quantum states of a spin-1/2 particle. Such column vectors can be added together to give a column vector of the same form: by definition of this operation, if a, a', b and b' are complex numbers,

$$\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} a + a' \\ b + b' \end{pmatrix}. \tag{2.9}$$

One can also multiply a column vector of complex numbers by a complex number, the result being a column vector of complex numbers. For example,

$$(2 + 3i) \begin{pmatrix} a \\ b \end{pmatrix} \equiv \begin{pmatrix} 2a + 3ia \\ 2b + 3ib \end{pmatrix}. \tag{2.10}$$

It is easy to see that these two operations have all the properties required by the definition of a vector space. The zero vector, here, is a column vector of zeros, since adding a column vector of zeros to another column vector does not change the latter:

$$\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}. \tag{2.11}$$

Moreover, any column vector has an additive inverse vector, which is obtained by multiplying each of its components by -1:

$$\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} -a \\ -b \end{pmatrix} = \begin{pmatrix} a - a \\ b - b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.12}$$

- N-component column vectors of real or complex numbers

  All what is said above in regards to 2-component column vectors of complex numbers also applies to column vectors of more than two components and to column vectors of real rather than complex numbers: For any $N \geq 1$, the set of N-component column vectors of real numbers is a real vector space and the set of N-component column vectors of complex numbers is a complex vector space.

- Functions of a real variable

  Here we consider the set of the functions $f(x)$ defined for all real x. This set is a vector space, defining vector addition as the ordinary sum of two such functions and multiplication by a scalar as the ordinary product by a number. To the risk of being pedantic:

  - The sum of a function $f_1(x)$ and a function $f_2(x)$ is the function $(f_1 + f_2)(x)$ such that

    $$(f_1 + f_2)(x) \equiv f_1(x) + f_2(x). \tag{2.13}$$

    That is to say, $(f_1 + f_2)(x)$ is the function whose value is given by the sum of the values of $f_1(x)$ and $f_2(x)$ at any x.

  - The product of a function $f(x)$ by a number α is the function $(\alpha f)(x)$ such that

    $$(\alpha f)(x) \equiv \alpha f(x). \tag{2.14}$$

  This vector space is real or complex according to whether the scalars α are real numbers or complex numbers.

> Again, check, as an exercise, that this set and these two operations fulfill the definition of a vector space. The function $0(x)$ whose value is zero for all values of x plays the role of the zero vector for this vector space, since
>
> $$f(x) + 0(x) \equiv f(x) \tag{2.15}$$
>
> for any function $f(x)$ if $0(x) \equiv 0$. Moreover, it is also the case that to any function $f(x)$ corresponds a function $-f(x) \equiv (-1)f(x)$ such that
>
> $$f(x) + [-f(x)] \equiv 0(x). \tag{2.16}$$
>
> [Don't be confused by the terminology: If $f(x)$ is regarded as an element of this vector space, then $-f(x)$ is its inverse element (its additive inverse), not its inverse function. The latter is the function $f^{-1}(y)$ such that if $f(x) = y$ then $f^{-1}(y) = x$.]

- Square-integrable functions of a real variable

  By a square-integrable function of a real variable we mean a function $f(x)$ (possibly taking complex values) such that the integral
$$\int_a^b |f(x)|^2 \,\mathrm{d}x$$
  exists and is finite (for functions square-integrable on a finite interval $[a, b]$) or such that the integral
$$\int_{-\infty}^{\infty} |f(x)|^2 \,\mathrm{d}x$$
  exists and is finite (for functions square-integrable on the infinite interval $(-\infty, \infty)$). Such functions also form a vector space, defining vector addition and multiplication by a scalar in the same way as in the previous example. They play an important role in Quantum Mechanics.

The question of whether the set of *all* square-integrable functions forms a vector space is rather subtle and cannot be fully addressed without using mathematical concepts outside the scope of this course. Recall that the axioms of a vector space require that the sum of any two elements of a vector space V is also an element of V, and likewise for the product of any element of V by any number. In other words, they require that V is closed under vector addition and multiplication by a scalar. It is clear that multiplying a square-integrable function $f(x)$ by a real or complex finite number α always results in a square-integrable function, in agreement with the axioms of a vector space: since

$$\int_{-\infty}^{\infty} |\alpha f(x)|^2 \,\mathrm{d}x = |\alpha|^2 \int_{-\infty}^{\infty} |f(x)|^2 \,\mathrm{d}x, \tag{2.17}$$

the integral of $|\alpha f(x)|^2$ exists and is finite if the integral of $|f(x)|^2$ exists and is finite.

For the set of all square-integrable functions to be a vector space, it is also necessary that the sum of any two square-integrable functions is a square-integrable function. This is not difficult to prove for the case where we would only consider functions that are continuous everywhere (which is not restrictive for us, as wave functions used in Quantum Mechanics are normally continuous).

Proof: The sum of two continuous functions $f(x)$ and $g(x)$ is a continuous function and so are the functions $|f(x)|^2$, $|g(x)|^2$ and $|f(x)+g(x)|^2$. The case where these functions are defined and continuous on a closed interval $[a, b]$ is simple, as continuity on a closed interval implies integrability on that interval; hence, $f(x) + g(x)$ is square-integrable on $[a, b]$. The case of of functions defined and continuous on the infinite interval $(-\infty, \infty)$ is not as simple, however, since not all functions that are continuous on $(-\infty, \infty)$ are also integrable on $(-\infty, \infty)$. Let us assume that $f(x)$ and $g(x)$ are both square-integrable and continuous. Hence, the integrals

$$\int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x \quad \text{and} \quad \int_{-\infty}^{\infty} |g(x)|^2 \, \mathrm{d}x$$

exist and are finite. We note that at any x,

$$\begin{aligned} |f(x) + g(x)|^2 &= [f(x) + g(x)][f^*(x) + g^*(x)] \\ &= |f(x)|^2 + |g(x)|^2 \\ &\qquad + f^*(x)g(x) + f(x)g^*(x). \end{aligned} \tag{2.18}$$

