2 Vectors and Linear Maps

The familiar two- and three-dimensional vectors can easily be generalized to higher dimensions. Representing vectors by their components, one can conceive of vectors having N components. This is the most immediate generalization of vectors in the plane and in space, and such vectors are called N-dimensional **Cartesian** vectors. Cartesian vectors are limited in two respects: Their components are real, and their dimensionality is finite. Some applications in physics require the removal of one or both of these limitations. It is therefore convenient to study vectors stripped of any dimensionality or reality of components. Such properties become consequences of more fundamental definitions. Although we will be concentrating on finite-dimensional vector spaces in this part of the book, many of the concepts and examples introduced here apply to infinite-dimensional spaces as well.

Cartesian vectors

2.1 Vector Spaces

Let us begin with the definition of an abstract (complex) vector space.[1]

Definition 2.1.1 A **vector space** $\mathcal{V}$ over $\mathbb{C}$ is a set of objects denoted by $|a\rangle$, $|b\rangle$, $|x\rangle$, and so on, called **vectors**, with the following properties:

vector space defined

1. To every pair of vectors $|a\rangle$ and $|b\rangle$ in $\mathcal{V}$ there corresponds a vector $|a\rangle + |b\rangle$, also in $\mathcal{V}$, called the *sum* of $|a\rangle$ and $|b\rangle$, such that
   (a) $|a\rangle + |b\rangle = |b\rangle + |a\rangle$,
   (b) $|a\rangle + (|b\rangle + |c\rangle) = (|a\rangle + |b\rangle) + |c\rangle$,
   (c) There exists a unique vector $|0\rangle \in \mathcal{V}$, called the **zero vector**, such that $|a\rangle + |0\rangle = |a\rangle$ for every vector $|a\rangle$,
   (d) To every vector $|a\rangle \in \mathcal{V}$ there corresponds a unique vector $-|a\rangle \in \mathcal{V}$ such that $|a\rangle + (-|a\rangle) = |0\rangle$.

[1] Keep in mind that $\mathbb{C}$ is the set of complex numbers and $\mathbb{R}$ the set of reals.

S. Hassani, *Mathematical Physics*, DOI 10.1007/978-3-319-01195-0_2,

scalars are numbers

2. To every complex number[2] α—also called a **scalar**—and every vector $|a\rangle$ there corresponds a vector $\alpha|a\rangle$ in $\mathcal{V}$ such that
   (a) $\alpha(\beta|a\rangle) = (\alpha\beta)|a\rangle$,
   (b) $1|a\rangle = |a\rangle$.
3. Multiplication involving vectors and scalars is distributive:
   (a) $\alpha(|a\rangle + |b\rangle) = \alpha|a\rangle + \alpha|b\rangle$.
   (b) $(\alpha + \beta)|a\rangle = \alpha|a\rangle + \beta|a\rangle$.

Dirac's bra and ket notation

The **bra**, $\langle\,|$, and **ket**, $|\,\rangle$, notation for vectors, invented by Dirac, is very useful when dealing with complex vector spaces. However, it is somewhat clumsy for certain topics such as norm and metrics and will therefore be abandoned in those discussions.

complex versus real vector space

The vector space defined above is also called a **complex vector space**. It is possible to replace $\mathbb{C}$ with $\mathbb{R}$—the set of real numbers—in which case the resulting space will be called a **real vector space**.

concept of field summarized

Real and complex numbers are prototypes of a mathematical structure called **field**. A field $\mathbb{F}$ is a set of objects with two binary operations called addition and multiplication. Multiplication distributes over addition, and each operation has an identity. The identity for addition is denoted by 0 and is called *additive identity*. The identity for multiplication is denoted by 1 and is called *multiplicative identity*. Furthermore, every element $\alpha \in \mathbb{F}$ has an additive inverse $-\alpha$, and every element except the additive identity has a multiplicative inverse α^{-1}.

Example 2.1.2 (Some vector spaces)

1. $\mathbb{R}$ is a vector space over the field of real numbers.
2. $\mathbb{C}$ is a vector space over the field of real numbers.
3. $\mathbb{C}$ is a vector space over the complex numbers.
4. Let $\mathcal{V} = \mathbb{R}$ and let the field of scalars be $\mathbb{C}$. This is *not* a vector space, because property 2 of Definition 2.1.1 is not satisfied: A complex number times a real number is not a real number and therefore does not belong to $\mathcal{V}$.
5. The set of "arrows" in the plane (or in space) forms a vector space over $\mathbb{R}$ under the parallelogram law of addition of planar (or spatial) vectors.
6. Let $\mathcal{P}^c[t]$ be the set of all polynomials with complex coefficients in a variable t. Then $\mathcal{P}^c[t]$ is a vector space under the ordinary addition of polynomials and the multiplication of a polynomial by a complex number. In this case the zero vector is the zero polynomial.
7. For a given positive integer n, let $\mathcal{P}^c_n[t]$ be the set of all polynomials with complex coefficients of degree less than or equal to n. Again it is easy to verify that $\mathcal{P}^c_n[t]$ is a vector space under the usual addition

[2]Complex numbers, particularly when they are treated as *variables*, are usually denoted by z, and we shall adhere to this convention in Part III. However, in the discussion of vector spaces, we have found it more convenient to use lower case Greek letters to denote complex numbers as scalars.

of polynomials and their multiplication by complex scalars. In particular, the sum of two polynomials of degree less than or equal to n is also a polynomial of degree less than or equal to n, and multiplying a polynomial with complex coefficients by a complex number gives another polynomial of the same type. Here the zero polynomial is the zero vector.

8. The set $\mathcal{P}_n^r[t]$ of polynomials of degree less than or equal to n with real coefficients is a vector space over the reals, but it is *not* a vector space over the complex numbers.
9. Let $\mathbb{C}^n$ consist of all complex n-tuples such as $|a\rangle = (\alpha_1, \alpha_2, \dots, \alpha_n)$ and $|b\rangle = (\beta_1, \beta_2, \dots, \beta_n)$. Let η be a complex number. Then we define

$$\begin{aligned} |a\rangle + |b\rangle &= (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n), \\ \eta|a\rangle &= (\eta\alpha_1, \eta\alpha_2, \dots, \eta\alpha_n), \\ |0\rangle &= (0, 0, \dots, 0), \\ -|a\rangle &= (-\alpha_1, -\alpha_2, \dots, -\alpha_n). \end{aligned}$$

It is easy to verify that $\mathbb{C}^n$ is a vector space over the complex numbers. It is called the *n-dimensional complex coordinate space*.

n-dimensional complex coordinate space

10. The set of all real n-tuples $\mathbb{R}^n$ is a vector space over the real numbers under the operations similar to that of $\mathbb{C}^n$. It is called the *n-dimensional real coordinate space*, or *Cartesian n-space*. It is not a vector space over the complex numbers.

n-dimensional real coordinate space, or Cartesian n-space

11. The set of all complex matrices with m rows and n columns $\mathcal{M}^{m\times n}$ is a vector space under the usual addition of matrices and multiplication by complex numbers. The zero vector is the $m \times n$ matrix with all entries equal to zero.
12. Let $\mathbb{C}^\infty$ be the set of all complex sequences $|a\rangle = \{\alpha_i\}_{i=1}^\infty$ such that $\sum_{i=1}^\infty |\alpha_i|^2 < \infty$. One can show that with addition and scalar multiplication defined componentwise, $\mathbb{C}^\infty$ is a vector space over the complex numbers.
13. The set of all complex-valued functions of a single real variable that are continuous in the real interval (a, b) is a vector space over the complex numbers.
14. The set $\mathcal{C}^n(a, b)$ of all real-valued functions of a single real variable defined on (a, b) that possess continuous derivatives of all orders up to n forms a vector space over the reals.
15. The set $\mathcal{C}^\infty(a, b)$ of all real-valued functions on (a, b) of a single real variable that possess derivatives of all orders forms a vector space over the reals.

It is clear from the example above that a vector space depends as much on the nature of the vectors as on the nature of the scalars.

Definition 2.1.3 The vectors $|a_1\rangle, |a_2\rangle, \dots, |a_n\rangle$, are said to be **linearly independent** if for $\alpha_i \in \mathbb{C}$, the relation $\sum_{i=1}^n \alpha_i |a_i\rangle = 0$ implies $\alpha_i = 0$ for all i. The sum $\sum_{i=1}^n \alpha_i |a_i\rangle$ is called a **linear combination** of $\{|a_i\rangle\}_{i=1}^n$.

linear independence and linear combination of vectors defined

2.1.1 Subspaces

Given a vector space $\mathcal{V}$, one can consider a collection $\mathcal{W}$ of vectors in $\mathcal{V}$, i.e., a subset of $\mathcal{V}$. Because $\mathcal{W}$ is a subset, it contains vectors, but there is no guarantee that it contains the linear combination of those vectors. We now investigate conditions under which it does.

subspace

Definition 2.1.4 A **subspace** $\mathcal{W}$ of a vector space $\mathcal{V}$ is a nonempty subset of $\mathcal{V}$ with the property that if $|a\rangle, |b\rangle \in \mathcal{W}$, then $\alpha|a\rangle + \beta|b\rangle$ also belongs to $\mathcal{W}$ for all $\alpha, \beta \in \mathbb{C}$.

The intersection of two subspaces is also a subspace.

The reader may verify that a subspace is a vector space in its own right, and that *the intersection of two subspaces is also a subspace*.

Example 2.1.5 The following are subspaces of some of the vector spaces considered in Example 2.1.2. The reader is urged to verify the validity of each case.

- The "space" of real numbers is a subspace of $\mathbb{C}$ over the reals.
- $\mathbb{R}$ is *not* a subspace of $\mathbb{C}$ over the complex numbers, because as explained in Example 2.1.2, $\mathbb{R}$ cannot be a vector space over the complex numbers.
- The set of all vectors along a given line *going through the origin* is a subspace of arrows in the plane (or space) over $\mathbb{R}$.
- $\mathcal{P}_n^c[t]$ is a subspace of $\mathcal{P}^c[t]$.
- $\mathbb{C}^{n-1}$ is a subspace of $\mathbb{C}^n$ when $\mathbb{C}^{n-1}$ is identified with all complex n-tuples with zero last entry. In general, $\mathbb{C}^m$ is a subspace of $\mathbb{C}^n$ for $m < n$ when $\mathbb{C}^m$ is identified with all n-tuples whose last $n - m$ elements are zero.
- $\mathcal{M}^{r\times s}$ is a subspace of $\mathcal{M}^{m\times n}$ for $r \le m$ and $s \le n$. Here, we identify an $r \times s$ matrix with an $m \times n$ matrix whose last $m - r$ rows and $n - s$ columns are all zero.
- $\mathcal{P}_m^c[t]$ is a subspace of $\mathcal{P}_n^c[t]$ for $m < n$.
- $\mathcal{P}_m^r[t]$ is a subspace of $\mathcal{P}_n^r[t]$ for $m < n$. Note that both $\mathcal{P}_n^r[t]$ and $\mathcal{P}_m^r[t]$ are vector spaces over the reals only.
- $\mathbb{R}^m$ is a subspace of $\mathbb{R}^n$ for $m < n$. Therefore, $\mathbb{R}^2$, the plane, is a subspace of $\mathbb{R}^3$, the Euclidean space. Also, $\mathbb{R}^1 \equiv \mathbb{R}$ is a subspace of both the plane $\mathbb{R}^2$ and the Euclidean space $\mathbb{R}^3$.
- Let $\mathbf{a}$ be along the x-axis (a subspace of $\mathbb{R}^2$) and $\mathbf{b}$ along the y-axis (also a subspace of $\mathbb{R}^2$). Then $\mathbf{a} + \mathbf{b}$ is neither along the x-axis nor along the y-axis. This shows that the union of two subspaces is not generally a subspace.

union of two subspaces is not a subspace

span of a subset of a vector space

Theorem 2.1.6 *If S is any nonempty set of vectors in a vector space $\mathcal{V}$, then the set $\mathcal{W}_S$ of all linear combinations of vectors in S is a subspace of $\mathcal{V}$. We say that $\mathcal{W}_S$ is the **span of** S, or that S* spans *$\mathcal{W}_S$, or that $\mathcal{W}_S$* is spanned by *S. $\mathcal{W}_S$ is often denoted by* Span$\{S\}$.

The proof of Theorem 2.1.6 is left as Problem 2.6.

Definition 2.1.7 A **basis** of a vector space $\mathcal{V}$ is a set B of linearly independent vectors that spans all of $\mathcal{V}$. A vector space that has a finite basis is called **finite-dimensional**; otherwise, it is **infinite-dimensional**.

basis defined

The definition of the dimensionality of a vector space based on a single basis makes sense because of the following theorem which we state without proof (see [Axle 96, page 31]):

Theorem 2.1.8 *All bases of a given finite-dimensional vector space have the same number of linearly independent vectors.*

Definition 2.1.9 The cardinality of a basis of a vector space $\mathcal{V}$ is called the **dimension** of $\mathcal{V}$ and denoted by $\dim \mathcal{V}$. To emphasize its dependence on the scalars, $\dim_{\mathbb{C}} \mathcal{V}$ and $\dim_{\mathbb{R}} \mathcal{V}$ are also used. A vector space of dimension N is sometimes denoted by $\mathcal{V}_N$.

If $|a\rangle$ is a vector in an N-dimensional vector space $\mathcal{V}$ and $B = \{|a_i\rangle\}_{i=1}^N$ a basis in that space, then by the definition of a basis, there exists a unique (see Problem 2.4) set of scalars $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that $|a\rangle = \sum_{i=1}^N \alpha_i |a_i\rangle$. The set $\{\alpha_i\}_{i=1}^N$ is called the **components** of $|a\rangle$ in the basis B.

components of a vector in a basis

Example 2.1.10 The following are bases for the vector spaces given in Example 2.1.2.

- The number 1 (or any nonzero real number) is a basis for $\mathbb{R}$, which is therefore one-dimensional.
- The numbers 1 and $i = \sqrt{-1}$ (or any pair of distinct nonzero complex numbers) are basis vectors for the vector space $\mathbb{C}$ over $\mathbb{R}$. Thus, this space is two-dimensional.
- The number 1 (or any nonzero complex number) is a basis for $\mathbb{C}$ over $\mathbb{C}$, and the space is one-dimensional. Note that although the vectors are the same as in the preceding item, changing the nature of the scalars changes the dimensionality of the space.
- The set $\{\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z\}$ of the unit vectors in the directions of the three axes forms a basis in space. The space is three-dimensional.
- A basis of $\mathcal{P}^c[t]$ can be formed by the monomials $1, t, t^2, \ldots$. It is clear that this space is *infinite-dimensional*.
- A basis of $\mathbb{C}^n$ is given by $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_n$, where $\hat{\mathbf{e}}_j$ is an n-tuple that has a 1 at the jth position and zeros everywhere else. This basis is called the **standard basis** of $\mathbb{C}^n$. Clearly, the space has n dimensions.

standard basis of $\mathbb{C}^n$

- A basis of $\mathcal{M}^{m\times n}$ is given by $\mathsf{e}_{11}, \mathsf{e}_{12}, \ldots, \mathsf{e}_{ij}, \ldots, \mathsf{e}_{mn}$, where e_{ij} is the $m \times n$ matrix with zeros everywhere except at the intersection of the ith row and jth column, where it has a one.
- A set consisting of the monomials $1, t, t^2, \ldots, t^n$ forms a basis of $\mathcal{P}_n^c[t]$. Thus, this space is $(n+1)$-dimensional.
- The standard basis of $\mathbb{C}^n$ is a basis of $\mathbb{R}^n$ as well. It is also called the *standard basis* of $\mathbb{R}^n$. Thus, $\mathbb{R}^n$ is n-dimensional.

- If we assume that $a < 0 < b$, then the set of monomials $1, x, x^2, \ldots$ forms a basis for $\mathcal{C}^\infty(a, b)$, because, by Taylor's theorem, any function belonging to $\mathcal{C}^\infty(a, b)$ can be expanded in an infinite power series about $x = 0$. Thus, this space is infinite-dimensional.

Remark 2.1.1 Given a space $\mathcal{V}$ with a basis $B = \{|a_i\rangle\}_{i=1}^n$, the span of any m vectors ($m < n$) of B is an m-dimensional subspace of $\mathcal{V}$.