Similarly,

$$|f(x) - g(x)|^2 = |f(x)|^2 + |g(x)|^2 - [f^*(x)g(x) + f(x)g^*(x)]. \tag{2.19}$$

From this last equation, and from the fact that $|f(x) - g(x)|^2 \geq 0$, we deduce that

$$f^*(x)g(x) + f(x)g^*(x) \leq |f(x)|^2 + |g(x)|^2. \tag{2.20}$$

Coming back to Eq. (2.18), we thus have that

$$|f(x) + g(x)|^2 \leq 2|f(x)|^2 + 2|g(x)|^2. \tag{2.21}$$

Hence

$$\int_{-\infty}^{\infty} |f(x) + g(x)|^2 \, \mathrm{d}x \leq 2 \int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x + 2 \int_{-\infty}^{\infty} |g(x)|^2 \, \mathrm{d}x. \tag{2.22}$$

The integral of $|f(x) + g(x)|^2$ is therefore finite (it is bounded above by the sum of two finite numbers). □

The matter turns out to be more complicated for more general functions, however. In fact, the mathematically theory of vector spaces of square-integrable functions is based not on the Riemann integral (the familiar integral you have learned at school) but on a generalization of this operation called the Lebesgue integral. However, the functions we are interested are normally square-integrable in terms of the familiar Riemann integral, and any function which is square-integrable in terms of the Riemann integral is also square-integrable in terms of the Lebesgue integral. While important for a rigorous mathematical study of functional spaces, the difference between these two types of integrals can thus be ignored at the level of this course. Square-integrable functions in the Lebesgue sense are often referred to as square-summable functions or as L^2 functions.

2.2 Subspaces

A subspace of a vector space V is a subset of V which itself forms a vector space under the same operations of vector addition and scalar multiplication as in V.

<u>**Examples**</u>

- We have seen that the set of all N-component column vectors of real numbers is a real vector space. For $N = 3$, these column vectors take the following form,
$$\begin{pmatrix} a \\ b \\ c \end{pmatrix},$$
where a, b and c are three real numbers. Amongst these vectors are those whose third component is zero, i.e., column vectors of the form
$$\begin{pmatrix} a \\ b \\ 0 \end{pmatrix}.$$
These particular column vectors form a subspace of the larger vector space formed by all 3-component real column vectors.

> It is clear that summing any two column vectors whose third component is zero gives a column vector whose third component is zero, and likewise, that multiplying any of them by a number also gives a column vector whose third component is zero. This set is therefore closed under vector addition and under multiplication by a scalar, as required by the axioms of a vector space.
>
> By contrast, the set of all 3-component column vectors whose third component is 1 is not closed under these two operations, and is therefore not a vector space.

- As seen at the end of the previous section, the set of all functions of a real variable is a vector space and the set of all square-integrable functions of a real variable is also a vector space. The set of all square-integrable functions is a subset of the set of all functions since all square-integrable functions are functions but not all functions are square-integrable. Correspondingly, the set formed by all square-integrable functions is a subspace of the vector space formed by all functions.

  Many other subspaces of this vector space can be considered, e.g., the subspace formed by all continuous functions, the subspace of that subspace formed by all functions which are both continuous and have a continuous first order derivative, the subspace formed by the periodic functions with period 2π, etc.

☞ Suppose that V_1 and V_2 are two subspaces of a certain vector space V, and that any vector v of V can be written in one and only one way as the sum of a vector v_1 of V_1 and a vector v_2 of V_2. The vector space V is then said to be the direct sum of V_1 and V_2. Direct sums of vectors spaces are denoted by the symbol $\oplus$: $V = V_1 \oplus V_2$.

For example, the space formed by the 3-component column vectors

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

is the direct sum of the space formed by the 3-component column vectors

$$\begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$$

and the space formed by the 3-component column vectors

$$\begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}.$$

2.3 Linear combinations

A linear combination of a finite number of vectors is a mathematical expression linear in the vectors involved. For example,

$$\mathbf{a} + \mathbf{b}$$

is a linear combination of the two geometric vectors **a** and **b**;

$$(2+3i)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (2-3i)\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

is a linear combination of two-component complex column vectors;

$$2\sin x + \cos 3x$$

is a linear combination of the functions $\sin x$ and $\cos 3x$; etc.

The general form of a linear combination of N vectors is

$$c_1 v_1 + c_2 v_2 + c_3 v_3 + \cdots + c_N v_N$$

where v_1, v_2, v_3, …, v_N are vectors and c_1, c_2, c_3, …, c_N are scalars.

> ☞ Extending this definition to a linear combination of an infinite number of vectors is fraught with mathematical difficulties. However, the extension is not impossible. For instance, a convergent Fourier series is in effect an infinite linear combination of vectors, each vector being a complex exponential (or a sine or cosine function).

2.4 The span of a set of vectors

The span of a set of N vectors is the set of all linear combinations of these vectors.

> The span of N vectors belonging to a vector space V is either V itself or a subspace of V not equal to V.
>
> Proof: By definition of this set, the span of N vectors is closed under vector addition and multiplication by a scalar. It is either V itself or a subspace of V not equal to V because any linear combination of elements of V must be an element of V since V is also closed under these two operations. □

Examples

- The span of the two column vectors

$$\begin{pmatrix}1\\0\\0\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}0\\1\\0\end{pmatrix}$$

  is the vector space formed by all 3-component column vectors of the form

$$\begin{pmatrix}a\\b\\0\end{pmatrix}.$$

☞ The two numbers a and b are real if we exclude linear combinations involving complex numbers. In this case, we can take these two numbers to be the x- and y-coordinates of a point in 3D space, the z-coordinate of this point being 0. If we do so, the column vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

then represent unit vectors starting at the origin and oriented, respectively, in the x- and y-directions, and saying that these two column vectors span the space formed by all column vectors of the form

$$\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

is the same as saying that these two unit vectors span the whole xy-plane.

- As you have seen in a previous course, the spherical harmonics $Y_{lm}(\theta, \phi)$ are certain functions of the polar angles θ and ϕ labelled by the quantum numbers l and m. You may remember that the quantum number m can only take the values -1, 0 and 1 for $l = 1$. These three functions span the set of all the functions of the form

$$f(\theta, \phi) = c_{-1}\, Y_{1-1}(\theta, \phi) + c_0\, Y_{10}(\theta, \phi) + c_1\, Y_{11}(\theta, \phi), \tag{2.23}$$

  where c_{-1}, c_0 and c_1 are three complex numbers.

2.5 Linear independence

A set formed by N vectors v_1, v_2, ..., v_N, and these N vectors themselves, are said to be linearly independent if none of these vectors can be written as a linear combination of the other vectors of the set. In the opposite case, one says that the set is linearly dependent (or that these N vectors are linearly dependent).

Examples

- The three column vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are linearly independent.

- By contrast, the three column vectors

$$\begin{pmatrix}1\\2\\1\end{pmatrix}, \begin{pmatrix}2\\1\\1\end{pmatrix}, \begin{pmatrix}5\\4\\3\end{pmatrix}$$

are linearly dependent since

$$\begin{pmatrix}5\\4\\3\end{pmatrix} = \begin{pmatrix}1\\2\\1\end{pmatrix} + 2\begin{pmatrix}2\\1\\1\end{pmatrix}. \tag{2.24}$$

☞ A necessary and sufficient condition for N non-zero vectors v_1, v_2, ..., v_N to be linearly independent is that the equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_N v_N = 0 \tag{2.25}$$

implies that the scalars c_1, c_2, ..., c_N are all zero.

Proof: Suppose that the equation was possible with one of the coefficients c_n non-zero, although these N vectors were both non-zero and linearly independent. Let us suppose that this non-zero coefficient was c_1 (if not c_1, we can always relabel the coefficients c_n and the vectors v_n so that $n = 1$ for this non-zero coefficient). We could then rewrite this equation as

$$v_1 = -(c_2 v_2 + \cdots + c_N v_N)/c_1, \tag{2.26}$$

which shows that v_1 must then be the zero vector or a non-zero linear combination of the vectors v_2, ..., v_N. These two possibilities are in contradiction with the hypothesis that the N vectors v_1, v_2, ..., v_N are both non-zero and linearly independent. The condition is thus necessary. The condition is also sufficient, as if, e.g., the vector v_1 was a linear combination of the vectors v_2, ..., v_N we could write

$$v_1 = a_2 v_2 + \cdots + a_N v_N \tag{2.27}$$

with at least one of the coefficients a_2, ..., a_N being non-zero. However, since this equation is the same as Eq. (2.25) with $c_1 = 1$ and $c_n = -a_n$ ($n \neq 1$), this would mean that the coefficients c_1, ..., c_N are not all zero. □

☞ The definition of linear independence given above applies to the case of a finite set of vectors. One says that an infinite set of vectors is linearly independent when every finite subset of this set is linearly independent.

The Exchange Theorem

A vector space spanned by N vectors does not contain linearly independent subsets of more than N vectors.

(Although this theorem is fairly intuitive, it cannot be proven in a few lines. The name of the theorem refers to an important step in its standard proof.)

☞ The Exchange Theorem implies, for example, that the three column vectors

$$\begin{pmatrix} a \\ b \\ 0 \end{pmatrix}, \begin{pmatrix} a' \\ b' \\ 0 \end{pmatrix}, \begin{pmatrix} a'' \\ b'' \\ 0 \end{pmatrix}$$

are always linearly dependent, irrespective of the values of a, a', a'', b, b' and b", since these three vectors belong to a vector space spanned by less than three vectors (see Section 2.4).

2.6 Dimension of a vector space

A vector space may be finite-dimensional of infinite-dimensional. A vector space is finite-dimensional and has a dimension N if it contains a linearly independent set of N vectors but no linearly independent set of more than N vectors. It is infinite-dimensional if it contains an arbitrarily large set of linearly independent vectors. (Note that the dimension of a vector space is unrelated to the number of elements this vector space contains. Finite-dimensional or not, vector spaces always contain an infinite number of vectors.[1])

A vector space spanned by N linearly independent vectors is finite-dimensional and its dimension is N.

Proof: The dimension of this vector space cannot be less than N since by construction it contains a linearly independent set of N vectors. Moreover, it cannot be larger than N since any set of more than N vectors belonging to this vector space is necessarily linearly dependent by virtue of the Exchange Theorem. □

A corollary of this theorem is that a vector space spanned by N linearly independent vectors cannot also be spanned by a set of fewer than N linearly independent vectors.

[1]This sweeping statement has an exception: the single-element set formed by the zero vector alone is a vector space in itself. This space contains only one vector…

Examples

- The vector space formed by all 3D geometric vectors is finite-dimensional and its dimension is 3 (unsurprisingly), since it is spanned by, e.g., the three unit vectors in the x-, y- and z-direction.

- The vector space spanned by the two column vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

  is finite-dimensional and its dimension is 2.

- As mentioned above, the set of all 2π-periodic functions is a vector space. This set includes, in particular, the complex exponentials $\exp(inx)$ ($n = 0, \pm 1, \pm 2, \ldots$). [These functions are 2π-periodic since, for any real x and any integer n, $\exp[in(x + 2\pi)] = \exp(inx)\exp(2n\pi i) = \exp(inx)$.] As we will see later, the set formed by the complex exponentials $\exp(inx)$ ($n = 0, \pm 1, \pm 2, \ldots, \pm n_{\text{max}}$) is linearly independent. This set is arbitrarily large since n_{max} can be as large as one wants. Therefore the vector space of all 2π-periodic functions is infinite dimensional.