2.1.2 Factor Space

Let $\mathcal{W}$ be a subspace of the vector space $\mathcal{V}$, and define a relation on $\mathcal{V}$ as follows. If $|a\rangle \in \mathcal{V}$ and $|b\rangle \in \mathcal{V}$, then we say that $|a\rangle$ is related to $|b\rangle$, and write $|a\rangle \bowtie |b\rangle$ if $|a\rangle - |b\rangle$ is in $\mathcal{W}$. It is easy to show that $\bowtie$ is an equivalence relation. Denote the equivalence class of $|a\rangle$ by $[\![a]\!]$, and the factor set (or quotient set) $\{[\![a]\!] \,|\, |a\rangle \in \mathcal{V}\}$ by $\mathcal{V}/\mathcal{W}$. We turn the factor *set* into a factor *space* by defining the combined addition of vectors and their multiplication by scalars as follows:

$$\alpha[\![a]\!] + \beta[\![b]\!] = [\![\alpha a + \beta b]\!] \tag{2.1}$$

where $[\![\alpha a + \beta b]\!]$ is the equivalence class of $\alpha|a\rangle + \beta|b\rangle$. For this equation to make sense, it must be independent of the choice of the representatives of the classes. If $[\![a']\!] = [\![a]\!]$ and $[\![b']\!] = [\![b]\!]$, then is it true that $[\![\alpha a' + \beta b']\!] = [\![\alpha a + \beta b]\!]$? For this to happen, we must have

$$\big(\alpha|a'\rangle + \beta|b'\rangle\big) - \big(\alpha|a\rangle + \beta|b\rangle\big) \in \mathcal{W}.$$

Now, since $|a'\rangle \in [\![a]\!]$, we must have $|a'\rangle = |a\rangle + |w_1\rangle$ for some $|w_1\rangle \in \mathcal{W}$. Similarly, $|b'\rangle = |b\rangle + |w_2\rangle$. Therefore,

$$\big(\alpha|a'\rangle + \beta|b'\rangle\big) - \big(\alpha|a\rangle + \beta|b\rangle\big) = \alpha|w_1\rangle + \beta|w_2\rangle$$

and the right-hand side is in $\mathcal{W}$ because $\mathcal{W}$ is a subspace.

Sometimes $[\![a]\!]$ is written as $|a\rangle + \mathcal{W}$. With this notation comes the equalities

$$|w\rangle + \mathcal{W} = \mathcal{W}, \qquad \mathcal{W} + \mathcal{W} = \mathcal{W}, \qquad \alpha\mathcal{W} = \mathcal{W}, \qquad \alpha\mathcal{W} + \beta\mathcal{W} = \mathcal{W},$$

which abbreviate the obvious fact that the sum of two vectors in $\mathcal{W}$ is a vector in $\mathcal{W}$, the product of a scalar and a vector in $\mathcal{W}$ is a vector in $\mathcal{W}$, and the linear combination of two vectors in $\mathcal{W}$ is a vector in $\mathcal{W}$.

How do we find a basis for $\mathcal{V}/\mathcal{W}$? Let $\{|a_i\rangle\}$ be a basis for $\mathcal{W}$. Extend it to a basis $\{|a_i\rangle, |b_j\rangle\}$ for $\mathcal{V}$. Then, $\{[\![b_j]\!]\}$ form a basis for $\mathcal{V}/\mathcal{W}$. Indeed, let $[\![a]\!] \in \mathcal{V}/\mathcal{W}$. Then, since $|a\rangle$ is in $\mathcal{V}$, we have

$$[\![a]\!] \equiv |a\rangle + \mathcal{W} = \overbrace{\sum_i \alpha_i|a_i\rangle}^{\in\mathcal{W}} + \sum_j \beta_j|b_j\rangle + \mathcal{W}$$

$$= \sum_j \beta_j |b_j\rangle + \mathcal{W} \quad \Rightarrow \quad [\![a]\!] = \Big[\!\!\Big[\sum_j \beta_j |b_j\rangle\Big]\!\!\Big] = \sum_j \beta_j [\![b_j]\!].$$

Thus, $\{[\![b_j]\!]\}$ span $\mathcal{V}/\mathcal{W}$. To form a basis, they also have to be linearly independent. So, suppose that $\sum_j \beta_j [\![b_j]\!] = [\![0]\!]$. This means that

$$\sum_j \beta_j |b_j\rangle + \mathcal{W} = |0\rangle + \mathcal{W} = \mathcal{W} \quad \Rightarrow \quad \sum_j \beta_j |b_j\rangle \in \mathcal{W}.$$

So the last sum must be a linear combination of $\{|a_i\rangle\}$:

$$\sum_j \beta_j |b_j\rangle = \sum_i \alpha_i |a_j\rangle \quad \text{or} \quad \sum_j \beta_j |b_j\rangle - \sum_i \alpha_i |a_j\rangle = 0.$$

This is a zero linear combination of the basis vectors of $\mathcal{V}$. Therefore, all coefficients, including all β_j must be zero. One consequence of the argument above is (with obvious notation)

$$\dim(\mathcal{V}/\mathcal{W}) = \dim \mathcal{V} - \dim \mathcal{W} \tag{2.2}$$

2.1.3 Direct Sums

Sometimes it is possible, and convenient, to break up a vector space into special (disjoint) subspaces. For instance, the study of the motion of a particle in $\mathbb{R}^3$ under the influence of a central force field is facilitated by decomposing the motion into its projections onto the direction of angular momentum and onto a plane perpendicular to the angular momentum. This corresponds to decomposing a vector in space into a vector, say in the xy-plane and a vector along the z-axis. We can generalize this to any vector space, but first some notation: Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of a vector space $\mathcal{V}$. Denote by $\mathcal{U} + \mathcal{W}$ the collection of all vectors in $\mathcal{V}$ that can be written as a sum of two vectors, one in $\mathcal{U}$ and one in $\mathcal{W}$. It is easy to show that $\mathcal{U} + \mathcal{W}$ is a subspace of $\mathcal{V}$.

Sum of two subspaces defined

Example 2.1.11 Let $\mathcal{U}$ be the xy-plane and $\mathcal{W}$ the yz-plane. These are both subspaces of $\mathbb{R}^3$, and so is $\mathcal{U} + \mathcal{W}$. In fact, $\mathcal{U} + \mathcal{W} = \mathbb{R}^3$, because given any vector (x, y, z) in $\mathbb{R}^3$, we can write it as

$$(x, y, z) = \underbrace{\left(x, \frac{1}{2}y, 0\right)}_{\in \mathcal{U}} + \underbrace{\left(0, \frac{1}{2}y, z\right)}_{\in \mathcal{W}}.$$

This decomposition is not unique: We could also write $(x, y, z) = (x, \frac{1}{3}y, 0) + (0, \frac{2}{3}y, z)$, and a host of other relations.

Definition 2.1.12 Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of a vector space $\mathcal{V}$ such that $\mathcal{V} = \mathcal{U} + \mathcal{W}$ and $\mathcal{U} \cap \mathcal{W} = \{|0\rangle\}$. Then we say that $\mathcal{V}$ is the **direct sum** of $\mathcal{U}$ and $\mathcal{W}$ and write $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$.

direct sum $\mathcal{U} \oplus \mathcal{W}$ defined uniqueness of direct sum

Proposition 2.1.13 *Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of $\mathcal{V}$ such that $\mathcal{V} = \mathcal{U} + \mathcal{W}$. Then $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$ if and only if any nonzero vector in $\mathcal{V}$ can be written uniquely as a vector in $\mathcal{U}$ plus a vector in $\mathcal{W}$.*

Proof Assume $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$, and let $|v\rangle \in \mathcal{V}$ be written as a sum of a vector in $\mathcal{U}$ and a vector in $\mathcal{W}$ *in two different ways*:

$$|v\rangle = |u\rangle + |w\rangle = |u'\rangle + |w'\rangle \quad \Leftrightarrow \quad |u\rangle - |u'\rangle = |w'\rangle - |w\rangle.$$

The LHS is in $\mathcal{U}$. Since it is equal to the RHS—which is in $\mathcal{W}$—it must be in $\mathcal{W}$ as well. Therefore, the LHS must equal zero, as must the RHS. Thus, $|u\rangle = |u'\rangle$, $|w'\rangle = |w\rangle$, and there is only one way that $|v\rangle$ can be written as a sum of a vector in $\mathcal{U}$ and a vector in $\mathcal{W}$.

Conversely, suppose that any vector in $\mathcal{V}$ can be written uniquely as a vector in $\mathcal{U}$ and a vector in $\mathcal{W}$. If $|a\rangle \in \mathcal{U}$ and also $|a\rangle \in \mathcal{W}$, then one can write

$$|a\rangle = \underbrace{\frac{1}{3}|a\rangle}_{\text{in } \mathcal{U}} + \underbrace{\frac{2}{3}|a\rangle}_{\text{in } \mathcal{W}} = \underbrace{\frac{1}{4}|a\rangle}_{\text{in } \mathcal{U}} + \underbrace{\frac{3}{4}|a\rangle}_{\text{in } \mathcal{W}}.$$

Hence $|a\rangle$ can be written in two different ways. By the uniqueness assumption $|a\rangle$ cannot be nonzero. Therefore, the only vector common to both $\mathcal{U}$ and $\mathcal{W}$ is the zero vector. This implies that $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$. □

More generally, we have the following situation:

Definition 2.1.14 Let $\{\mathcal{U}_i\}_{i=1}^r$ be subspaces of $\mathcal{V}$ such that

$$\mathcal{V} = \mathcal{U}_1 + \cdots + \mathcal{U}_r \quad \text{and} \quad \mathcal{U}_i \cap \mathcal{U}_j = \{|0\rangle\} \quad \text{for all } i, j = 1, \ldots, r.$$

Then we say that $\mathcal{V}$ is the direct sum of $\{\mathcal{U}_i\}_{i=1}^r$ and write

$$\mathcal{V} = \mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_r = \bigoplus_{i=1}^r \mathcal{U}_i.$$

Let $\mathcal{W} = \mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_s$ be a direct sum of s subspaces (they need not span the entire $\mathcal{V}$). Write $\mathcal{W}$ as $\mathcal{W} = \mathcal{U}_1 \oplus \mathcal{W}'$, with $\mathcal{W}' = \mathcal{U}_2 \oplus \cdots \oplus \mathcal{U}_s$. Let $\{|u_i\rangle\}_{i=1}^s$ be nonzero vectors with $|u_i\rangle \in \mathcal{U}_i$ and suppose that

$$\alpha_1|u_1\rangle + \alpha_2|u_2\rangle + \cdots + \alpha_s|u_s\rangle = |0\rangle, \tag{2.3}$$

or

$$\alpha_1|u_1\rangle + \alpha|w'\rangle = |0\rangle \quad \Rightarrow \quad \alpha_1|u_1\rangle = -\alpha|w'\rangle,$$

with $|w'\rangle \in \mathcal{W}'$. Since $\alpha_1|u_1\rangle \in \mathcal{U}_1$ from the left-hand side, and $\alpha_1|u_1\rangle \in \mathcal{W}'$ from the right-hand side, we must have $\alpha_1|u_1\rangle = |0\rangle$. Hence, $\alpha_1 = 0$ because $|u_1\rangle \neq |0\rangle$. Equation (2.3) now becomes

$$\alpha_2|u_2\rangle + \alpha_3|u_3\rangle + \cdots + \alpha_s|u_s\rangle = |0\rangle.$$

Write this as

$$\alpha_2|u_2\rangle + \beta\big|w''\big\rangle = |0\rangle \quad \Rightarrow \quad \alpha_2|u_2\rangle = -\beta\big|w''\big\rangle,$$

where $\mathcal{W}' = \mathcal{U}_2 \oplus \mathcal{W}''$ with $\mathcal{W}'' = \mathcal{U}_3 \oplus \cdots \oplus \mathcal{U}_s$ and $|w''\rangle \in \mathcal{W}''$. An argument similar to above shows that $\alpha_2 = 0$. Continuing in this way, we have

Proposition 2.1.15 *The vectors in different subspaces of Definition 2.1.14 are linearly independent.*

Proposition 2.1.16 *Let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Then there exist a subspace $\mathcal{W}$ of $\mathcal{V}$ such that $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$.*

Proof Let $\{|u_i\rangle\}_{i=1}^m$ be a basis of $\mathcal{U}$. Extend this basis to a basis $\{|u_i\rangle\}_{i=1}^N$ of $\mathcal{V}$. Then $\mathcal{W} = \text{Span}\{|u_j\rangle\}_{j=m+1}^N$. □

Example 2.1.17 Let $\mathcal{U}$ be the xy-plane and $\mathcal{W}$ the z-axis. These are both subspaces of $\mathbb{R}^3$, and so is $\mathcal{U}+\mathcal{W}$. Furthermore, it is clear that $\mathcal{U}+\mathcal{W} = \mathbb{R}^3$, because given any vector (x, y, z) in $\mathbb{R}^3$, we can write it as

$$(x, y, z) = \underbrace{(x, y, 0)}_{\in\mathcal{U}} + \underbrace{(0, 0, z)}_{\in\mathcal{W}}.$$

This decomposition is obviously unique. Therefore, $\mathbb{R}^3 = \mathcal{U} \oplus \mathcal{W}$.

Proposition 2.1.18 *If $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$, then* $\dim\mathcal{V} = \dim\mathcal{U} + \dim\mathcal{W}$.

dimensions in a direct sum

Proof Let $\{|u_i\rangle\}_{i=1}^m$ be a basis for $\mathcal{U}$ and $\{|w_i\rangle\}_{i=1}^k$ a basis for $\mathcal{W}$. Then it is easily verified that $\{|u_1\rangle, |u_2\rangle, \ldots, |u_m\rangle, |w_1\rangle, |w_2\rangle, \ldots, |w_k\rangle\}$ is a basis for $\mathcal{V}$. The details are left as an exercise. □

Let $\mathcal{U}$ and $\mathcal{V}$ be any two vector spaces over $\mathbb{R}$ or $\mathbb{C}$. Consider the Cartesian product $\mathcal{W} \equiv \mathcal{U} \times \mathcal{V}$ of their underlying set. Define a scalar multiplication and a sum on $\mathcal{W}$ by

$$\begin{aligned} \alpha\big(|u\rangle, |v\rangle\big) &= \big(\alpha|u\rangle, \alpha|v\rangle\big) \\ \big(|u_1\rangle, |v_1\rangle\big) + \big(|u_2\rangle, |v_2\rangle\big) &= \big(|u_1\rangle + |u_2\rangle, |v_1\rangle + |v_2\rangle\big). \end{aligned} \tag{2.4}$$

With $|0\rangle_W = (|0\rangle_U, |0\rangle_V)$, $\mathcal{W}$ becomes a vector space. Furthermore, if we identify $\mathcal{U}$ and $\mathcal{V}$ with vectors of the form $(|u\rangle, |0\rangle_V)$ and $(|0\rangle_U, |v\rangle)$, respectively, then $\mathcal{U}$ and $\mathcal{V}$ become subspaces of $\mathcal{W}$. If a vector $|w\rangle \in \mathcal{W}$ belongs to both $\mathcal{U}$ and $\mathcal{V}$, then it can be written as both $(|u\rangle, |0\rangle_V)$ and $(|0\rangle_U, |v\rangle)$, i.e., $(|u\rangle, |0\rangle_V) = (|0\rangle_U, |v\rangle)$. But this can happen only if $|u\rangle = |0\rangle_U$ and $|v\rangle = |0\rangle_V$, or $|w\rangle = |0\rangle_W$. Thus, the only common vector in $\mathcal{U}$ and $\mathcal{V}$ is the zero vector. Therefore,

Proposition 2.1.19 *Let $\mathcal{U}$ and $\mathcal{V}$ be any two vector spaces over $\mathbb{R}$ or $\mathbb{C}$. Then their Cartesian product $\mathcal{W} \equiv \mathcal{U} \times \mathcal{V}$ together with the operations de-*

fined in Eq. (2.4) *becomes a vector space. Furthermore,* $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$ *if* $\mathcal{U}$ *and* $\mathcal{V}$ *are identified with vectors of the form* $(|u\rangle, |0\rangle_V)$ *and* $(|0\rangle_U, |v\rangle)$, *respectively.*

Let $\{|a_i\rangle\}_{i=1}^M$ be a basis of $\mathcal{U}$ and $\{|b_j\rangle\}_{j=1}^N$ a basis of $\mathcal{V}$. Define the vectors $\{|c_k\rangle\}_{k=1}^{M+N}$ in $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$ by

$$\begin{aligned} |c_k\rangle &= \big(|a_k\rangle, |0\rangle_V\big) && \text{if } 1 \le k \le M \\ |c_k\rangle &= \big(|0\rangle_U, |b_{k-M}\rangle\big) && \text{if } M+1 \le k \le M+N. \end{aligned} \tag{2.5}$$

Then $\{|c_k\rangle\}_{k=1}^{M+N}$ are linearly independent. In fact,

$$\sum_{k=1}^{M+N} \gamma_k |c_k\rangle = |0\rangle_W \quad \text{iff}$$

$$\sum_{k=1}^{M} \gamma_k \big(|a_k\rangle, |0\rangle_V\big) + \sum_{j=1}^{N} \gamma_{M+j} \big(|0\rangle_U, |b_j\rangle\big) = \big(|0\rangle_U, |0\rangle_V\big),$$

or

$$\left(\sum_{k=1}^{M} \gamma_k |a_k\rangle, |0\rangle_V\right) + \left(|0\rangle_U, \sum_{j=1}^{N} \gamma_{M+j} |b_j\rangle\right) = \big(|0\rangle_U, |0\rangle_V\big),$$

or

$$\left(\sum_{k=1}^{M} \gamma_k |a_k\rangle, \sum_{j=1}^{N} \gamma_{M+j} |b_j\rangle\right) = \big(|0\rangle_U, |0\rangle_V\big),$$

or

$$\sum_{k=1}^{M} \gamma_k |a_k\rangle = |0\rangle_U \quad \text{and} \quad \sum_{j=1}^{N} \gamma_{M+j} |b_j\rangle = |0\rangle_V.$$

Linear independence of $\{|a_i\rangle\}_{i=1}^M$ and $\{|b_j\rangle\}_{j=1}^N$ imply that $\gamma_k = 0$ for $1 \le k \le M+N$.

It is not hard to show that $\mathcal{W} = \mathrm{Span}\{|c_k\rangle\}_{k=1}^{M+N}$. Hence, we have the following

Theorem 2.1.20 *Let* $\{|a_i\rangle\}_{i=1}^M$ *be a basis of* $\mathcal{U}$ *and* $\{|b_j\rangle\}_{j=1}^N$ *a basis of* $\mathcal{V}$. *The set of vectors* $\{|c_k\rangle\}_{k=1}^{M+N}$ *defined by Eq.* (2.5) *form a basis of the direct sum* $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$. *In particular,* $\mathcal{W}$ *has dimension* $M + N$.

2.1.4 Tensor Product of Vector Spaces

Direct sum is one way of constructing a new vector space out of two. There is another procedure. Let $\mathcal{U}$ and $\mathcal{V}$ be vector spaces. On their Cartesian product, impose the scalar product and bilinearity conditions:

$$\begin{aligned}
\alpha\big(|u\rangle,|v\rangle\big) &= \big(\alpha|u\rangle,|v\rangle\big) = \big(|u\rangle,\alpha|v\rangle\big) \\
\big(\alpha_1|u_1\rangle+\alpha_2|u_2\rangle,|v\rangle\big) &= \alpha_1\big(|u_1\rangle,|v\rangle\big)+\alpha_2\big(|u_2\rangle,|v\rangle\big) \\
\big(|u\rangle,\beta_1|v_1\rangle+\beta_2|v_2\rangle\big) &= \beta_1\big(|u\rangle,|v_1\rangle\big)+\beta_2\big(|u\rangle,|v_2\rangle\big).
\end{aligned} \tag{2.6}$$

These properties turn $\mathcal{U}\times\mathcal{V}$ into a vector space called the **tensor product** of $\mathcal{U}$ and $\mathcal{V}$ and denoted by $\mathcal{U}\otimes\mathcal{V}$.[3] The vectors in the tensor product space are denoted by $|u\rangle\otimes|v\rangle$, (or occasionally by $|uv\rangle$). If $\{|a_i\rangle\}_{i=1}^M$ and $\{|b_j\rangle\}_{j=1}^N$ are bases in $\mathcal{U}$ and $\mathcal{V}$, respectively, and

$$|u\rangle=\sum_{i=1}^M\alpha_i|a_i\rangle \quad \text{and} \quad |v\rangle=\sum_{j=1}^N\beta_j|b_j\rangle,$$

then Eq. (2.6) yields

$$|u\rangle\otimes|v\rangle=\left(\sum_{i=1}^M\alpha_i|a_i\rangle\right)\otimes\left(\sum_{j=1}^N\beta_j|b_j\rangle\right)=\sum_{i=1}^M\sum_{j=1}^N\alpha_i\beta_j|a_i\rangle\otimes|b_j\rangle.$$

Therefore, $\{|a_i\rangle\otimes|b_j\rangle\}$ is a basis of $\mathcal{U}\otimes\mathcal{V}$ and $\dim(\mathcal{U}\otimes\mathcal{V})=\dim\mathcal{U}\dim\mathcal{V}$.

From (2.6), we have

$$|0\rangle_U\otimes|v\rangle=\big(|u\rangle-|u\rangle\big)\otimes|v\rangle=|u\rangle\otimes|v\rangle-|u\rangle\otimes|v\rangle=|0\rangle_{U\otimes V}$$

Similarly, $|u\rangle\otimes|0\rangle_V=|0\rangle_{U\otimes V}$.

2.2 Inner Product

A vector space, as given by Definition 2.1.1, is too general and structureless to be of much physical interest. One useful structure introduced on a vector space is a scalar product. Recall that the scalar (dot) product of vectors in the plane or in space is a rule that associates with two vectors **a** and **b**, a real number. This association, denoted symbolically by $g:\mathcal{V}\times\mathcal{V}\to\mathbb{R}$, with $g(\mathbf{a},\mathbf{b})=\mathbf{a}\cdot\mathbf{b}$, is symmetric: $g(\mathbf{a},\mathbf{b})=g(\mathbf{b},\mathbf{a})$, is linear in the first (and by symmetry, in the second) factor:[4]

$$g(\alpha\mathbf{a}+\beta\mathbf{b},\mathbf{c})=\alpha g(\mathbf{a},\mathbf{c})+\beta g(\mathbf{b},\mathbf{c}) \quad \text{or} \quad (\alpha\mathbf{a}+\beta\mathbf{b})\cdot\mathbf{c}=\alpha\mathbf{a}\cdot\mathbf{c}+\beta\mathbf{b}\cdot\mathbf{c},$$

gives the "length" of a vector: $|\mathbf{a}|^2=g(\mathbf{a},\mathbf{a})=\mathbf{a}\cdot\mathbf{a}\ge 0$, and ensures that the only vector with zero length[5] is the zero vector: $g(\mathbf{a},\mathbf{a})=0$ if and only if $\mathbf{a}=\mathbf{0}$.

[3] A detailed discussion of tensor products and tensors in general is given in Chap. 26.

[4] A function that is linear in both of its arguments is called a **bilinear** function.

[5] In our present discussion, we are avoiding situations in which a nonzero vector can have zero "length". Such occasions arise in relativity, and we shall discuss them in Part VIII.

We want to generalize these properties to abstract vector spaces for which the scalars are complex numbers. A verbatim generalization of the foregoing properties, however, leads to a contradiction. Using the linearity in both arguments and a nonzero $|a\rangle$, we obtain

$$g\big(i|a\rangle, i|a\rangle\big) = i^2 g\big(|a\rangle, |a\rangle\big) = -g\big(|a\rangle, |a\rangle\big). \tag{2.7}$$

Either the right-hand side (RHS) or left-hand side (LHS) of this equation must be negative! But this is inconsistent with the positivity of the "length" of a vector, which requires $g(|a\rangle, |a\rangle)$ to be positive for *all* nonzero vectors, including $i|a\rangle$. The source of the problem is the linearity in *both* arguments. If we can change this property in such a way that one of the i's in Eq. (2.7) comes out complex-conjugated, the problem may go away. This requires linearity in one argument and complex-conjugate linearity in the other. Which argument is to be complex-conjugate linear is a matter of convention. We choose the first argument to be so.[6] We thus have

$$g\big(\alpha|a\rangle + \beta|b\rangle, |c\rangle\big) = \alpha^* g\big(|a\rangle, |c\rangle\big) + \beta^* g\big(|b\rangle, |c\rangle\big),$$

where α^* denotes the complex conjugate. Consistency then requires us to change the symmetry property as well. In fact, we must demand that $g(|a\rangle, |b\rangle) = (g(|b\rangle, |a\rangle))^*$, from which the *reality* of $g(|a\rangle, |a\rangle)$—a necessary condition for its positivity—follows immediately.

The question of the existence of an inner product on a vector space is a deep problem in higher analysis. Generally, if an inner product exists, there may be many ways to introduce one on a vector space. However, as we shall see in Sect. 2.2.4, a *finite-dimensional* vector space always has an inner product and this inner product is unique.[7] So, for all practical purposes we can speak of *the* inner product on a finite-dimensional vector space, and as with the two- and three-dimensional cases, we can omit the letter g and use a notation that involves only the vectors. There are several such notations in use, but the one that will be employed in this book is the *Dirac bra(c)ket notation*, whereby $g(|a\rangle, |b\rangle)$ is denoted by $\langle a|b\rangle$. Using this notation, we have

Dirac "bra," $\langle\ |$, and "ket" $|\ \rangle$, notation is used for inner products.

inner product defined

Definition 2.2.1 The **inner product** of two vectors, $|a\rangle$ and $|b\rangle$, in a vector space $\mathcal{V}$ is a complex number, $\langle a|b\rangle \in \mathbb{C}$, such that

1. $\langle a|b\rangle = \langle b|a\rangle^*$
2. $\langle a|(\beta|b\rangle + \gamma|c\rangle) = \beta\langle a|b\rangle + \gamma\langle a|c\rangle$
3. $\langle a|a\rangle \geq 0$, and $\langle a|a\rangle = 0$ if and only if $|a\rangle = |0\rangle$.

positive definite, or Euclidean inner product

The last relation is called the **positive definite** property of the inner prod-

[6] In some books, particularly in the mathematical literature, the second argument is chosen to be conjugate linear.

[7] This uniqueness holds up to a certain equivalence of inner products that we shall not get into here.

uct.[8] A positive definite real inner product is also called a **Euclidean** inner product, otherwise it is called **pseudo-Euclidean**.

Note that linearity in the first argument is absent in the definition above, because, as explained earlier, it would be inconsistent with the first property, which expresses the "symmetry" of the inner product. The extra operation of complex conjugation renders the true linearity in the first argument impossible. Because of this complex conjugation, the inner product on a complex vector space is not truly bilinear; it is commonly called **sesquilinear** or **hermitian**.

sesquilinear or hermitian inner product

A shorthand notation will be useful when dealing with the inner product of a linear combination of vectors.

Box 2.2.2 *We write the LHS of the second equation in the definition above as* $\langle a|\beta b+\gamma c\rangle$.

This has the advantage of treating a linear combination as a single vector. The second property then states that if the complex scalars happen to be in a *ket*, they "split out" unaffected:

$$\langle a|\beta b+\gamma c\rangle = \beta\langle a|b\rangle + \gamma\langle a|c\rangle. \tag{2.8}$$

On the other hand, if the complex scalars happen to be in the first factor (the bra), then they should be conjugated when they are "split out":

$$\langle \beta b+\gamma c|a\rangle = \beta^*\langle b|a\rangle + \gamma^*\langle c|a\rangle. \tag{2.9}$$

A vector space $\mathcal{V}$ on which an inner product is defined is called an **inner product space**. As mentioned above, a finite-dimensional vector space can always be turned into an inner product space.

Example 2.2.3 In this example, we introduce some of the most common inner products. The reader is urged to verify that in all cases, we indeed have an inner product.

natural inner product for $\mathbb{C}^n$

- Let $|a\rangle, |b\rangle \in \mathbb{C}^n$, with $|a\rangle = (\alpha_1, \alpha_2, \dots, \alpha_n)$ and $|b\rangle = (\beta_1, \beta_2, \dots, \beta_n)$, and define an inner product on $\mathbb{C}^n$ as

$$\langle a|b\rangle \equiv \alpha_1^*\beta_1 + \alpha_2^*\beta_2 + \cdots + \alpha_n^*\beta_n = \sum_{i=1}^n \alpha_i^*\beta_i.$$

  That this product satisfies all the required properties of an inner product is easily checked. For example, if $|b\rangle = |a\rangle$, we obtain $\langle a|a\rangle = |\alpha_1|^2 + |\alpha_2|^2 + \cdots + |\alpha_n|^2$, which is clearly nonnegative.

[8] The positive definiteness must be relaxed in the space-time of relativity theory, in which nonzero vectors can have zero "length".

- Similarly, for $|a\rangle, |b\rangle \in \mathbb{R}^n$ the same definition (without the complex conjugation) satisfies all the properties of an inner product.
- For $|a\rangle, |b\rangle \in \mathbb{C}^\infty$ the natural inner product is defined as $\langle a|b\rangle = \sum_{i=1}^{\infty} \alpha_i^* \beta_i$. The question of the convergence of this infinite sum is the subject of Problem 2.18.
- Let $x(t), y(t) \in \mathcal{P}^c[t]$, the space of all polynomials in t with complex coefficients. Define

weight function of an inner product defined in terms of integrals

$$\langle x|y\rangle \equiv \int_a^b w(t)x^*(t)y(t)\,dt, \tag{2.10}$$

where a and b are real numbers—or infinity—for which the integral exists, and $w(t)$, called the **weight function**, is a real-valued, continuous function that is *always strictly positive* in the interval (a, b). Then Eq. (2.10) defines an inner product. Depending on the weight function $w(t)$, there can be many different inner products defined on the infinite-dimensional space $\mathcal{P}^c[t]$.
- Let $f, g \in \mathbb{C}(a, b)$ and define their inner product by

natural inner product for complex functions

$$\langle f|g\rangle \equiv \int_a^b w(x)f^*(x)g(x)\,dx.$$

It is easily shown that $\langle f|g\rangle$ satisfies all the requirements of the inner product if, as in the previous case, the weight function $w(x)$ is always positive in the interval (a, b). This is called the *standard inner product* on $\mathbb{C}(a, b)$.

2.2.1 Orthogonality

The vectors of analytic geometry and calculus are often expressed in terms of unit vectors along the axes, i.e., vectors that are of unit length and perpendicular to one another. Such vectors are also important in abstract inner product spaces.

orthogonality defined

orthonormal basis

Definition 2.2.4 Vectors $|a\rangle, |b\rangle \in \mathcal{V}$ are **orthogonal** if $\langle a|b\rangle = 0$. A **normal** vector, or *normalized vector*, $|e\rangle$ is one for which $\langle e|e\rangle = 1$. A basis $B = \{|e_i\rangle\}_{i=1}^N$ in an N-dimensional vector space $\mathcal{V}$ is an **orthonormal basis** if

$$\langle e_i|e_j\rangle = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \tag{2.11}$$

Kronecker delta

where δ_{ij}, defined by the last equality, is called the **Kronecker delta**.

Example 2.2.5 Let $\mathcal{U}$ and $\mathcal{V}$ be inner product vector spaces. Let $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$. Then an inner product can be defined on $\mathcal{W}$ in terms of those on $\mathcal{U}$ and $\mathcal{V}$. In fact, it can be easily shown that if $|w_i\rangle = (|u_i\rangle, |v_i\rangle)$, $i = 1, 2$, then

$$\langle w_1|w_2\rangle = \langle u_1|u_2\rangle + \langle v_1|v_2\rangle \tag{2.12}$$

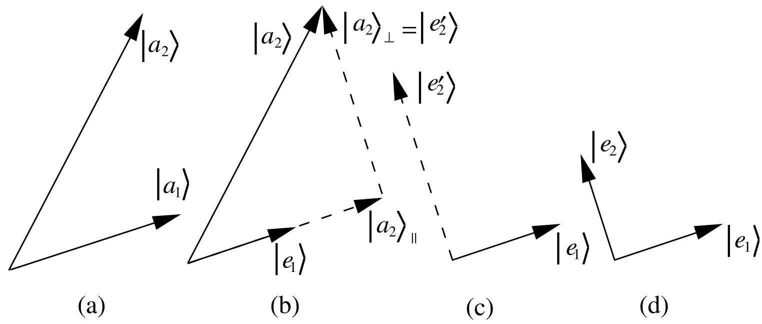

Fig. 2.1 The essence of the Gram–Schmidt process is neatly illustrated by the process in two dimensions. This figure, depicts the stages of the construction of two orthonormal vectors

defines an inner product on $\mathcal{W}$. Moreover, with the identification

$$\mathcal{U} = \big\{\big(|u\rangle, |0\rangle_V\big) \mid |u\rangle \in \mathcal{U}\big\} \quad \text{and} \quad \mathcal{V} = \big\{\big(|0\rangle_U, |v\rangle\big) \mid |v\rangle \in \mathcal{V}\big\},$$

any vector in $\mathcal{U}$ is orthogonal to any vector in $\mathcal{V}$.

Example 2.2.6 Here are examples of orthonormal bases:

- The standard basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$)

$$|e_1\rangle = (1, 0, \dots, 0),\ |e_2\rangle = (0, 1, \dots, 0),\ \dots,\ |e_n\rangle = (0, 0, \dots, 1)$$

  is orthonormal under the usual inner product of those spaces.
- Let $|e_k\rangle = e^{ikx}/\sqrt{2\pi}$ be functions in $\mathbb{C}(0, 2\pi)$ with $w(x) = 1$. Then

$$\langle e_k|e_k\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{-ikx}e^{ikx}\,dx = 1,$$

  and for $l \neq k$,

$$\langle e_l|e_k\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{-ilx}e^{ikx}\,dx = \frac{1}{2\pi}\int_0^{2\pi} e^{i(k-l)x}\,dx = 0.$$

  Thus, $\langle e_l|e_k\rangle = \delta_{lk}$.