- The vector space of all square-integrable functions of a real variable is also infinite-dimensional. (You have seen in the Term 1 course that the Hamiltonian of a linear harmonic oscillator has infinitely many orthonormal eigenfunctions. You know that these eigenfunctions are square-integrable functions, as otherwise they could not be normalized in the usual way. We will see later that these eigenfunctions are linearly independent. Hence this vector space contains an arbitarily large set of linearly independent vectors.)

2.7 Bases

A basis of a finite-dimensional vector space is a set of linearly independent vectors such that any vector belonging to this vector space can be written as a linear combination of these basis vectors. (The concept of basis for an infinite-dimensional vector space is more complicated; it is addressed in Section 2.12.)

> Although certain bases are more convenient than others, the choice of basis vectors is arbitrary as long as these vectors are linearly independent. In fact, any set of N linearly independent vectors belonging to a vector space of dimension N is a basis for this vector space. Infinitely many such bases can therefore be constructed.

Proof: By definition of the dimension of a vector space, a vector space of dimension N cannot contain a linearly independent set of more than N vectors. Suppose that a vector space V of dimension N contains a linearly independent set of N vectors but that this set would not be a basis for V. I.e., suppose that at least one of the elements of V could not be written as a linear superposition of these N basis vectors. Then that element could be joined to these N basis vectors to form a linearly independent set of $N + 1$ vectors, which is in contradiction with the hypothesis that the dimension of V is N. □

Given a basis of a vector space, there is one and only way of writing each element of that space as a linear combination of these basis vectors.

Proof: Suppose that there would be more than one way of writing a vector a as a linear combination of basis vectors v_1, v_2,..., v_N. I.e., one could write

$$a = c_1 v_1 + c_2 v_2 + \cdots + c_N v_N \tag{2.28}$$

and also

$$a = c_1' v_1 + c_2' v_2 + \cdots + c_N' v_N \tag{2.29}$$

with $c_n \neq c_n'$ for at least one value of n. However, subtracting these two equations gives

$$0 = (c_1 - c_1')v_1 + (c_2 - c_2')v_2 + \cdots + (c_N - c_N')v_N. \tag{2.30}$$

Since the vectors v_1, v_2,..., v_N form a basis, they must be linearly independent, and therefore this last equation is possible only if $c_n - c_n' = 0$ for all n. Hence there is only one way of writing the vector a as a linear combination of these basis vectors. □

Examples

- Any 3D geometric vector can be written as a linear combination of a unit vector oriented in the positive x-direction, a unit vector oriented in the positive y-direction and a unit vector oriented in the positive z-direction. These three unit vectors thus form a basis of this vector space. (Note that they are linearly independent: none of these three unit vectors is a linear combination of the other two.)

- The two column vectors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

form a basis for the vector space of the two-component column vectors since

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.31}$$

for any complex numbers a and b.

This basis is not unique. For instance, the two column vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

also form a basis for this vector space since for any complex numbers a and b one can always find a number α and a number β such that

$$\begin{pmatrix} a \\ b \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ 4 \end{pmatrix}. \tag{2.32}$$

(In fact, $\alpha = (3b - 4a)/2$ and $\beta = (2a - b)/2$.)

In each of these two cases there is only one way of writing a vector as a linear combination of the basis vectors. In contrast, there are infinitely many ways of writing a vector as a linear combination of linearly dependent vectors. For example, take the three column vectors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

These vectors are linearly dependent since

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.33}$$

Any 2-component column vector can be written as a linear combination of these three column vectors, as if they formed a basis; however, there are infinitely many ways of doing so since

$$\begin{pmatrix} a \\ b \end{pmatrix} = (a - \lambda) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (b - \lambda) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{2.34}$$

for any value of λ.

2.8 Inner product

The vector spaces most relevant to Quantum Mechanics are inner product spaces, namely vector spaces equipped with an operation called inner product (or scalar product) which extends the familiar dot product between geometric vectors to more general vector spaces.

We will denote the dot product of two geometric vectors $\mathbf{v}$ and $\mathbf{w}$ by the usual symbol $\mathbf{v} \cdot \mathbf{w}$, and the inner product of two other vectors v and w by (v, w) or a similar symbol. Other notations are sometimes used in Mathematics.

More precisely, for a complex vector field, an inner product is an operation which associates a complex number (v, w) to any pair of vectors v, w subject to the following axioms.

1. For any vectors v and w,
$$(v, w) = (w, v)^*. \tag{2.35}$$
(The order of the vectors in the pair thus matters.)

2. For any vectors v_1, v_2 and w and for any complex numbers α and β,
$$(\alpha\, v_1 + \beta\, v_2, w) = \alpha^*(v_1, w) + \beta^*(v_2, w). \tag{2.36}$$
(As can be seen if one sets $\alpha = \beta = 0$ in this equation, this axiom implies that the inner product of any vector with the zero vector is 0: $(0, w) = (w, 0) = 0$ for any vector w.)

3. $(v, v) = 0$ if and only if v is the zero vector, and $(v, v) > 0$ if v is not the zero vector.

Note the complex conjugations in Eq. (2.36). There is no complex conjugation if the vector on the right of the inner product is multiplied by a complex number:
$$(v, \alpha\, w_1 + \beta\, w_2) = \alpha(v, w_1) + \beta(v, w_2) \tag{2.37}$$
for any vectors v, w_1 and w_2 and any complex numbers α and β.

Proof: In view of Eqs. (2.35) and (2.36),
$$(v, \alpha\, w_1 + \beta\, w_2) = (\alpha\, w_1 + \beta\, w_2, v)^* \tag{2.38}$$
$$= [\alpha^*(w_1, v) + \beta^*(w_2, v)]^* \tag{2.39}$$
$$= \alpha(w_1, v)^* + \beta(w_2, v)^*. \tag{2.40}$$
Eq. (2.37) follow. □

The inner product is defined in the same way for real vector spaces, the only difference being that (v, w) and the scalars α and β are real numbers in the case of a real vector space.

Warning: The definition of the inner product used in this course differs from a similar notation widely used in Mathematics. As defined here, the inner product (v, w) is antilinear in the vector written on the left whereas in Mathematics (v, w) is usually taken to be antilinear in the vector written on the right. I.e., for us, $(\alpha\, v_1 + \beta\, v_2, w) = \alpha^*(v_1, w) + \beta^*(v_2, w)$ and $(v, \alpha\, w_1 + \beta\, w_2) = \alpha(v, w_1) + \beta(v, w_2)$, whereas mathematicians would instead write $(\alpha\, v_1 + \beta\, v_2, w) = \alpha(v_1, w) + \beta(v_2, w)$ and $(v, \alpha\, w_1 + \beta\, w_2) = \alpha^*(v, w_1) + \beta^*(v, w_2)$.

Examples

- The familiar dot product between geometric vectors is an inner product: For any two 3D geometric vectors, $\mathbf{v}$ and $\mathbf{w}$, $(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$. (There is no complex conjugation because this vector space is real.)

- For the 2-component complex column vectors used to represent spin states, the inner product is defined in the following way: If

$$v = \begin{pmatrix} a \\ b \end{pmatrix} \tag{2.41}$$

and

$$w = \begin{pmatrix} a' \\ b' \end{pmatrix}, \tag{2.42}$$

then

$$(v, w) = \begin{pmatrix} a^* & b^* \end{pmatrix} \begin{pmatrix} a' \\ b' \end{pmatrix} = a^* a' + b^* b'. \tag{2.43}$$

See Appendix A for a reminder of how to multiply a column vector by a row vector. Note that the components of the column vector v, which is on the left of (v, w), are complex conjugated when this inner product is calculated, and also that this column vector is written as a row vector in the product.

The rule is the same for column vectors of any number of components: We will always calculate the inner product (v, w) by transforming the complex conjugate of the column vector v into a row vector and multiplying the column vector w by this row vector. (Other ways of calculating the inner product are possible in principle but are not often used in Quantum Mechanics.)

Changing the column vector appearing on the left into a row vector can be taken as a convention. It differs from the one you probably follow when calculating the dot product of two geometric vectors, in which this calculation is written as a dot product of two column vectors formed by the x-, y- and z-components of the respective geometric vectors. These two conventions are entirely equivalent, although we will see that there are mathematical reasons for preferring the one adopted in this course.

- The inner product of two square-integrable functions $f(x)$ and $g(x)$ is defined as follows:

$$(f,g) = \int_{-\infty}^{\infty} f^*(x)\, g(x)\, \mathrm{d}x, \tag{2.44}$$

or, for functions which are square-integrable on a finite interval $[a,b]$,

$$(f,g) = \int_a^b f^*(x)\, g(x)\, \mathrm{d}x. \tag{2.45}$$

- The inner product of two square-integrable functions $f(\theta,\phi)$ and $g(\theta,\phi)$ of the polar angles θ and ϕ is defined as follows:

$$(f,g) = \int_0^{\pi} \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi\, f^*(\theta,\phi)\, g(\theta,\phi). \tag{2.46}$$

For functions $f(r,\theta,\phi)$ and $g(r,\theta,\phi)$ of the spherical polar coordinates r, θ and ϕ,

$$(f,g) = \int_0^{\infty} \mathrm{d}r\, r^2 \int_0^{\pi} \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi\, f^*(r,\theta,\phi)\, g(r,\theta,\phi). \tag{2.47}$$

2.9 Norm of a vector

The norm of a vector v, which we will represent by the symbol $||v||$ (or $|\mathbf{v}|$ for 3D geometric vectors) is the real number defined by the following equation:

$$||v|| = \sqrt{(v,v)} \tag{2.48}$$

(or $|\mathbf{v}| = \sqrt{\mathbf{v}\cdot\mathbf{v}}$ for geometric vectors).

A vector is said to be normalized if it has unit norm. (A normalized vector is also called a unit vector.)

Any non-zero vector can be normalized by multiplying it by the inverse of its norm: If $u = v/||v||$, then,

$$||u|| = \sqrt{(v,v)}/||v|| = 1. \tag{2.49}$$

(Clearly, the zero vector has a zero norm and cannot be normalized. By definition of an inner product, the zero vector is the only vector which has a zero norm.)

Example

The vector

$$v = \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix} \tag{2.50}$$

is a unit vector since

$$(v,v) = \begin{pmatrix} 1/\sqrt{2} & -i/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix} = 1/2 + (-i)(i)/2 = 1. \tag{2.51}$$

☞ As you know, $\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}|\,|\mathbf{w}| \cos\theta$ where θ is the angle between the vectors $\mathbf{v}$ and $\mathbf{w}$. Since $\cos\theta$ is a number between -1 and 1, its absolute value is never larger than 1. Hence $|\mathbf{v} \cdot \mathbf{w}| \leq |\mathbf{v}|\,|\mathbf{w}|$. This result is a particular case of the more general inequality

$$|(v,w)| \leq ||v||\,||w||, \tag{2.52}$$

called the Schwarz inequality (or Cauchy-Schwarz inequality), which applies to any inner product space.

Proof: If $v = 0$ or $w = 0$ (i.e., v or w is the zero vector), then $(v,w) = 0$, $||v|| = 0$ or $||w|| = 0$, and the inequality reduces to $0 \leq 0$, which is true. If neither v nor w is the zero vector, then we can define the normalized vectors $v' = v/||v||$ and $w' = w/||w||$ such that $(v',v') = (w',w') = 1$, and set $u' = v' - (w',v')\,w'$. Now, since $(w',v') = (v',w')^*$,

$$\begin{aligned} (u',u') &= (v' - (w',v')\,w'\,,\; v' - (w',v')\,w') \qquad (2.53)\\ &= 1 - (w',v')^*(w',v') - (w',v')(v',w') \\ &\qquad\qquad + (w',v')^*(w',v') \qquad (2.54)\\ &= 1 - |(v',w')|^2. \qquad (2.55) \end{aligned}$$

However, $(u',u') \geq 0$ by virtue of the third axiom of the inner product. Thus $1 - |(v',w')|^2 \geq 0$, and therefore $||v||\,||w|| \geq |(v,w)|$. □

2.10 Orthogonal vectors

Two vectors are said to be orthogonal if their inner product is zero. Two orthogonal unit vectors are said to be orthonormal.