2.2.2 The Gram-Schmidt Process

It is always possible to convert—by taking appropriate linear combinations—any basis in $\mathcal{V}$ into an orthonormal basis. A process by which this may be accomplished is called **Gram–Schmidt orthonormalization**. Consider a basis $B = \{|a_i\rangle\}_{i=1}^N$. We intend to take linear combinations of $|a_i\rangle$ in such a way that the resulting vectors are orthonormal. First, we let $|e_1\rangle = |a_1\rangle/\sqrt{\langle a_1|a_1\rangle}$ and note that $\langle e_1|e_1\rangle = 1$. If we subtract from $|a_2\rangle$ its projection along $|e_1\rangle$, we obtain a vector that is orthogonal to $|e_1\rangle$ (see Fig. 2.1).

The Gram–Schmidt process explained

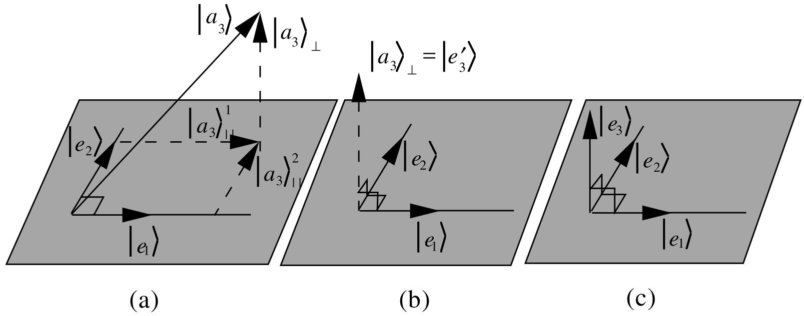

Fig. 2.2 Once the orthonormal vectors in the plane of two vectors are obtained, the third orthonormal vector is easily constructed

Calling the resulting vector $|e_2'\rangle$, we have $|e_2'\rangle = |a_2\rangle - \langle e_1|a_2\rangle|e_1\rangle$, which can be written more symmetrically as $|e_2'\rangle = |a_2\rangle - |e_1\rangle\langle e_1|a_2\rangle$. Clearly, this vector is orthogonal to $|e_1\rangle$. In order to normalize $|e_2'\rangle$, we divide it by $\sqrt{\langle e_2'|e_2'\rangle}$. Then $|e_2\rangle = |e_2'\rangle/\sqrt{\langle e_2'|e_2'\rangle}$ will be a normal vector orthogonal to $|e_1\rangle$. Subtracting from $|a_3\rangle$ its projections along the first and second unit vectors obtained so far will give the vector

$$|e_3'\rangle = |a_3\rangle - |e_1\rangle\langle e_1|a_3\rangle - |e_2\rangle\langle e_2|a_3\rangle = |a_3\rangle - \sum_{i=1}^{2} |e_i\rangle\langle e_i|a_3\rangle,$$

which is orthogonal to both $|e_1\rangle$ and $|e_2\rangle$ (see Fig. 2.2):

$$\langle e_1|e_3'\rangle = \langle e_1|a_3\rangle - \overbrace{\langle e_1|e_1\rangle}^{=1}\langle e_1|a_3\rangle - \overbrace{\langle e_1|e_2\rangle}^{=0}\langle e_2|a_3\rangle = 0.$$

Similarly, $\langle e_2|e_3'\rangle = 0$.

Historical Notes

Erhard Schmidt
1876–1959

Erhard Schmidt (1876–1959) obtained his doctorate under the supervision of David Hilbert. His main interest was in integral equations and Hilbert spaces. He is the "Schmidt" of the **Gram–Schmidt orthogonalization process**, which takes a basis of a space and constructs an orthonormal one from it. (Laplace had presented a special case of this process long before Gram or Schmidt.)

In 1908 Schmidt worked on infinitely many equations in infinitely many unknowns, introducing various geometric notations and terms that are still in use for describing spaces of functions. Schmidt's ideas were to lead to the geometry of Hilbert spaces. This was motivated by the study of integral equations (see Chap. 18) and an attempt at their abstraction.

Earlier, Hilbert regarded a function as given by its Fourier coefficients. These satisfy the condition that $\sum_{k=1}^{\infty} a_k^2$ is finite. He introduced sequences of real numbers $\{x_n\}$ such that $\sum_{n=1}^{\infty} x_n^2$ is finite. Riesz and Fischer showed that there is a one-to-one correspondence between square-integrable functions and square-summable sequences of their Fourier coefficients. In 1907 Schmidt and Fréchet showed that a consistent theory could be obtained if the square-summable sequences were regarded as the coordinates of points in an infinite-dimensional space that is a generalization of n-dimensional Euclidean space. Thus *functions can be regarded as points of a space*, now called a **Hilbert space**.

In general, if we have calculated m orthonormal vectors $|e_1\rangle, \dots, |e_m\rangle$, with $m < N$, then we can find the next one using the following relations:

$$\begin{aligned} |e'_{m+1}\rangle &= |a_{m+1}\rangle - \sum_{i=1}^{m} |e_i\rangle\langle e_i|a_{m+1}\rangle, \\ |e_{m+1}\rangle &= \frac{|e'_{m+1}\rangle}{\sqrt{\langle e'_{m+1}|e'_{m+1}\rangle}}. \end{aligned} \tag{2.13}$$

Even though we have been discussing finite-dimensional vector spaces, the process of Eq. (2.13) can continue for infinite-dimensions as well. The reader is asked to pay attention to the fact that, at each stage of the Gram–Schmidt process, one is taking linear combinations of the original vectors.

2.2.3 The Schwarz Inequality

Let us now consider an important inequality that is valid in both finite and infinite dimensions and whose restriction to two and three dimensions is equivalent to the fact that the cosine of the angle between two vectors is always less than one.

Schwarz inequality

Theorem 2.2.7 *For any pair of vectors $|a\rangle$, $|b\rangle$ in an inner product space $\mathcal{V}$, the* ***Schwarz inequality*** *holds*: $\langle a|a\rangle\langle b|b\rangle \geq |\langle a|b\rangle|^2$. *Equality holds when $|a\rangle$ is proportional to $|b\rangle$.*

Proof Let $|c\rangle = |b\rangle - (\langle a|b\rangle/\langle a|a\rangle)|a\rangle$, and note that $\langle a|c\rangle = 0$. Write $|b\rangle = (\langle a|b\rangle/\langle a|a\rangle)|a\rangle + |c\rangle$ and take the inner product of $|b\rangle$ with itself:

$$\langle b|b\rangle = \left|\frac{\langle a|b\rangle}{\langle a|a\rangle}\right|^2 \langle a|a\rangle + \langle c|c\rangle = \frac{|\langle a|b\rangle|^2}{\langle a|a\rangle} + \langle c|c\rangle.$$

Since $\langle c|c\rangle \geq 0$, we have

$$\langle b|b\rangle \geq \frac{|\langle a|b\rangle|^2}{\langle a|a\rangle} \Rightarrow \langle a|a\rangle\langle b|b\rangle \geq \left|\langle a|b\rangle\right|^2.$$

Equality holds iff $\langle c|c\rangle = 0$, i.e., iff $|c\rangle = 0$. From the definition of $|c\rangle$, we conclude that for the equality to hold, $|a\rangle$ and $|b\rangle$ must be proportional. □

Notice the power of abstraction: We have derived the Schwarz inequality solely from the basic assumptions of inner product spaces independent of the specific nature of the inner product. Therefore, we do not have to prove the Schwarz inequality every time we encounter a new inner product space.

Historical Notes

Karl Herman Amandus Schwarz (1843–1921) the son of an architect, was born in what is now Sobiecin, Poland. After gymnasium, Schwarz studied chemistry in Berlin for a time before switching to mathematics, receiving his doctorate in 1864. He was greatly influenced by the reigning mathematicians in Germany at the time, especially Kummer

Karl Herman Amandus Schwarz 1843–1921

and Weierstrass. The lecture notes that Schwarz took while attending Weierstrass's lectures on the integral calculus still exist. Schwarz received an initial appointment at Halle and later appointments in Zurich and Göttingen before being named as Weierstrass's successor at Berlin in 1892. These later years, filled with students and lectures, were not Schwarz's most productive, but his early papers assure his place in mathematics history.

Schwarz's favorite tool was geometry, which he soon turned to the study of analysis. He conclusively proved some of Riemann's results that had been previously (and justifiably) challenged. The primary result in question was the assertion that every simply connected region in the plane could be conformally mapped onto a circular area. From this effort came several well-known results now associated with Schwarz's name, including the principle of reflection and Schwarz's lemma. He also worked on surfaces of minimal area, the branch of geometry beloved by all who dabble with soap bubbles.

Schwarz's most important work, for the occasion of Weierstrass's seventieth birthday, again dealt with minimal area, specifically whether a minimal surface yields a minimal area. Along the way, Schwarz demonstrated second variation in a multiple integral, constructed a function using successive approximation, and demonstrated the existence of a "least" eigenvalue for certain differential equations. This work also contained the most famous inequality in mathematics, which bears his name.

Schwarz's success obviously stemmed from a matching of his aptitude and training to the mathematical problems of the day. One of his traits, however, could be viewed as either positive or negative—his habit of treating all problems, whether trivial or monumental, with the same level of attention to detail. This might also at least partly explain the decline in productivity in Schwarz's later years.

Schwarz had interests outside mathematics, although his marriage was a mathematical one, since he married Kummer's daughter. Outside mathematics he was the captain of the local voluntary fire brigade, and he assisted the stationmaster at the local railway station by closing the doors of the trains!

2.2.4 Length of a Vector

In dealing with objects such as directed line segments in the plane or in space, the intuitive idea of the length of a vector is used to define the dot product. However, sometimes it is more convenient to introduce the inner product first and then define the length, as we shall do now.

norm of a vector defined

Definition 2.2.8 The **norm**, or *length*, of a vector $|a\rangle$ in an inner product space is denoted by $\|a\|$ and defined as $\|a\| \equiv \sqrt{\langle a|a\rangle}$. We use the notation $\|\alpha a + \beta b\|$ for the norm of the vector $\alpha|a\rangle + \beta|b\rangle$.

One can easily show that the norm has the following properties:

1. The norm of the zero vector is zero: $\|0\| = 0$.
2. $\|a\| \geq 0$, and $\|a\| = 0$ if and only if $|a\rangle = |0\rangle$.
3. $\|\alpha a\| = |\alpha|\|a\|$ for any[9] complex α.
4. $\|a + b\| \leq \|a\| + \|b\|$. This property is called the **triangle inequality**.

triangle inequality

Any function on a vector space satisfying the four properties above is called a **norm**, and the vector space on which a norm is defined is called a **normed linear space**. One does not need an inner product to have a norm.

normed linear space

One can introduce the idea of the "distance" between two vectors in a normed linear space. The distance between $|a\rangle$ and $|b\rangle$—denoted by $d(a, b)$—is simply the norm of their difference: $d(a, b) \equiv \|a - b\|$. It can

natural distance in a normed linear space

[9]The first property follows from this by letting $\alpha = 0$.

be readily shown that this has all the properties one expects of the distance (or metric) function introduced in Chap. 1. However, one does not need a normed space to define distance. For example, as explained in Chap. 1, one can define the distance between two points on the surface of a sphere, but the addition of two points on a sphere—a necessary operation for vector space structure—is not defined. Thus the points on a sphere form a metric space, but not a vector space.

Inner product spaces are automatically normed spaces, but the converse is not, in general, true: There are normed spaces, i.e., spaces satisfying properties 1–4 above that cannot be promoted to inner product spaces. However, if the norm satisfies the **parallelogram law**,

parallelogram law

$$\|a+b\|^2+\|a-b\|^2=2\|a\|^2+2\|b\|^2, \tag{2.14}$$

then one can define

$$\langle a|b\rangle \equiv \frac{1}{4}\bigl\{\|a+b\|^2-\|a-b\|^2-i\bigl(\|a+ib\|^2-\|a-ib\|^2\bigr)\bigr\} \tag{2.15}$$

and show that it is indeed an inner product. In fact, we have (see [Frie 82, pp. 203–204] for a proof) the following theorem.

Theorem 2.2.9 *A normed linear space is an inner product space if and only if the norm satisfies the parallelogram law.*

Now consider any N-dimensional vector space $\mathcal{V}$. Choose a basis $\{|a_i\rangle\}_{i=1}^N$ in $\mathcal{V}$, and for any vector $|a\rangle$ whose components are $\{\alpha_i\}_{i=1}^N$ in this basis, define

$$\|a\|^2 \equiv \sum_{i=1}^N |\alpha_i|^2.$$

The reader may check that this defines a norm, and that the norm satisfies the parallelogram law. From Theorem 2.2.9 we have the following:

Theorem 2.2.10 *Every finite-dimensional vector space can be turned into an inner product space.*

$\mathbb{C}^n$ has many different distance functions

Example 2.2.11 Let the space be $\mathbb{C}^n$. The natural inner product of $\mathbb{C}^n$ gives rise to a norm, which, for the vector $|a\rangle=(\alpha_1,\alpha_2,\dots,\alpha_n)$ is

$$\|a\|=\sqrt{\langle a|a\rangle}=\sqrt{\sum_{i=1}^n |\alpha_i|^2}.$$

This norm yields the following distance between $|a\rangle$ and $|b\rangle=(\beta_1,\beta_2,\dots,\beta_n)$:

$$d(a,b)=\|a-b\|=\sqrt{\langle a-b|a-b\rangle}=\sqrt{\sum_{i=1}^n |\alpha_i-\beta_i|^2}.$$

One can define other norms, such as $\|a\|_1 \equiv \sum_{i=1}^n |\alpha_i|$, which has all the required properties of a norm, and leads to the distance

$$d_1(a, b) = \|a - b\|_1 = \sum_{i=1}^{n} |\alpha_i - \beta_i|.$$

Another norm defined on $\mathbb{C}^n$ is given by

$$\|a\|_p \equiv \left(\sum_{i=1}^{n} |\alpha_i|^p\right)^{1/p},$$

where p is a positive integer. It is proved in higher mathematical analysis that $\|\cdot\|_p$ has all the properties of a norm. (The nontrivial part of the proof is to verify the triangle inequality.) The associated distance is

$$d_p(a, b) = \|a - b\|_p = \left(\sum_{i=1}^{n} |\alpha_i - \beta_i|^p\right)^{1/p}.$$

The other two norms introduced above are special cases, for $p = 2$ and $p = 1$.

2.3 Linear Maps

We have made progress in enriching vector spaces with structures such as norms and inner products. However, this enrichment, although important, will be of little value if it is imprisoned in a single vector space. We would like to give vector space properties freedom of movement, so they can go from one space to another. The vehicle that carries these properties is a linear map or linear transformation which is the subject of this section. First it is instructive to review the concept of a map (discussed in Chap. 1) by considering some examples relevant to the present discussion.

Example 2.3.1 The following are a few familiar examples of mappings.

1. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$.
2. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be given by $g(x, y) = x^2 + y^2 - 4$.
3. Let $F : \mathbb{R}^2 \to \mathbb{C}$ be given by $F(x, y) = U(x, y) + iV(x, y)$, where $U : \mathbb{R}^2 \to \mathbb{R}$ and $V : \mathbb{R}^2 \to \mathbb{R}$.
4. Let $T : \mathbb{R} \to \mathbb{R}^2$ be given by $T(t) = (t + 3, 2t - 5)$.
5. Motion of a point particle in space can be considered as a mapping $M : [a, b] \to \mathbb{R}^3$, where $[a, b]$ is an interval of the real line. For each $t \in [a, b]$, we define $M(t) = (x(t), y(t), z(t))$, where $x(t)$, $y(t)$, and $z(t)$ are real-valued functions of t. If we identify t with time, which is assumed to have a value in the interval $[a, b]$, then $M(t)$ describes the path of the particle as a function of time, and a and b are the beginning and the end of the motion, respectively.

Let us consider an arbitrary mapping $F : \mathcal{V} \to \mathcal{W}$ from a vector space $\mathcal{V}$ to another vector space $\mathcal{W}$. It is assumed that the two vector spaces are over the same scalars, say $\mathbb{C}$. Consider $|a\rangle$ and $|b\rangle$ in $\mathcal{V}$ and $|x\rangle$ and $|y\rangle$ in $\mathcal{W}$ such that $F(|a\rangle) = |x\rangle$ and $F(|b\rangle) = |y\rangle$. In general, F does not preserve the vector space structure. That is, the image of a linear combination of vectors is not the same as the linear combination of the images:

$$F\big(\alpha|a\rangle + \beta|b\rangle\big) \neq \alpha F\big(|x\rangle\big) + \beta F\big(|y\rangle\big).$$

This is the case for all the mappings of Example 2.3.1. There are many applications in which the preservation of the vector space structure (preservation of the linear combination) is desired.

linear map (or transformation), linear operator, endomorphism

Definition 2.3.2 A **linear map** (or **transformation**) from the complex vector space $\mathcal{V}$ to the complex vector space $\mathcal{W}$ is a mapping $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ such that

$$\mathbf{T}\big(\alpha|a\rangle + \beta|b\rangle\big) = \alpha\mathbf{T}\big(|a\rangle\big) + \beta\mathbf{T}\big(|b\rangle\big) \quad \forall|a\rangle, |b\rangle \in \mathcal{V} \text{ and } \alpha, \beta \in \mathbb{C}.$$

A linear transformation $\mathbf{T} : \mathcal{V} \to \mathcal{V}$ is called an **endomorphism** of $\mathcal{V}$ or a **linear operator** on $\mathcal{V}$. The action of a linear transformation on a vector is written without the parentheses: $\mathbf{T}(|a\rangle) \equiv \mathbf{T}|a\rangle$.