Examples

- The two vectors

$$v = \begin{pmatrix} 1 \\ i \end{pmatrix} \text{ and } w = \begin{pmatrix} 1 \\ -i \end{pmatrix} \tag{2.56}$$

are orthogonal since

$$(v,w) = \begin{pmatrix} 1 & -i \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = 1 + (-i)(-i) = 1 - 1 = 0. \tag{2.57}$$

- The two complex exponentials $\exp(inx)$ and $\exp(imx)$ are orthogonal on the interval $[0, 2\pi]$ for any integer n and $m \neq n$. Indeed, if n and m are two integers and $m \neq n$, then $m - n$ is a non-zero integer and therefore

$$\int_0^{2\pi} [\exp(inx)]^* \exp(imx)\,\mathrm{d}x = \int_0^{2\pi} \exp[i(m-n)x]\,\mathrm{d}x \tag{2.58}$$

$$= \left.\frac{\exp[i(m-n)x]}{i(m-n)}\right|_0^{2\pi} \tag{2.59}$$

$$= \frac{\exp[2(m-n)\pi i] - 1}{i(m-n)} \tag{2.60}$$

$$= \frac{1-1}{i(m-n)} \tag{2.61}$$

$$= 0. \tag{2.62}$$

(We used the fact that $\exp(2k\pi i) = 1$ for any integer k.)

Gram-Schmidt orthogonalization

A set of linearly independent vectors can always be transformed into a set of vectors othogonal to each other by using a method known as Gram-Schmidt orthogonalization.

> ☞ The ortho*gonalization* method we are talking about differs from the similar method of Gram-Schmidt ortho*normalization* you may have studied in other courses.

The method is perhaps best explained by way of an example. Suppose that we want to form three vectors a', b' and c' orthogonal to each other, starting from three linearly independent non-zero vectors a, b and c. For the sake of the illustration, let us imagine that

$$a = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{2.63}$$

We can decide to include one of the latter amongst our set of orthogonal vectors. For instance, let us take a' to be the vector a:

$$a' = a = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \tag{2.64}$$

Then we form a vector b' orthogonal to a' by subtracting from b the vector a' multiplied by (a', b) and divided by (a', a'):

$$b' = b - [(a', b)/(a', a')]\, a'. \tag{2.65}$$

Here

$$(a', b) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 2 \quad \text{and} \quad (a', a') = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 3. \tag{2.66}$$

Thus, in our case,

$$b' = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ -2/3 \end{pmatrix}. \tag{2.67}$$

The vector b' so defined is always orthogonal to a' since

$$(a', b') = (a', b) - [(a', b)/(a', a')]\,(a', a') = (a', b) - (a', b) = 0. \tag{2.68}$$

Note that b' cannot be the zero-vector, as otherwise the vectors a and b would not be linearly independent. We then form a vector c' orthogonal to both a' and b' by the same process. I.e., we set

$$c' = c - [(a', c)/(a', a')]\,a' - [(b', c)/(b', b')]\,b'. \tag{2.69}$$

Here $(a', c) = 2$, $(a', a) = 3$, $(b', c) = -1/3$ and $(b', b') = 6/9 = 2/3$, and thus

$$c' = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1/3 \\ 1/3 \\ -2/3 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix}. \tag{2.70}$$

As can be checked easily, the three vectors so obtained,

$$a' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad b' = \begin{pmatrix} 1/3 \\ 1/3 \\ -2/3 \end{pmatrix} \quad \text{and} \quad c' = \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix}, \tag{2.71}$$

are orthogonal to each other.

This orthogonalization process can be iterated further to orthogonalize as many vectors as one wants, as long as these vectors are linearly independent.

Orthonormal bases

Bases formed of normalized vectors orthogonal to each other are particularly convenient. Suppose that the vectors u_1, u_2,..., u_N form a basis. Then, for any vectors v and w there exist a set of scalars c_1, c_2,..., c_N and a set of scalars d_1, d_2,..., d_N such that

$$v = \sum_{j=1}^{N} c_j u_j \quad \text{and} \quad w = \sum_{j=1}^{N} d_j u_j. \tag{2.72}$$

Suppose further that the vectors u_j are orthonormal — i.e., that $(u_i, u_j) = \delta_{ij}$ where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases} \tag{2.73}$$

The coefficients c_j and d_j can then be obtained as the inner product of the respective vector with the corresponding basis vector: $c_j = (u_j, v)$ and $d_j = (u_j, w)$. Moreover,

$$(v, w) = \sum_{j=1}^{N} c_j^* d_j \tag{2.74}$$

and therefore

$$||v||^2 = (v, v) = \sum_{j=1}^{N} |c_j|^2 \quad \text{and} \quad ||w||^2 = (w, w) = \sum_{j=1}^{N} |d_j|^2. \tag{2.75}$$

Proof: Taking the inner product of v with u_1, we see that

$$(u_1, v) = \sum_{j=1}^{N} c_j (u_1, u_j) = \sum_{j=1}^{N} c_j \delta_{1j} = c_1. \tag{2.76}$$

Similarly $(u_2, v) = c_2$, and in general $c_j = (u_j, v)$ and $d_j = (u_j, w)$. Moreover,

$$(v, w) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i^* d_j (u_i, u_j) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i^* d_j \delta_{ij}. \tag{2.77}$$

Eqs. (2.74) follows. □

Orthogonal subspaces

A subspace V' of a vector space V is said to be orthogonal to another subspace V'' of V if all the vectors of V' are orthogonal to all the vectors of V''.