The same definition applies to real vector spaces. Note that the definition demands that both vector spaces have the same set of scalars: The same scalars must multiply vectors in $\mathcal{V}$ on the LHS and those in $\mathcal{W}$ on the RHS.

The set of linear maps from $\mathcal{V}$ to $\mathcal{W}$ is denoted by $\mathcal{L}(\mathcal{V}, \mathcal{W})$, and this set happens to be a vector space. The *zero transformation*, $\mathbf{0}$, is defined to take every vector in $\mathcal{V}$ to the zero vector of $\mathcal{W}$. The sum of two linear transformations $\mathbf{T}$ and $\mathbf{U}$ is the linear transformation $\mathbf{T} + \mathbf{U}$, whose action on a vector $|a\rangle \in \mathcal{V}$ is defined to be $(\mathbf{T} + \mathbf{U})|a\rangle \equiv \mathbf{T}|a\rangle + \mathbf{U}|a\rangle$. Similarly, define $\alpha\mathbf{T}$ by $(\alpha\mathbf{T})|a\rangle \equiv \alpha(\mathbf{T}|a\rangle) = \alpha\mathbf{T}|a\rangle$. The set of endomorphisms of $\mathcal{V}$ is denoted by $\mathcal{L}(\mathcal{V})$ or $\mathrm{End}(\mathcal{V})$ rather than $\mathcal{L}(\mathcal{V}, \mathcal{V})$. We summarize these observations in

$\mathcal{L}(\mathcal{V}, \mathcal{W})$ is a vector space

Box 2.3.3 *$\mathcal{L}(\mathcal{V}, \mathcal{W})$ is a vector space. In particular, so is the set of endomorphisms of a single vector space $\mathcal{L}(\mathcal{V}) \equiv \mathrm{End}(\mathcal{V}) \equiv \mathcal{L}(\mathcal{V}, \mathcal{V})$.*

Definition 2.3.4 Let $\mathcal{V}$ and $\mathcal{U}$ be inner product spaces. A linear map $\mathbf{T} : \mathcal{V} \to \mathcal{U}$ is called an **isometric** map if[10]

$$\langle \mathbf{T}a|\mathbf{T}b\rangle = \langle a|b\rangle, \quad \forall|a\rangle, |b\rangle \in \mathcal{V}.$$

[10] It is convenient here to use the notation $|\mathbf{T}a\rangle$ for $\mathbf{T}|a\rangle$. This would then allow us to write the dual (see below) of the vector as $\langle\mathbf{T}a|$, emphasizing that it is indeed the bra associated with $\mathbf{T}|a\rangle$.

isometry

If $\mathcal{U} = \mathcal{V}$, then $\mathbf{T}$ is called a linear **isometry** or simply an isometry of $\mathcal{V}$. It is common to call an isometry of a complex (real) $\mathcal{V}$ a **unitary** (**orthogonal**) operator

Example 2.3.5 The following are some examples of linear operators in various vector spaces. The proofs of linearity are simple in all cases and are left as exercises for the reader.

1. Let $\mathcal{V}$ be a one-dimensional space (e.g., $\mathcal{V} = \mathbb{C}$). Then any linear endomorphism $\mathbf{T}$ of $\mathcal{V}$ is of the form $\mathbf{T}|x\rangle = \alpha|x\rangle$ with α a scalar. In particular, if $\mathbf{T}$ is an isometry, then $|\alpha|^2 = 1$. If $\mathcal{V} = \mathbb{R}$ and $\mathbf{T}$ is an isometry, then $\mathbf{T}|x\rangle = \pm|x\rangle$.
2. Let π be a permutation (shuffling) of the integers $\{1, 2, \dots, n\}$. If $|x\rangle = (\eta_1, \eta_2, \dots, \eta_n)$ is a vector in $\mathbb{C}^n$, we can write
$$\mathbf{A}_\pi|x\rangle = (\eta_{\pi(1)}, \eta_{\pi(2)}, \dots, \eta_{\pi(n)}).$$
Then $\mathbf{A}_\pi$ is a linear operator.
3. For any $|x\rangle \in \mathcal{P}^c[t]$, with $x(t) = \sum_{k=0}^n \alpha_k t^k$, write $|y\rangle = \mathbf{D}|x\rangle$, where $|y\rangle$ is defined as $y(t) = \sum_{k=1}^n k\alpha_k t^{k-1}$. Then $\mathbf{D}$ is a linear operator, the **derivative operator**.

derivative operator

4. For every $|x\rangle \in \mathcal{P}^c[t]$, with $x(t) = \sum_{k=0}^n \alpha_k t^k$, write $|y\rangle = \mathbf{S}|x\rangle$, where $|y\rangle \in \mathcal{P}^c[t]$ is defined as $y(t) = \sum_{k=0}^n [\alpha_k/(k+1)]t^{k+1}$. Then $\mathbf{S}$ is a linear operator, the **integration operator**.

integration operator

5. Let $\mathcal{C}^n(a, b)$ be the set of real-valued functions defined in the interval $[a, b]$ whose first n derivatives exist and are continuous. For any $|f\rangle \in \mathcal{C}^n(a, b)$ define $|u\rangle = \mathbf{G}|f\rangle$, with $u(t) = g(t)f(t)$ and $g(t)$ a fixed function in $\mathcal{C}^n(a, b)$. Then $\mathbf{G}$ is linear. In particular, the operation of multiplying by t, whose operator is denoted by $\mathbf{T}$, is linear.

An immediate consequence of Definition 2.3.2 is the following:

Box 2.3.6 *Two linear transformations* $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ *and* $\mathbf{U} : \mathcal{V} \to \mathcal{W}$ *are equal if and only if* $\mathbf{T}|a_i\rangle = \mathbf{U}|a_i\rangle$ *for all* $|a_i\rangle$ *in some basis of* $\mathcal{V}$. *Thus, a linear transformation is uniquely determined by its action on some basis of its domain space.*

The equality in this box is simply the set-theoretic equality of maps discussed in Chap. 1.

The equality of operators can also be established by other, more convenient, methods when an inner product is defined on the vector space. The following two theorems contain the essence of these alternatives.

Theorem 2.3.7 *An endomorphism* $\mathbf{T}$ *of an inner product space is* $\mathbf{0}$ *if and only if* $\langle b|\mathbf{T}|a\rangle \equiv \langle b|\mathbf{T}a\rangle = 0$ *for all* $|a\rangle$ *and* $|b\rangle$.

Proof Clearly, if $\mathbf{T} = \mathbf{0}$ then $\langle b|\mathbf{T}|a\rangle = 0$. Conversely, if $\langle b|\mathbf{T}|a\rangle = 0$ for all $|a\rangle$ and $|b\rangle$, then, choosing $|b\rangle = \mathbf{T}|a\rangle = |\mathbf{T}a\rangle$, we obtain

$$\langle \mathbf{T}a|\mathbf{T}a\rangle = 0 \quad \forall |a\rangle \quad \Leftrightarrow \quad \mathbf{T}|a\rangle = 0 \quad \forall |a\rangle \quad \Leftrightarrow \quad \mathbf{T} = \mathbf{0}$$

by positive definiteness of the inner product. □

Theorem 2.3.8 *A linear operator* $\mathbf{T}$ *on an inner product space is* $\mathbf{0}$ *if and only if* $\langle a|\mathbf{T}|a\rangle = 0$ *for all* $|a\rangle$.

Proof Obviously, if $\mathbf{T} = \mathbf{0}$, then $\langle a|\mathbf{T}|a\rangle = 0$. Conversely, choose a vector $\alpha|a\rangle + \beta|b\rangle$, sandwich $\mathbf{T}$ between this vector and its bra, and rearrange terms to obtain what is known as the **polarization identity**

polarization identity

$$\alpha^*\beta\langle a|\mathbf{T}|b\rangle + \alpha\beta^*\langle b|\mathbf{T}|a\rangle = \langle \alpha a + \beta b|\mathbf{T}|\alpha a + \beta b\rangle \\ - |\alpha|^2\langle a|\mathbf{T}|a\rangle - |\beta|^2\langle b|\mathbf{T}|b\rangle.$$

According to the assumption of the theorem, the RHS is zero. Thus, if we let $\alpha = \beta = 1$ we obtain $\langle a|\mathbf{T}|b\rangle + \langle b|\mathbf{T}|a\rangle = 0$. Similarly, with $\alpha = 1$ and $\beta = i$ we get $i\langle a|\mathbf{T}|b\rangle - i\langle b|\mathbf{T}|a\rangle = 0$. These two equations give $\langle a|\mathbf{T}|b\rangle = 0$ for all $|a\rangle, |b\rangle$. By Theorem 2.3.7, $\mathbf{T} = \mathbf{0}$. □

To show that two operators $\mathbf{U}$ and $\mathbf{T}$ on an inner product space are equal, one can either have them act on an arbitrary vector and show that they give the same result, or one verifies that $\mathbf{U} - \mathbf{T}$ is the zero operator by means of one of the theorems above. Equivalently, one shows that $\langle a|\mathbf{T}|b\rangle = \langle a|\mathbf{U}|b\rangle$ or $\langle a|\mathbf{T}|a\rangle = \langle a|\mathbf{U}|a\rangle$ for all $|a\rangle, |b\rangle$.

2.3.1 Kernel of a Linear Map

It follows immediately from Definition 2.3.2 that the image of the zero vector in $\mathcal{V}$ is the zero vector in $\mathcal{W}$. This is not true for a general mapping, but it is necessarily true for a linear mapping. As the zero vector of $\mathcal{V}$ is mapped onto the zero vector of $\mathcal{W}$, other vectors of $\mathcal{V}$ may also be dragged along. In fact, we have the following theorem.

Theorem 2.3.9 *The set of vectors in* $\mathcal{V}$ *that are mapped onto the zero vector of* $\mathcal{W}$ *under the linear transformation* $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ *form a* subspace *of* $\mathcal{V}$ *called the* ***kernel****, or* ***null space****, of* $\mathbf{T}$ *and denoted by* $\ker \mathbf{T}$.

kernel of a linear transformation

Proof The proof is left as an exercise. □

The dimension of $\ker \mathbf{T}$ is also called the **nullity** of $\mathcal{V}$.

nullity

The proof of the following is also left as an exercise.

Theorem 2.3.10 *The range* $\mathbf{T}(\mathcal{V})$ *of a linear map* $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ *is a subspace of* $\mathcal{W}$. *The dimension of* $\mathbf{T}(\mathcal{V})$ *is called the* ***rank*** *of* $\mathbf{T}$.

rank of a linear transformation

Theorem 2.3.11 *A linear transformation is* 1–1 (*injective*) *iff its kernel is zero.*

Proof The "only if" part is trivial. For the "if" part, suppose $\mathbf{T}|a_1\rangle = \mathbf{T}|a_2\rangle$; then linearity of $\mathbf{T}$ implies that $\mathbf{T}(|a_1\rangle - |a_2\rangle) = 0$. Since $\ker \mathbf{T} = 0$,[11] we must have $|a_1\rangle = |a_2\rangle$. □

Theorem 2.3.12 *A linear isometric map is injective.*

Proof Let $\mathbf{T} : \mathcal{V} \to \mathcal{U}$ be a linear isometry. Let $|a\rangle \in \ker \mathbf{T}$, then

$$\langle a|a\rangle = \langle \mathbf{T}a|\mathbf{T}a\rangle = \langle 0|0\rangle = 0.$$

Therefore, $|a\rangle = |0\rangle$. By Theorem 2.3.11, $\mathbf{T}$ is injective. □

Suppose we start with a basis of $\ker \mathbf{T}$ and add enough linearly independent vectors to it to get a basis for $\mathcal{V}$. Without loss of generality, let us assume that the first n vectors in this basis form a basis of $\ker \mathbf{T}$. So let $B = \{|a_1\rangle, |a_2\rangle, \dots, |a_N\rangle\}$ be a basis for $\mathcal{V}$ and $B' = \{|a_1\rangle, |a_2\rangle, \dots, |a_n\rangle\}$ be a basis for $\ker \mathbf{T}$. Here $N = \dim \mathcal{V}$ and $n = \dim \ker \mathbf{T}$. It is straightforward to show that $\{\mathbf{T}|a_{n+1}\rangle, \dots, \mathbf{T}|a_N\rangle\}$ is a basis for $\mathbf{T}(\mathcal{V})$. We therefore have the following result (see also the end of this subsection).

dimension theorem

Theorem 2.3.13 *Let* $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ *be a linear transformation. Then*[12]

$$\dim \mathcal{V} = \dim \ker \mathbf{T} + \dim \mathbf{T}(\mathcal{V})$$

This theorem is called the **dimension theorem**. One of its consequences is that an injective endomorphism is automatically surjective, and vice versa:

Proposition 2.3.14 *An endomorphism of a finite-dimensional vector space is bijective if it is either injective or surjective.*

The dimension theorem is obviously valid only for finite-dimensional vector spaces. In particular, neither surjectivity nor injectivity implies bijectivity for infinite-dimensional vector spaces.

Example 2.3.15 Let us try to find the kernel of $\mathbf{T} : \mathbb{R}^4 \to \mathbb{R}^3$ given by

$$\mathbf{T}(x_1, x_2, x_3, x_4) = (2x_1 + x_2 + x_3 - x_4, x_1 + x_2 + 2x_3 + 2x_4, x_1 - x_3 - 3x_4).$$

---

[11] Since $\ker \mathbf{T}$ is a set, we should write the equality as $\ker \mathbf{T} = \{|0\rangle\}$, or at least as $\ker \mathbf{T} = |0\rangle$. However, when there is no danger of confusion, we set $\{|0\rangle\} = |0\rangle = 0$.

[12] Recall that the dimension of a vector space depends on the scalars used in that space. Although we are dealing with two different vector spaces here, since they are both over the same set of scalars (complex or real), no confusion in the concept of dimension arises.

We must look for (x_1, x_2, x_3, x_4) such that $\mathbf{T}(x_1, x_2, x_3, x_4) = (0, 0, 0)$, or

$$\begin{aligned} 2x_1 + x_2 + x_3 - x_4 &= 0, \\ x_1 + x_2 + 2x_3 + 2x_4 &= 0, \\ x_1 - x_3 - 3x_4 &= 0. \end{aligned}$$

The "solution" to these equations is $x_1 = x_3 + 3x_4$ and $x_2 = -3x_3 - 5x_4$. Thus, to be in $\ker \mathbf{T}$, a vector in $\mathbb{R}^4$ must be of the form

$$(x_3 + 3x_4, -3x_3 - 5x_4, x_3, x_4) = x_3(1, -3, 1, 0) + x_4(3, -5, 0, 1),$$

where x_3 and x_4 are arbitrary real numbers. It follows that $\ker \mathbf{T}$ consists of vectors that can be written as linear combinations of the two *linearly independent* vectors $(1, -3, 1, 0)$ and $(3, -5, 0, 1)$. Therefore, $\dim \ker \mathbf{T} = 2$. Theorem 2.3.13 then says that $\dim \mathbf{T}(\mathcal{V}) = 2$; that is, the range of $\mathbf{T}$ is two-dimensional. This becomes clear when one notes that

$$\begin{aligned} &\mathbf{T}(x_1, x_2, x_3, x_4) \\ &\quad = (2x_1 + x_2 + x_3 - x_4)(1, 0, 1) + (x_1 + x_2 + 2x_3 + 2x_4)(0, 1, -1), \end{aligned}$$

and therefore $\mathbf{T}(x_1, x_2, x_3, x_4)$, an arbitrary vector in the range of $\mathbf{T}$, is a linear combination of *only* two linearly independent vectors, $(1, 0, 1)$ and $(0, 1, -1)$.

2.3.2 Linear Isomorphism

In many cases, two vector spaces may "look" different, while in reality they are very much the same. For example, the set of complex numbers $\mathbb{C}$ is a two-dimensional vector space *over the reals*, as is $\mathbb{R}^2$. Although we call the vectors of these two spaces by different names, they have very similar properties. This notion of "similarity" is made precise in the following definition.

isomorphism and automorphism

Definition 2.3.16 A vector space $\mathcal{V}$ is said to be **isomorphic** to another vector space $\mathcal{W}$, and written $\mathcal{V} \cong \mathcal{W}$, if there exists a bijective linear map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$. Then $\mathbf{T}$ is called an **isomorphism**.[13] A bijective linear map of $\mathcal{V}$ onto itself is called an **automorphism** of $\mathcal{V}$. An automorphism is also called an **invertible** linear map. The set of automorphisms of $\mathcal{V}$ is denoted by $GL(\mathcal{V})$.

An immediate consequence of the injectivity of an isometry and Proposition 2.3.14 is the following:

[13]The word "isomorphism", as we shall see, is used in conjunction with many algebraic structures. To distinguish them, qualifiers need to be used. In the present context, we speak of **linear isomorphism**. We shall use qualifiers when necessary. However, the context usually makes the meaning of isomorphism clear.

Proposition 2.3.17 *An isometry of a finite-dimensional vector space is an automorphism of that vector space.*

For all practical purposes, two isomorphic vector spaces are different manifestations of the "same" vector space. In the example discussed above, the correspondence $\mathbf{T} : \mathbb{C} \to \mathbb{R}^2$, with $\mathbf{T}(x + iy) = (x, y)$, establishes an isomorphism between the two vector spaces. It should be emphasized that *only as vector spaces* are $\mathbb{C}$ and $\mathbb{R}^2$ isomorphic. If we go beyond the vector space structures, the two sets are quite different. For example, $\mathbb{C}$ has a natural multiplication for its elements, but $\mathbb{R}^2$ does not. The following three theorems give a working criterion for isomorphism. The proofs are simple and left to the reader.

Theorem 2.3.18 *A linear* surjective *map* $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ *is an isomorphism if and only if its nullity is zero.*

Theorem 2.3.19 *An injective linear transformation* $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ *carries linearly independent sets of vectors onto linearly independent sets of vectors.*

Theorem 2.3.20 *Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.*

only two N-dimensional vector spaces

A consequence of Theorem 2.3.20 is that all N-dimensional vector spaces over $\mathbb{R}$ are isomorphic to $\mathbb{R}^N$ and all complex N-dimensional vector spaces are isomorphic to $\mathbb{C}^N$. So, for all practical purposes, we have only two N-dimensional vector spaces, $\mathbb{R}^N$ and $\mathbb{C}^N$.

Suppose that $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$ and that $\mathbf{T}$ is an automorphism of $\mathcal{V}$ which leaves $\mathcal{V}_1$ invariant, i.e., $\mathbf{T}(\mathcal{V}_1) = \mathcal{V}_1$. Then $\mathbf{T}$ leaves $\mathcal{V}_2$ invariant as well. To see this, first note that if $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$ and $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}'_2$, then $\mathcal{V}_2 = \mathcal{V}'_2$. This can be readily established by looking at a basis of $\mathcal{V}$ obtained by extending a basis of $\mathcal{V}_1$. Now note that since $\mathbf{T}(\mathcal{V}) = \mathcal{V}$ and $\mathbf{T}(\mathcal{V}_1) = \mathcal{V}_1$, we must have

$$\mathcal{V}_1 \oplus \mathcal{V}_2 = \mathcal{V} = \mathbf{T}(\mathcal{V}) = \mathbf{T}(\mathcal{V}_1 \oplus \mathcal{V}_2) = \mathbf{T}(\mathcal{V}_1) \oplus \mathbf{T}(\mathcal{V}_2) = \mathcal{V}_1 \oplus \mathbf{T}(\mathcal{V}_2).$$

Hence, by the argument above, $\mathbf{T}(\mathcal{V}_2) = \mathcal{V}_2$. We summarize the discussion as follows:

Proposition 2.3.21 *If* $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$, *then an automorphism of* $\mathcal{V}$ *which leaves one of the summands invariant leaves the other invariant as well.*

Example 2.3.22 (Another proof of the dimension theorem) Let $\mathbf{T}$, $\mathcal{V}$, and $\mathcal{W}$ be as in Theorem 2.3.13. Let $\mathbf{T}' : \mathcal{V}/\ker \mathbf{T} \to \mathbf{T}(\mathcal{V})$ be a linear map defined as follows. If $[\![a]\!]$ is represented by $|a\rangle$, then $\mathbf{T}'([\![a]\!]) = \mathbf{T}|a\rangle$. First, we have to show that this map is well defined, i.e., that if $[\![a']\!] = [\![a]\!]$, then $\mathbf{T}'([\![a']\!]) = \mathbf{T}|a\rangle$. But this is trivially true, because $[\![a']\!] = [\![a]\!]$ implies that $|a'\rangle = |a\rangle + |z\rangle$ with $|z\rangle \in \ker \mathbf{T}$. So,

$$\mathbf{T}'\big([\![a']\!]\big) \equiv \mathbf{T}\big|a'\big\rangle = \mathbf{T}\big(|a\rangle + |z\rangle\big) = \mathbf{T}\big(|a\rangle\big) + \underbrace{\mathbf{T}\big(|z\rangle\big)}_{=|0\rangle} = \mathbf{T}\big(|a\rangle\big).$$

One can also easily show that $\mathbf{T}'$ is linear.

We now show that $\mathbf{T}'$ is an isomorphism. Suppose that $|x\rangle \in \mathbf{T}(\mathcal{V})$. Then there is $|y\rangle \in \mathcal{V}$ such that $|x\rangle = \mathbf{T}|y\rangle = \mathbf{T}'([\![y]\!])$. This shows that $\mathbf{T}'$ is surjective. To show that it is injective, let $\mathbf{T}'([\![y]\!]) = \mathbf{T}'([\![x]\!])$; then $\mathbf{T}|y\rangle = \mathbf{T}|x\rangle$ or $\mathbf{T}(|y\rangle - |x\rangle) = 0$. This shows that $|y\rangle - |x\rangle \in \ker \mathbf{T}$, i.e., $[\![y]\!] = [\![x]\!]$. This isomorphism implies that $\dim(\mathcal{V}/\ker \mathbf{T}) = \dim \mathbf{T}(\mathcal{V})$. Equation (2.2) now yields the result of the dimension theorem.

The result of the preceding example can be generalized as follows

Theorem 2.3.23 *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ a linear map. Let $\mathcal{U}$ be a subspace of $\mathcal{V}$. Define $\mathbf{T}' : \mathcal{V}/\mathcal{U} \to \mathbf{T}(\mathcal{V})$ by $\mathbf{T}'([\![a]\!]) = \mathbf{T}|a\rangle$, where $|a\rangle$ is assumed to represent $[\![a]\!]$. Then $\mathbf{T}'$ is a well defined isomorphism.*

Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be complex vector spaces. Consider the linear map

$$\mathbf{T} : (\mathcal{U} \oplus \mathcal{V}) \otimes \mathcal{W} \to (\mathcal{U} \otimes \mathcal{W}) \oplus (\mathcal{V} \otimes \mathcal{W})$$

given by

$$\mathbf{T}\big(\big(|u\rangle + |v\rangle\big) \otimes |w\rangle\big) = |u\rangle \otimes |w\rangle + |v\rangle \otimes |w\rangle.$$

It is trivial to show that $\mathbf{T}$ is an isomorphism. We thus have

$$(\mathcal{U} \oplus \mathcal{V}) \otimes \mathcal{W} \cong (\mathcal{U} \otimes \mathcal{W}) \oplus (\mathcal{V} \otimes \mathcal{W}). \tag{2.16}$$

From the fact that $\dim(\mathcal{U} \otimes \mathcal{V}) = \dim \mathcal{U} \dim \mathcal{V}$, we have

$$\mathcal{U} \otimes \mathcal{V} \cong \mathcal{V} \otimes \mathcal{U}. \tag{2.17}$$

Moreover, since $\dim \mathbb{C} = 1$ we have $\dim(\mathbb{C} \otimes \mathcal{V}) = \dim \mathcal{V}$. Hence,

$$\mathbb{C} \otimes \mathcal{V} \cong \mathcal{V} \otimes \mathbb{C} \cong \mathcal{V}. \tag{2.18}$$

Similarly

$$\mathbb{R} \otimes \mathcal{V} \cong \mathcal{V} \otimes \mathbb{R} \cong \mathcal{V}. \tag{2.19}$$

for a real vector space $\mathcal{V}$.

2.4 Complex Structures

Thus far in our treatment of vector spaces, we have avoided changing the nature of scalars. When we declared that a vector space was complex, we kept the scalars of that vector space complex, and if we used real numbers in that vector space, they were treated as a subset of complex numbers.

In this section, we explore the possibility of changing the scalars, and the corresponding changes in the other structures of the vector space that may ensue. The interesting case is changing the reals to complex numbers.

In the discussion of changing the scalars, as well as other formal treatments of other topics, it is convenient to generalize the concept of inner products. While the notion of positive definiteness is crucial for the physical applications of an inner product, for certain other considerations, it is too restrictive. So, we relax that requirement and define our inner product anew. However, except in this subsection,

> **Box 2.4.1** *Unless otherwise indicated, all complex inner products are assumed to be sesquilinear as in Definition 2.2.1.*

complex bilinear inner product

Definition 2.4.2 Let $\mathbb{F}$ be either $\mathbb{C}$ or $\mathbb{R}$. An inner product on an $\mathbb{F}$-linear space $\mathcal{V}$ is a map $g : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$ with the following properties:

(a) symmetry: $g\big(|a\rangle, |b\rangle\big) = g\big(|b\rangle, |a\rangle\big)$;

(b) bilinearity: $g\big(|x\rangle, \alpha|a\rangle + \beta|b\rangle\big) = \alpha g\big(|x\rangle, |a\rangle\big) + \beta g\big(|x\rangle, |b\rangle\big)$,
$g\big(\alpha|a\rangle + \beta|b\rangle, |x\rangle\big) = \alpha g\big(|a\rangle, |x\rangle\big) + \beta g\big(|b\rangle, |x\rangle\big)$;

(c) nondegeneracy: $g\big(|x\rangle, |a\rangle\big) = 0 \quad \forall |x\rangle \in \mathcal{V} \quad \Rightarrow \quad |a\rangle = |0\rangle$;

with $\alpha, \beta \in \mathbb{F}$ and $|a\rangle, |b\rangle, |x\rangle \in \mathcal{V}$.

Non-degeneracy can be restated by saying that for any nonzero $|a\rangle \in \mathcal{V}$, there is at least one vector $|x\rangle \in \mathcal{V}$ such that $g(|x\rangle, |a\rangle) \neq 0$. It is the statement of the fact that the only vector orthogonal to all vectors of an inner product space is the zero vector.

Once again we use the Dirac bra and ket notation for the inner product. However, to distinguish it from the previous inner product, we subscript the notation with $\mathbb{F}$. Thus the three properties in the definition above are denoted by

$$
\begin{aligned}
&\text{(a) symmetry:} && \langle a|b\rangle_{\mathbb{F}} = \langle b|a\rangle_{\mathbb{F}};\\
&\text{(b) bilinearity:} && \langle x|\alpha a + \beta b\rangle_{\mathbb{F}} = \alpha\langle x|a\rangle_{\mathbb{F}} + \beta\langle x|b\rangle_{\mathbb{F}},\\
& && \langle \alpha a + \beta b|x\rangle_{\mathbb{F}} = \alpha\langle a|x\rangle_{\mathbb{F}} + \beta\langle b|x\rangle_{\mathbb{F}};\\
&\text{(c) non-degeneracy:} && \langle x|a\rangle_{\mathbb{F}} = 0 \quad \forall |x\rangle \in \mathcal{V} \quad \Rightarrow \quad |a\rangle = |0\rangle.
\end{aligned}
\tag{2.20}
$$

Note that $\langle\,|\,\rangle_{\mathbb{F}} = \langle\,|\,\rangle$ when $\mathbb{F} = \mathbb{R}$.

adjoint, self-adjoint, skew

Definition 2.4.3 The **adjoint** of an operator $\mathbf{A} \in \mathrm{End}(\mathcal{V})$, denoted by $\mathbf{A}^{\mathsf{T}}$, is defined by

$$\langle \mathbf{A}a|b\rangle_{\mathbb{F}} = \langle a|\mathbf{A}^{\mathsf{T}}b\rangle_{\mathbb{F}} \quad \text{or} \quad \langle a|\mathbf{A}^{\mathsf{T}}|b\rangle_{\mathbb{F}} = \langle b|\mathbf{A}|a\rangle_{\mathbb{F}}.$$

An operator $\mathbf{A}$ is called **self-adjoint** if $\mathbf{A}^{\mathsf{T}} = \mathbf{A}$, and **skew** if $\mathbf{A}^{\mathsf{T}} = -\mathbf{A}$.

From this definition and the non-degeneracy of $\langle\,|\,\rangle_{\mathbb{F}}$ it follows that

$$\big(\mathbf{A}^{\mathsf{T}}\big)^{\mathsf{T}} = \mathbf{A}. \tag{2.21}$$

Proposition 2.4.4 *An operator* $\mathbf{A} \in \operatorname{End}(\mathcal{V})$ *is skew iff* $\langle x|\mathbf{A}x\rangle_{\mathbb{F}} \equiv \langle x|\mathbf{A}|x\rangle_{\mathbb{F}} = 0$ *for all* $|x\rangle \in \mathcal{V}$.

Proof If $\mathbf{A}$ is skew, then

$$\langle x|\mathbf{A}|x\rangle_{\mathbb{F}} = \langle x|\mathbf{A}^{\mathsf{T}}|x\rangle_{\mathbb{F}} = -\langle x|\mathbf{A}|x\rangle_{\mathbb{F}} \quad \Rightarrow \quad \langle x|\mathbf{A}|x\rangle_{\mathbb{F}} = 0.$$

Conversely, suppose that $\langle x|\mathbf{A}|x\rangle_{\mathbb{F}} = 0$ for all $|x\rangle \in \mathcal{V}$, then for nonzero $\alpha, \beta \in \mathbb{F}$ and nonzero $|a\rangle, |b\rangle \in \mathcal{V}$,

$$\begin{aligned} 0 &= \langle \alpha a + \beta b|\mathbf{A}|\alpha a + \beta b\rangle_{\mathbb{F}} \\ &= \alpha^2 \underbrace{\langle a|\mathbf{A}|a\rangle_{\mathbb{F}}}_{=0} + \alpha\beta\langle a|\mathbf{A}|b\rangle_{\mathbb{F}} + \alpha\beta\langle b|\mathbf{A}|a\rangle_{\mathbb{F}} + \beta^2 \underbrace{\langle b|\mathbf{A}|b\rangle_{\mathbb{F}}}_{=0} \\ &= \alpha\beta\big(\langle b|\mathbf{A}|a\rangle_{\mathbb{F}} + \langle b|\mathbf{A}^{\mathsf{T}}|a\rangle_{\mathbb{F}}\big). \end{aligned}$$

Since $\alpha\beta \neq 0$, we must have $\langle b|(\mathbf{A} + \mathbf{A}^{\mathsf{T}})|a\rangle_{\mathbb{F}} = 0$ for all nonzero $|a\rangle, |b\rangle \in \mathcal{V}$. By non-degeneracy of the inner product, $(\mathbf{A} + \mathbf{A}^{\mathsf{T}})|a\rangle = |0\rangle$. Since this is true for all $|a\rangle \in \mathcal{V}$, we must have $\mathbf{A}^{\mathsf{T}} = -\mathbf{A}$. □

Comparing this proposition with Theorem 2.3.8 shows how strong a restriction the positive definiteness imposes on the inner product.

Definition 2.4.5 A **complex structure J** on a real vector space $\mathcal{V}$ is a linear operator which satisfies $\mathbf{J}^2 = -\mathbf{1}$ and $\langle \mathbf{J}a|\mathbf{J}b\rangle = \langle a|b\rangle$ for all $|a\rangle, |b\rangle \in \mathcal{V}$.

complex structure

Proposition 2.4.6 *The complex structure* $\mathbf{J}$ *is skew.*

Proof Let $|a\rangle \in \mathcal{V}$ and $|b\rangle = \mathbf{J}|a\rangle$. Then recalling that $\langle\,|\,\rangle_{\mathbb{R}} = \langle\,|\,\rangle$, on the one hand,

$$\langle a|\mathbf{J}a\rangle = \langle a|b\rangle = \langle \mathbf{J}a|\mathbf{J}b\rangle = \langle \mathbf{J}a|\mathbf{J}^2 a\rangle = -\langle \mathbf{J}a|a\rangle.$$

On the other hand,

$$\langle a|\mathbf{J}a\rangle = \langle a|b\rangle = \langle b|a\rangle = \langle \mathbf{J}a|a\rangle.$$

These two equations show that $\langle a|\mathbf{J}a\rangle = 0$ for all $|a\rangle \in \mathcal{V}$. Hence, by Proposition 2.4.4, $\mathbf{J}$ is skew. □

Let $|a\rangle$ be any vector in the N-dimensional real inner product space. Normalize $|a\rangle$ to get the unit vector $|e_1\rangle$. By Propositions 2.4.4 and 2.4.6, $\mathbf{J}|e_1\rangle$ is orthogonal to $|e_1\rangle$. Normalize $\mathbf{J}|e_1\rangle$ to get $|e_2\rangle$. If $N > 2$, let $|e_3\rangle$ be any unit vector orthogonal to $|e_1\rangle$ and $|e_2\rangle$. Then $|a_3\rangle \equiv \mathbf{J}|e_3\rangle$ is obviously orthogonal to $|e_3\rangle$. We claim that it is also orthogonal to both $|e_1\rangle$ and $|e_2\rangle$:

$$\begin{aligned} \langle e_1|a_3\rangle &= \langle \mathbf{J}e_1|\mathbf{J}a_3\rangle = \langle \mathbf{J}e_1|\mathbf{J}^2 e_3\rangle \\ &= -\langle \mathbf{J}e_1|e_3\rangle = -\langle e_2|e_3\rangle = 0 \\ \langle e_2|a_3\rangle &= \langle \mathbf{J}e_1|\mathbf{J}e_3\rangle = \langle e_1|e_3\rangle = 0. \end{aligned}$$

Continuing this process, we can prove the following:

Theorem 2.4.7 *The vectors* $\{|e_i\rangle, \mathbf{J}|e_i\rangle\}_{i=1}^m$ *with* $N = 2m$ *form an orthonormal basis for the real vector space* $\mathcal{V}$ *with inner product* $\langle\,|\,\rangle_{\mathbb{R}} = \langle\,|\,\rangle$. *In particular,* $\mathcal{V}$ *must be even-dimensional for it to have a complex structure* $\mathbf{J}$.