For example, the space spanned by the two column vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

is orthogonal to the space spanned by the column vector

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

since

$$\begin{pmatrix} a^* & b^* & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} = 0 \tag{2.78}$$

for any complex numbers a, b and c.

☞ We have seen, in Section 2.4, that the space spanned by these first two column vectors can be represented geometrically by the xy-plane. Similarly, the one-dimensional space spanned by the third column vector can be represented by the z-axis, and saying that these two spaces are orthogonal is the same as saying that the z-axis is orthogonal to the xy-plane.

Orthogonality and linear independence

A set of non-zero orthogonal vectors is always linearly independent.

Proof: As seen in Section 2.5, we can be sure that N vectors are linearly independent if the equation

$$c_1\, v_1 + c_2\, v_2 + \cdots + c_N\, v_N = 0 \tag{2.79}$$

is possible only if all the coefficients c_n are zero. Suppose that the N vectors v_n are mutually orthogonal and non-zero. Suppose further that the equation would be possible with some of the coefficients being non-zero. Let c_j be a non-zero coefficient. Taking the inner product of the vector v_j with each side of the equation gives

$$c_1\,(v_j, v_1) + c_2\,(v_j, v_2) + \cdots + c_N\,(v_j, v_N) = (v_j, 0). \tag{2.80}$$

However, $(v_j, v_n) = 0$ if $j \neq n$ and $(v_j, 0) = 0$. Eq. (2.80) thus reduces to

$$c_j\,(v_j, v_j) = 0. \tag{2.81}$$

Since v_j is not the zero vector, $(v_j, v_j) \neq 0$. Thus c_j must be zero, in contradiction with the hypothesis that c_j is non-zero. Hence all the coefficients must be zero for the equation to be possible. □

For example, taken as functions defined on the interval $[0, 2\pi]$, the complex exponentials $\exp(inx)$ and $\exp(imx)$ are linearly independent for any integer n and $m \neq n$ since, as established above, these functions are orthogonal on that interval.

2.11 Isomorphic vector spaces

You are familiar with the fact that 3D geometric vectors representing, e.g., velocities, can be written as linear combinations of three orthogonal unit vectors. For example,

$$\mathbf{v} = v_x\,\hat{\mathbf{x}} + v_y\,\hat{\mathbf{y}} + v_z\,\hat{\mathbf{z}}, \tag{2.82}$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are the unit vectors in the x-, y- and z-directions. Given these three unit vectors, the components v_x, v_y and v_z are in one-to-one correspondence with the vector $\mathbf{v}$: Each geometric vector corresponds to one and only one set of components, and each set of three components corresponds to one and only one geometric vector. Since these sets of components can be arranged in column vectors, e.g.,

$$\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix},$$

Eq. (2.82) defines a one-to-one correspondence between the elements of the vector space of 3D geometric vectors and the vector space of 3-component column vectors of real numbers. In fact, adding geometric vectors or multiplying them by a number is equivalent to adding the corresponding column vectors or multiplying these by the same number. Also, the dot product of any two geometric vectors can be calculated as the inner product of the corresponding column vectors: If $\mathbf{v} = v_x\,\hat{\mathbf{x}} + v_y\,\hat{\mathbf{y}} + v_z\,\hat{\mathbf{z}}$ and $\mathbf{w} = w_x\,\hat{\mathbf{x}} + w_y\,\hat{\mathbf{y}} + w_z\,\hat{\mathbf{z}}$, then

$$\mathbf{v}\cdot\mathbf{w} = v_x w_x + v_y w_y + v_z w_z = \begin{pmatrix} v_x & v_y & v_z \end{pmatrix} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}. \tag{2.83}$$

(In writing the row vector, we have assumed that the components v_x, v_y and v_z are real numbers and therefore equal to their complex conjugate.)

In recognition of these facts, one says that the vector space of 3D geometric vectors and the vector space of 3-component column vectors of real numbers are isomorphic (which means, literally, that they have the same form). In a sense, there is only one such vector space, and its elements can be represented equally well by arrow vectors as by column vectors.

As another example, take the vector space spanned by the three spherical harmonics $Y_{1-1}(\theta,\phi)$, $Y_{10}(\theta,\phi)$ and $Y_{11}(\theta,\phi)$ (see Section 2.4). You have seen in the Term 1 course that these three functions are orthonormal; in fact, for any l, l', m and m',

$$\int_0^\pi \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi\, Y^*_{lm}(\theta,\phi)\, Y_{l'm'}(\theta,\phi) = \delta_{ll'}\delta_{mm'}. \tag{2.84}$$

These three spherical harmonics are thus linearly independent. Any element $f(\theta,\phi)$ of this vector space is a linear combination of the form

$$f(\theta,\phi) = c_{-1}\,Y_{1-1}(\theta,\phi) + c_0\,Y_{10}(\theta,\phi) + c_1\,Y_{11}(\theta,\phi), \tag{2.85}$$

where c_{-1}, c_0 and c_1 are three complex numbers. Since the three spherical harmonics are linearly independent, each of these functions can be written in only one way as a linear combination of that form. These functions are therefore in one-to-one correspondence with the column vectors

$$\begin{pmatrix} c_{-1} \\ c_0 \\ c_1 \end{pmatrix}$$

and the vector space spanned by these three spherical harmonics is isomorphic to the vector space of 3-component column vectors of complex numbers.

Exercise: Let

$$f(\theta,\phi) = c_{-1}\,Y_{1-1}(\theta,\phi) + c_0\,Y_{10}(\theta,\phi) + c_1\,Y_{11}(\theta,\phi), \tag{2.86}$$
$$g(\theta,\phi) = d_{-1}\,Y_{1-1}(\theta,\phi) + d_0\,Y_{10}(\theta,\phi) + d_1\,Y_{11}(\theta,\phi). \tag{2.87}$$

Show that the inner product of these two functions, defined as the integral

$$\int_0^\pi \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi\, f^*(\theta,\phi)\, g(\theta,\phi),$$

is $c_{-1}^* d_{-1} + c_0^* d_0 + c_1^* d_1$. Show that the same result is also obtained by taking the inner product of the corresponding column vectors,

$$\begin{pmatrix} c_{-1} \\ c_0 \\ c_1 \end{pmatrix} \text{ and } \begin{pmatrix} d_{-1} \\ d_0 \\ d_1 \end{pmatrix}.$$

2.12 Hilbert spaces

The vector spaces of greatest importance in Quantum Mechanics are inner product spaces which have a further mathematical property called completeness. Such vector spaces are called Hilbert spaces. (Recall that inner product spaces are vector spaces in which an inner product is defined. For the mathematical property called completeness, see below.)

We have already encountered several Hilbert spaces: The vector spaces formed by N-component column vector of complex numbers (e.g., column vectors representing spin states) are Hilbert spaces, and so are the vector spaces formed by square-integrable functions (e.g., wave functions).

What completeness is is not important for this course. Briefly, if one takes an infinite set of vectors belonging a vector space V and this infinite set forms a convergent sequence (in a precise mathematical sense, see the example in the note below for more information), then for V to be complete the limit of this sequence must also a vector belonging to V. This property plays an important role in the mathematical theory of infinite-dimensional vector spaces.

One can show that any finite-dimensional inner product space is complete and is therefore a Hilbert space. However, not all infinite-dimensional inner product spaces are Hilbert spaces.

☞ An example of an *in*complete inner product space is the space formed by all the functions that are continuous on the closed interval $[0,1]$, defining the inner product as per Eq. (2.45):

$$(f,g) = \int_0^1 f^*(x)g(x)\,\mathrm{d}x. \tag{2.88}$$

Accordingly, the norm $||f||$ of a function $f(x)$ belonging to that space is given by the equation

$$||f||^2 = \int_0^1 |f(x)|^2\,\mathrm{d}x. \tag{2.89}$$

Now, consider the following sequence of functions: $f_1(x) = x$, $f_2(x) = x^{1/2}$, $f_3(x) = x^{1/3}$, $f_4(x) = x^{1/4}$,..., $f_n(x) = x^{1/n}$,... All these functions are continuous on $[0,1]$. A short calculation shows that

$$||f_n - f_m||^2 = \left(\frac{2}{n}+1\right)^{-1} + \left(\frac{2}{m}+1\right)^{-1} - 2\left(\frac{1}{n}+\frac{1}{m}+1\right)^{-1}. \tag{2.90}$$

Thus $||f_n - f_m||$ goes to zero for n and $m \to \infty$, which means that the sequence converges (loosely speaking, the difference between the functions $f_n(x)$ and $f_m(x)$ becomes vanishingly small when n and m increase). If you have taken a course in Analysis, you may have recognized that these functions actually form a Cauchy sequence, which is the mathematically rigorous way convergence is defined in this context: for any positive number ϵ one can find an integer N such that $||f_m - f_n|| < \epsilon$ for all m and $n > N$. Although this sequence converges and only contains functions continuous everywhere on $[0,1]$, however, the function it converges to is *not* continuous on $[0,1]$: the limit of $x^{1/n}$ for $n \to \infty$ is indeed the discontinuous function

$$f(x) = \begin{cases} 1 & 0 < x \leq 1, \\ 0 & x = 0. \end{cases} \tag{2.91}$$

Therefore the vector space of all the functions continuous on $[0,1]$ and equipped with this inner product is not complete and does not qualify as a Hilbert space.

One needs to enlarge the space in order to complete it, in particular by including functions that are not continuous everywhere. How to do this is well outside the scope of the course, but the result can be stated relatively simply: the space of *all* square-integrable functions on an interval $[a, b]$ or on the infinite interval $(-\infty, \infty)$ is complete with respect to the inner products defined by Eqs. (2.44) and (2.45). (For this result to hold, however, these integrals must be understood as being Lebesgue integrals — see page 13 of these notes. Recall that square-integrable functions are called L^2 functions.) The mathematical theory of the corresponding Hilbert spaces underpins the whole of wave quantum mechanics.

☞ As seen in previous sections, in a finite-dimensional vector space V any linear combination of vectors of V is an element of V and any element of V can be written in a unique way as a linear combination of given basis vectors. This is also the case in many important infinite-dimensional Hilbert spaces, although with significant differences.

Let us take, for example, the Hilbert space of L^2 functions on the interval $[0, 2\pi]$. We have seen that the complex exponentials $\exp(inx)$ ($n = 0, \pm 1, \pm 2, \ldots$) are mutually orthogonal on that interval. We can thus form sequences of linear combinations of an increasingly large number of such functions, e.g., $s_0(x)$, $s_1(x)$, $s_2(x)$,..., with

$$s_N(x) = \sum_{n=-N}^{N} c_n \exp(inx). \tag{2.92}$$

Each of these functions is a linear combination of a finite number of linearly-independent L^2 functions, and is therefore a L^2 function. However, this sequence of $s_N(x)$ functions converges to a well defined L^2 function $s(x)$ only if the coefficients c_n make this possible (e.g., these coefficients should go to zero for $n \to \pm\infty$). Furthermore, converging to $s(x)$ here means that $||s_N(x) - s(x)|| \to 0$ for $N \to \infty$, not that $s_N(x) \to s(x)$ at any value of x. However, within this definition of convergence, we write

$$s(x) = \sum_n c_n \exp(inx), \tag{2.93}$$

being understood that this equality does not necessarily hold at all values of x. It turns out that for any L^2 function on $[0, 2\pi]$ one can find a set of coefficients c_n such that the function can be expanded in such a way. In this sense, the complex exponentials $\exp(inx)$ form a basis for this Hilbert space. (There would be much more to say about the concept of basis on infinite-dimensional vector space, but saying much more would go well beyond the scope of this course.)