Definition 2.4.8 If $\mathcal{V}$ is a real vector space, then $\mathbb{C} \otimes \mathcal{V}$, together with the complex multiplication rule

$$\alpha\big(\beta \otimes |a\rangle\big) = (\alpha\beta) \otimes |a\rangle, \quad \alpha, \beta \in \mathbb{C},$$

complexification

is a complex vector space called the **complexification** of $\mathcal{V}$ and denoted by $\mathcal{V}^{\mathbb{C}}$. In particular, $(\mathbb{R}^n)^{\mathbb{C}} \equiv \mathbb{C} \otimes \mathbb{R}^n \cong \mathbb{C}^n$.

Note that $\dim_{\mathbb{C}} \mathcal{V}^{\mathbb{C}} = \dim_{\mathbb{R}} \mathcal{V}$ and $\dim_{\mathbb{R}} \mathcal{V}^{\mathbb{C}} = 2\dim_{\mathbb{R}} \mathcal{V}$. In fact, if $\{|a_k\rangle\}_{k=1}^N$ is a basis of $\mathcal{V}$, then it is also a basis of $\mathcal{V}^{\mathbb{C}}$ as a *complex* vector space, while $\{|a_k\rangle, i|a_k\rangle\}_{k=1}^N$ is a basis of $\mathcal{V}^{\mathbb{C}}$ as a *real* vector space.

After complexifying a real vector space $\mathcal{V}$ with inner product $\langle\,|\,\rangle_{\mathbb{R}} = \langle\,|\,\rangle$, we can define an inner product on it which is sesquilinear (or hermitian) as follows

$$\langle \alpha \otimes a|\beta \otimes b\rangle \equiv \bar{\alpha}\beta\langle a|b\rangle.$$

It is left to the reader to show that this inner product satisfies all the properties given in Definition 2.2.1.

To complexify a real vector space $\mathcal{V}$, we have to "multiply" it by the set of complex numbers: $\mathcal{V}^{\mathbb{C}} = \mathbb{C} \otimes \mathcal{V}$. As a result, we get a *real* vector space of twice the original dimension. Is there a reverse process, a "division" of a (necessarily even-dimensional) real vector space? That is, is there a way of getting a complex vector space of half complex dimension, starting with an even-dimensional real vector space?

Let $\mathcal{V}$ be a $2m$-dimensional real vector space. Let $\mathbf{J}$ be a complex structure on $\mathcal{V}$, and $\{|e_i\rangle, \mathbf{J}|e_i\rangle\}_{i=1}^m$ a basis of $\mathcal{V}$. On the subspace $\mathcal{V}_1 \equiv \mathrm{Span}\{|e_i\rangle\}_{i=1}^m$, define the multiplication by a complex number by

$$(\alpha + i\beta) \otimes |v_1\rangle \equiv (\alpha\mathbf{1} + \beta\mathbf{J})|v_1\rangle, \quad \alpha, \beta \in \mathbb{R},\ |v_1\rangle \in \mathcal{V}_1. \tag{2.22}$$

It is straightforward to show that this process turns the $2m$-dimensional real vector space $\mathcal{V}$ into the m-dimensional complex vector space $\mathcal{V}_1^{\mathbb{C}}$.

2.5 Linear Functionals

linear functional

dual vector space $\mathcal{V}^*$

An important example of a linear transformation occurs when the second vector space, $\mathcal{W}$, happens to be the set of scalars, $\mathbb{C}$ or $\mathbb{R}$, in which case the linear transformation is called a **linear functional**. The set of linear functionals $\mathcal{L}(\mathcal{V}, \mathbb{C})$—or $\mathcal{L}(\mathcal{V}, \mathbb{R})$ if $\mathcal{V}$ is a real vector space—is denoted by $\mathcal{V}^*$ and is called the **dual space** of $\mathcal{V}$.

Example 2.5.1 Here are some examples of linear functionals:

(a) Let $|a\rangle = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ be in $\mathbb{C}^n$. Define $\boldsymbol{\phi} : \mathbb{C}^n \to \mathbb{C}$ by

$$\boldsymbol{\phi}\big(|a\rangle\big) = \sum_{k=1}^{n} \alpha_k.$$

Then it is easy to show that $\boldsymbol{\phi}$ is a linear functional.

(b) Let μ_{ij} denote the elements of an $m \times n$ matrix M. Define $\boldsymbol{\omega} : \mathcal{M}^{m\times n} \to \mathbb{C}$ by

$$\boldsymbol{\omega}(\mathsf{M}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_{ij}.$$

Then it is easy to show that $\boldsymbol{\omega}$ is a linear functional.

(c) Let μ_{ij} denote the elements of an $n \times n$ matrix M. Define $\boldsymbol{\theta} : \mathcal{M}^{n\times n} \to \mathbb{C}$ by

$$\boldsymbol{\theta}(\mathsf{M}) = \sum_{j=1}^{n} \mu_{jj},$$

the sum of the diagonal elements of M. Then it is routine to show that $\boldsymbol{\theta}$ is a linear functional.

(d) Define the operator **int** : $\mathcal{C}^0(a, b) \to \mathbb{R}$ by

integration is a linear functional on the space of continuous functions

$$\mathbf{int}(f) = \int_a^b f(t)\,dt.$$

Then **int** is a linear functional on the vector space $\mathcal{C}^0(a, b)$.

(e) Let $\mathcal{V}$ be a complex inner product space. Fix $|a\rangle \in \mathcal{V}$, and let $\boldsymbol{\gamma}_a : \mathcal{V} \to \mathbb{C}$ be defined by

$$\boldsymbol{\gamma}_a\big(|b\rangle\big) = \langle a|b\rangle.$$

Then one can show that $\boldsymbol{\gamma}_a$ is a linear functional.

(f) Let $\{|a_1\rangle, |a_2\rangle, \ldots, |a_m\rangle\}$ be an arbitrary finite set of vectors in $\mathcal{V}$, and $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_m\}$ an arbitrary set of linear functionals on $\mathcal{V}$. Let

$$\mathbf{A} \equiv \sum_{k=1}^{m} |a_k\rangle\boldsymbol{\phi}_k \in \mathrm{End}(\mathcal{V})$$

be defined by

$$\mathbf{A}|x\rangle = \sum_{k=1}^{m} |a_k\rangle\boldsymbol{\phi}_k\big(|x\rangle\big) = \sum_{k=1}^{m} \boldsymbol{\phi}_k\big(|x\rangle\big)|a_k\rangle.$$

Then $\mathbf{A}$ is a linear operator on $\mathcal{V}$.

An example of linear isomorphism is that between a vector space and its dual space, which we discuss now. Consider an N-dimensional vector space with a basis $B = \{|a_1\rangle, |a_2\rangle, \ldots, |a_N\rangle\}$. For any given set of N scalars,

$\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$, define the linear functional $\boldsymbol{\phi}_\alpha$ by $\boldsymbol{\phi}_\alpha|a_i\rangle = \alpha_i$. When $\boldsymbol{\phi}_\alpha$ acts on any arbitrary vector $|b\rangle = \sum_{i=1}^N \beta_i|a_i\rangle$ in $\mathcal{V}$, the result is

$$\boldsymbol{\phi}_\alpha|b\rangle = \boldsymbol{\phi}_\alpha\left(\sum_{i=1}^N \beta_i|a_i\rangle\right) = \sum_{i=1}^N \beta_i\boldsymbol{\phi}_\alpha|a_i\rangle = \sum_{i=1}^N \beta_i\alpha_i. \tag{2.23}$$

This expression suggests that $|b\rangle$ can be represented as a column vector with entries $\beta_1, \beta_2, \ldots, \beta_N$ and $\boldsymbol{\phi}_\alpha$ as a row vector with entries $\alpha_1, \alpha_2, \ldots, \alpha_N$. Then $\boldsymbol{\phi}_\alpha|b\rangle$ is merely the matrix product[14] of the row vector (on the left) and the column vector (on the right).

$\boldsymbol{\phi}_\alpha$ is uniquely determined by the set $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$. In other words, corresponding to every set of N scalars there exists a unique linear functional. This leads us to a particular set of functionals, $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N$ corresponding, respectively, to the sets of scalars $\{1, 0, 0, \ldots, 0\}$, $\{0, 1, 0, \ldots, 0\}$, $\ldots$, $\{0, 0, 0, \ldots, 1\}$. This means that

Every set of N scalars defines a linear functional.

$$\begin{aligned}
\boldsymbol{\phi}_1|a_1\rangle = 1 \quad &\text{and} \quad \boldsymbol{\phi}_1|a_j\rangle = 0 \quad \text{for } j \neq 1,\\
\boldsymbol{\phi}_2|a_2\rangle = 1 \quad &\text{and} \quad \boldsymbol{\phi}_2|a_j\rangle = 0 \quad \text{for } j \neq 2,\\
\vdots \qquad & \qquad\quad \vdots \qquad\qquad \vdots\\
\boldsymbol{\phi}_N|a_N\rangle = 1 \quad &\text{and} \quad \boldsymbol{\phi}_N|a_j\rangle = 0 \quad \text{for } j \neq N,
\end{aligned}$$

or that

$$\boldsymbol{\phi}_i|a_j\rangle = \delta_{ij}, \tag{2.24}$$

where δ_{ij} is the Kronecker delta.

The functionals of Eq. (2.24) form a basis of the dual space $\mathcal{V}^*$. To show this, consider an arbitrary $\boldsymbol{\gamma} \in \mathcal{V}^*$, which is uniquely determined by its action on the vectors in a basis $B = \{|a_1\rangle, |a_2\rangle, \ldots, |a_N\rangle\}$. Let $\boldsymbol{\gamma}|a_i\rangle = \gamma_i \in \mathbb{C}$. Then we claim that $\boldsymbol{\gamma} = \sum_{i=1}^N \gamma_i\boldsymbol{\phi}_i$. In fact, consider an arbitrary vector $|a\rangle$ in $\mathcal{V}$ with components $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ with respect to B. Then, on the one hand,

$$\boldsymbol{\gamma}|a\rangle = \boldsymbol{\gamma}\left(\sum_{i=1}^N \alpha_i|a_i\rangle\right) = \sum_{i=1}^N \alpha_i\boldsymbol{\gamma}|a_i\rangle = \sum_{i=1}^N \alpha_i\gamma_i.$$

On the other hand,

$$\begin{aligned}
\left(\sum_{i=1}^N \gamma_i\boldsymbol{\phi}_i\right)|a\rangle &= \left(\sum_{i=1}^N \gamma_i\boldsymbol{\phi}_i\right)\left(\sum_{j=1}^N \alpha_j|a_j\rangle\right)\\
&= \sum_{i=1}^N \gamma_i \sum_{j=1}^N \alpha_j\boldsymbol{\phi}_i|a_j\rangle = \sum_{i=1}^N \gamma_i \sum_{j=1}^N \alpha_j\delta_{ij} = \sum_{i=1}^N \gamma_i\alpha_i.
\end{aligned}$$

[14]Matrices will be taken up in Chap. 5. Here, we assume only a nodding familiarity with elementary matrix operations.

Since the actions of $\boldsymbol{\gamma}$ and $\sum_{i=1}^{N} \gamma_i \boldsymbol{\phi}_i$ yield equal results for arbitrary $|a\rangle$, we conclude that $\boldsymbol{\gamma} = \sum_{i=1}^{N} \gamma_i \boldsymbol{\phi}_i$, i.e., $\{\boldsymbol{\phi}_i\}_{i=1}^{N}$ span $\mathcal{V}^*$. Thus, we have the following result.

Theorem 2.5.2 *For every basis $B = \{|a_j\rangle\}_{j=1}^{N}$ in $\mathcal{V}$, there corresponds a unique basis $B^* = \{\boldsymbol{\phi}_i\}_{i=1}^{N}$ in $\mathcal{V}^*$ with the property that $\boldsymbol{\phi}_i|a_j\rangle = \delta_{ij}$.*

By this theorem the dual space of an N-dimensional vector space is also N-dimensional, and thus isomorphic to it. The basis B^* is called the **dual basis** of B. A corollary to Theorem 2.5.2 is that to every vector in $\mathcal{V}$ there corresponds a *unique* linear functional in $\mathcal{V}^*$. This can be seen by noting that every vector $|a\rangle$ is uniquely determined by its components $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ in a basis B. The unique linear functional $\boldsymbol{\phi}_a$ corresponding to $|a\rangle$, also called the dual of $|a\rangle$, is simply $\sum_{i=1}^{N} \alpha_i \boldsymbol{\phi}_i$, with $\boldsymbol{\phi}_i \in B^*$.

dual basis

Definition 2.5.3 An **annihilator** of $|a\rangle \in \mathcal{V}$ is a linear functional $\boldsymbol{\phi} \in \mathcal{V}^*$ such that $\boldsymbol{\phi}|a\rangle = 0$. Let $\mathcal{W}$ be a subspace of $\mathcal{V}$. The set of linear functionals in $\mathcal{V}^*$ that annihilate all vectors in $\mathcal{W}$ is denoted by $\mathcal{W}^0$.

annihilator of a vector and a subspace

The reader may check that $\mathcal{W}^0$ is a subspace of $\mathcal{V}^*$. Moreover, if we extend a basis $\{|a_i\rangle\}_{i=1}^{k}$ of $\mathcal{W}$ to a basis $B = \{|a_i\rangle\}_{i=1}^{N}$ of $\mathcal{V}$, then we can show that the functionals $\{\boldsymbol{\phi}_j\}_{j=k+1}^{N}$, chosen from the basis $B^* = \{\boldsymbol{\phi}_j\}_{j=1}^{N}$ dual to B, span $\mathcal{W}^0$. It then follows that

$$\dim \mathcal{V} = \dim \mathcal{W} + \dim \mathcal{W}^0. \tag{2.25}$$

We shall have occasions to use annihilators later on when we discuss symplectic geometry.

We have "dualed" a vector, a basis, and a complete vector space. The only object remaining is a linear transformation.

dual, or pull back, of a linear transformation

Definition 2.5.4 Let $\mathbf{T} : \mathcal{V} \to \mathcal{U}$ be a linear map. Define $\mathbf{T}^* : \mathcal{U}^* \to \mathcal{V}^*$ by[15]

$$\big[\mathbf{T}^*(\boldsymbol{\gamma})\big]|a\rangle = \boldsymbol{\gamma}\big(\mathbf{T}|a\rangle\big) \quad \forall |a\rangle \in \mathcal{V},\ \boldsymbol{\gamma} \in \mathcal{U}^*,$$

$\mathbf{T}^*$ is called the **dual** or **pullback**, of $\mathbf{T}$.

One can readily verify that $\mathbf{T}^* \in \mathcal{L}(\mathcal{U}^*, \mathcal{V}^*)$, i.e., that $\mathbf{T}^*$ is a *linear* operator on $\mathcal{U}^*$. Some of the mapping properties of $\mathbf{T}^*$ are tied to those of $\mathbf{T}$. To see this we first consider the kernel of $\mathbf{T}^*$. Clearly, $\boldsymbol{\gamma}$ is in the kernel of $\mathbf{T}^*$ if and only if $\boldsymbol{\gamma}$ annihilates all vectors of the form $\mathbf{T}|a\rangle$, i.e., all vectors in $\mathbf{T}(\mathcal{V})$. It follows that $\boldsymbol{\gamma}$ is in $\mathbf{T}(\mathcal{V})^0$. In particular, if $\mathbf{T}$ is surjective, $\mathbf{T}(\mathcal{V}) = \mathcal{U}$, and $\boldsymbol{\gamma}$ annihilates all vectors in $\mathcal{U}$, i.e., it is the zero linear functional. We conclude that $\ker \mathbf{T}^* = 0$, and therefore, $\mathbf{T}^*$ is injective. Similarly, one can show that if $\mathbf{T}$ is injective, then $\mathbf{T}^*$ is surjective. We summarize the discussion above:

[15] Do not confuse this "*" with complex conjugation.

Proposition 2.5.5 *Let* $\mathbf{T}$ *be a linear transformation and* $\mathbf{T}^*$ *its pull back. Then* $\ker \mathbf{T}^* = \mathbf{T}(\mathcal{V})^0$. *If* $\mathbf{T}$ *is surjective (injective), then* $\mathbf{T}^*$ *is injective (surjective). In particular,* $\mathbf{T}^*$ *is an isomorphism if* $\mathbf{T}$ *is.*

duals and inner products

It is useful to make a connection between the inner product and linear functionals. To do this, consider a basis $\{|a_1\rangle, |a_2\rangle, \dots, |a_N\rangle\}$ and let $\alpha_i = \langle a|a_i\rangle$. As noted earlier, the set of scalars $\{\alpha_i\}_{i=1}^N$ defines a unique linear functional $\boldsymbol{\gamma}_a$ (see Example 2.5.1) such that $\boldsymbol{\gamma}_a|a_i\rangle = \alpha_i$. Since $\langle a|a_i\rangle$ is also equal to α_i, it is natural to identify $\boldsymbol{\gamma}_a$ with the symbol $\langle a|$, and write $\boldsymbol{\gamma}_a \mapsto \langle a|$.

It is also convenient to introduce the notation[16]

$$\big(|a\rangle\big)^\dagger \equiv \langle a|, \tag{2.26}$$

dagger of a linear combination of vectors

where the symbol † means "dual, or dagger of". Now we ask: How does this dagger operation act on a linear combination of vectors? Let $|c\rangle = \alpha|a\rangle + \beta|b\rangle$ and take the inner product of $|c\rangle$ with an arbitrary vector $|x\rangle$ using linearity in the second factor: $\langle x|c\rangle = \alpha\langle x|a\rangle + \beta\langle x|b\rangle$. Now complex conjugate both sides and use the (sesqui)symmetry of the inner product:

$$\begin{aligned}
(\text{LHS})^* &= \langle x|c\rangle^* = \langle c|x\rangle, \\
(\text{RHS})^* &= \alpha^*\langle x|a\rangle^* + \beta^*\langle x|b\rangle^* = \alpha^*\langle a|x\rangle + \beta^*\langle b|x\rangle \\
&= \big(\alpha^*\langle a| + \beta^*\langle b|\big)|x\rangle.
\end{aligned}$$

Since this is true for all $|x\rangle$, we must have $(|c\rangle)^\dagger \equiv \langle c| = \alpha^*\langle a| + \beta^*\langle b|$. Therefore, in a duality "operation" the complex scalars must be conjugated. So, we have

$$\big(\alpha|a\rangle + \beta|b\rangle\big)^\dagger = \alpha^*\langle a| + \beta^*\langle b|. \tag{2.27}$$

Thus, unlike the association $|a\rangle \mapsto \boldsymbol{\gamma}_a$ which is linear, the association $\boldsymbol{\gamma}_a \mapsto \langle a|$ is not linear, but sesquilinear:

$$\boldsymbol{\gamma}_{\alpha a+\beta b} \mapsto \alpha^*\langle a| + \beta^*\langle b|.$$

It is convenient to represent $|a\rangle \in \mathbb{C}^n$ as a column vector

$$|a\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Then the definition of the complex inner product suggests that the dual of $|a\rangle$ must be represented as a row vector with complex conjugate entries:

$$\langle a| = \begin{pmatrix} \alpha_1^* & \alpha_2^* & \dots & \alpha_n^* \end{pmatrix}, \tag{2.28}$$

[16] The significance of this notation will become clear in Sect. 4.3.

and the inner product can be written as the (matrix) product

$$\langle a|b\rangle = \begin{pmatrix}\alpha_1^* & \alpha_2^* & \cdots & \alpha_n^*\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{pmatrix} = \sum_{i=1}^{n}\alpha_i^*\beta_i.$$

Compare (2.28) with the comments after (2.23). The complex conjugation in (2.28) is the result of the sesquilinearity of the association $|a\rangle \leftrightarrow \langle a|$.

2.6 Multilinear Maps

There is a very useful generalization of the linear functionals that becomes essential in the treatment of tensors later in the book. However, a limited version of its application is used in the discussion of determinants, which we shall start here.

Definition 2.6.1 Let $\mathcal{V}$ and $\mathcal{U}$ be vector spaces. Let $\mathcal{V}^p$ denote the p-fold Cartesian product of $\mathcal{V}$. A **p-linear map** from $\mathcal{V}$ to $\mathcal{U}$ is a map $\boldsymbol{\theta} : \mathcal{V}^p \to \mathcal{U}$ which is linear with respect to each of its arguments:

p-linear map

$$\begin{aligned}&\boldsymbol{\theta}\big(|a_1\rangle, \dots, \alpha|a_j\rangle + \beta|b_j\rangle, \dots, |a_p\rangle\big)\\ &\quad = \alpha\boldsymbol{\theta}\big(|a_1\rangle, \dots, |a_j\rangle, \dots, |a_p\rangle\big) + \beta\boldsymbol{\theta}\big(|a_1\rangle, \dots, |b_j\rangle, \dots, |a_p\rangle\big).\end{aligned}$$

A p-linear map from $\mathcal{V}$ to $\mathbb{C}$ or $\mathbb{R}$ is called a **p-linear function** in $\mathcal{V}$.

p-linear function

As an example, let $\{\boldsymbol{\phi}_i\}_{i=1}^p$ be linear functionals on $\mathcal{V}$. Define $\boldsymbol{\theta}$ by

$$\boldsymbol{\theta}\big(|a_1\rangle, \dots, |a_p\rangle\big) = \boldsymbol{\phi}_1\big(|a_1\rangle\big)\dots\boldsymbol{\phi}_p\big(|a_p\rangle\big), \quad |a_i\rangle \in \mathcal{V}.$$

Clearly $\boldsymbol{\theta}$ is p-linear.

Let σ denote a permutation of $1, 2, \dots, p$. Define the p-linear map $\sigma\boldsymbol{\omega}$ by

$$\sigma\boldsymbol{\omega}\big(|a_1\rangle, \dots, |a_p\rangle\big) = \boldsymbol{\omega}\big(|a_{\sigma(1)}\rangle, \dots, |a_{\sigma(p)}\rangle\big)$$

Definition 2.6.2 A p-linear map $\boldsymbol{\omega}$ from $\mathcal{V}$ to $\mathcal{U}$ is **skew-symmetric** if $\sigma\boldsymbol{\omega} = \epsilon_\sigma \cdot \boldsymbol{\omega}$, i.e., if

skew-symmetric p-linear map

$$\boldsymbol{\omega}\big(|a_{\sigma(1)}\rangle, \dots, |a_{\sigma(p)}\rangle\big) = \epsilon_\sigma\boldsymbol{\omega}\big(|a_1\rangle, \dots, |a_p\rangle\big)$$

where ϵ_σ is the sign of σ, which is $+1$ if σ is even and -1 if it is odd. The set of p-linear skew-symmetric maps from $\mathcal{V}$ to $\mathcal{U}$ is denoted by $\Lambda^p(\mathcal{V}, \mathcal{U})$. The set of p-linear skew-symmetric functions in $\mathcal{V}$ is denoted by $\Lambda^p(\mathcal{V})$.

The permutation sign ϵ_σ is sometimes written as

$$\epsilon_\sigma = \epsilon_{\sigma(1)\sigma(2)\dots\sigma(p)} \equiv \epsilon_{i_1 i_2 \dots i_p}, \tag{2.29}$$

where $i_k \equiv \sigma(k)$.

Any p-linear map can be turned into a skew-symmetric p-linear map. In fact, if $\boldsymbol{\theta}$ is a p-linear map, then

$$\boldsymbol{\omega} \equiv \sum_{\pi} \epsilon_\pi \cdot \pi\boldsymbol{\theta} \tag{2.30}$$

is skew-symmetric:

$$\begin{aligned}\sigma\boldsymbol{\omega} &= \sigma\sum_{\pi} \epsilon_\pi \cdot \pi\boldsymbol{\theta} = \sum_{\pi} \epsilon_\pi \cdot (\sigma\pi)\boldsymbol{\theta} = (\epsilon_\sigma)^2 \sum_{\pi} \epsilon_\pi \cdot (\sigma\pi)\boldsymbol{\theta} \\ &= \epsilon_\sigma \sum_{\pi} (\epsilon_\sigma\epsilon_\pi) \cdot (\sigma\pi)\boldsymbol{\theta} = \epsilon_\sigma \sum_{\sigma\pi} \epsilon_{\sigma\pi} \cdot (\sigma\pi)\boldsymbol{\theta} = \epsilon_\sigma \cdot \boldsymbol{\omega},\end{aligned}$$

where we have used the fact that the sign of the product is the product of the signs of two permutations, and if $\sum_\pi$ sums over all permutations, then so does $\sum_{\sigma\pi}$.

The following theorem can be proved using properties of permutations:

Theorem 2.6.3 *Let $\boldsymbol{\omega} \in \Lambda^p(\mathcal{V}, \mathcal{U})$. Then the following statements are equivalent*:

1. $\boldsymbol{\omega}(|a_1\rangle, \ldots, |a_p\rangle) = 0$ *whenever* $|a_i\rangle = |a_j\rangle$ *for some pair* $i \neq j$.
2. $\boldsymbol{\omega}(|a_{\sigma(1)}\rangle, \ldots, |a_{\sigma(p)}\rangle) = \epsilon_\sigma \boldsymbol{\omega}(|a_1\rangle, \ldots, |a_p\rangle)$, *for any permutation* σ *of* $1, 2, \ldots, p$, *and any* $|a_1\rangle, \ldots, |a_p\rangle$ *in* $\mathcal{V}$.
3. $\boldsymbol{\omega}(|a_1\rangle, \ldots, |a_p\rangle) = 0$ *whenever* $\{|a_k\rangle\}_{k=1}^p$ *are linearly dependent*.

Proposition 2.6.4 *Let $N = \dim \mathcal{V}$ and $\boldsymbol{\omega} \in \Lambda^N(\mathcal{V}, \mathcal{U})$. Then $\boldsymbol{\omega}$ is determined uniquely by its value on a basis of $\mathcal{V}$. In particular, if $\boldsymbol{\omega}$ vanishes on a basis*, *then* $\boldsymbol{\omega} = \mathbf{0}$.

Proof Let $\{|e_k\rangle\}_{k=1}^N$ be a basis of $\mathcal{V}$. Let $\{|a_j\rangle\}_{j=1}^N$ be any set of vectors in $\mathcal{V}$ and write $|a_j\rangle = \sum_{k=1}^N \alpha_{jk}|e_k\rangle$ for $j = 1, \ldots, N$. Then

$$\begin{aligned}\boldsymbol{\omega}\bigl(|a_1\rangle, \ldots, |a_N\rangle\bigr) &= \sum_{k_1 \ldots k_N=1}^{N} \alpha_{1k_1} \ldots \alpha_{Nk_N} \boldsymbol{\omega}\bigl(|e_{k_1}\rangle, \ldots, |e_{k_N}\rangle\bigr) \\ &\equiv \sum_{\pi} \alpha_{1\pi(1)} \ldots \alpha_{N\pi(N)} \boldsymbol{\omega}\bigl(|e_{\pi(1)}\rangle, \ldots, |e_{\pi(N)}\rangle\bigr) \\ &= \biggl(\sum_{\pi} \epsilon_\pi \alpha_{1\pi(1)} \ldots \alpha_{N\pi(N)}\biggr) \boldsymbol{\omega}\bigl(|e_1\rangle, \ldots, |e_N\rangle\bigr).\end{aligned}$$

Since the term in parentheses is a constant, we are done. □

Determinant function

Definition 2.6.5 A skew symmetric N-linear function in $\mathcal{V}$, i.e., a member of $\Lambda^N(\mathcal{V})$ is called a **determinant function** in $\mathcal{V}$.

Let $B = \{|e_k\rangle\}_{k=1}^N$ be a basis of $\mathcal{V}$ and $B^* = \{\boldsymbol{\epsilon}_j\}_{j=1}^N$ a basis of $\mathcal{V}^*$, dual to B. For any set of N vectors $\{|a_k\rangle\}_{k=1}^N$ in $\mathcal{V}$, define the N-linear function $\boldsymbol{\theta}$ by

$$\boldsymbol{\theta}\bigl(|a_1\rangle, \ldots, |a_N\rangle\bigr) = \boldsymbol{\epsilon}_1\bigl(|a_1\rangle\bigr) \ldots \boldsymbol{\epsilon}_N\bigl(|a_N\rangle\bigr),$$

and note that

$$\pi\boldsymbol{\theta}\big(|e_1\rangle,\dots,|e_N\rangle\big)\equiv\boldsymbol{\theta}\big(|e_{\pi(1)}\rangle,\dots,|e_{\pi(N)}\rangle\big)=\delta_{\iota\pi},$$

where ι is the identity permutation and $\delta_{\iota\pi}=1$ if $\pi=\iota$ and $\delta_{\iota\pi}=0$ if $\pi\neq\iota$. Now let $\boldsymbol{\Delta}$ be defined by $\boldsymbol{\Delta}\equiv\sum_\pi\epsilon_\pi\cdot\pi\boldsymbol{\theta}$. Then, by Eq. (2.30), $\boldsymbol{\Delta}\in\Lambda^N(\mathcal{V})$, i.e., $\boldsymbol{\Delta}$ is a determinant function. Furthermore,

$$\boldsymbol{\Delta}\big(|e_1\rangle,\dots,|e_N\rangle\big)=\sum_\pi\epsilon_\pi\cdot\pi\boldsymbol{\theta}\big(|e_1\rangle,\dots,|e_N\rangle\big)=\sum_\pi\epsilon_\pi\delta_{\iota\pi}=\epsilon_\iota=1$$

Therefore, we have the following:

> **Box 2.6.6** *In every finite-dimensional vector space, there are determinant functions which are not identically zero.*

Proposition 2.6.7 *Let $\boldsymbol{\omega}\in\Lambda^N(\mathcal{V},\mathcal{U})$. Let $\boldsymbol{\Delta}$ be a fixed nonzero determinant function in $\mathcal{V}$. Then $\boldsymbol{\omega}$ determines a unique $|u_\Delta\rangle\in\mathcal{U}$ such that*

$$\boldsymbol{\omega}\big(|v_1\rangle,\dots,|v_N\rangle\big)=\boldsymbol{\Delta}\big(|v_1\rangle,\dots,|v_N\rangle\big)\cdot|u_\Delta\rangle.$$

Proof Let $\{|v_k\rangle\}_{k=1}^N$ be a basis of $\mathcal{V}$ such that $\boldsymbol{\Delta}(|v_1\rangle,\dots,|v_N\rangle)\neq 0$. By dividing one of the vectors (or $\boldsymbol{\Delta}$) by a constant, we can assume that $\boldsymbol{\Delta}(|v_1\rangle,\dots,|v_N\rangle)=1$. Denote $\boldsymbol{\omega}(|v_1\rangle,\dots,|v_N\rangle)$ by $|u_\Delta\rangle$. Now note that $\boldsymbol{\omega}-\boldsymbol{\Delta}\cdot|u_\Delta\rangle$ yields zero on the basis $\{|v_k\rangle\}_{k=1}^N$. By Proposition 2.6.4, it must be identically zero. □

Corollary 2.6.8 *Let $\boldsymbol{\Delta}$ be a fixed nonzero determinant function in $\mathcal{V}$. Then every determinant function is a scalar multiple of $\boldsymbol{\Delta}$.*

Proof Let $\mathcal{U}$ be $\mathbb{C}$ or $\mathbb{R}$ in Proposition 2.6.7. □

Proposition 2.6.9 *Let $\boldsymbol{\Delta}$ be a determinant function in the N-dimensional vector space $\mathcal{V}$. Let $|v\rangle$ and $\{|v_k\rangle\}_{k=1}^N$ be vectors in $\mathcal{V}$. Then*

$$\sum_{j=1}^N(-1)^{j-1}\boldsymbol{\Delta}\big(|v\rangle,|v_1\rangle,\dots,\widehat{|v_j\rangle},\dots,|v_N\rangle\big)\cdot|v_j\rangle=\boldsymbol{\Delta}\big(|v_1\rangle,\dots,|v_N\rangle\big)\cdot|v\rangle$$

where a hat on a vector means that particular vector is missing.

Proof See Problem 2.37. □

2.6.1 Determinant of a Linear Operator

Let $\mathbf{A}$ be a linear operator on an N-dimensional vector space $\mathcal{V}$. Choose a nonzero determinant function $\boldsymbol{\Delta}$. For a basis $\{|v_i\rangle\}_{i=1}^N$ define the function $\boldsymbol{\Delta}_A$ by

$$\boldsymbol{\Delta}_A\big(|v_1\rangle,\dots,|v_N\rangle\big)\equiv\boldsymbol{\Delta}\big(\mathbf{A}|v_1\rangle,\dots,\mathbf{A}|v_N\rangle\big).\tag{2.31}$$

Clearly, $\mathbf{\Delta}_A$ is also a determinant function. By Corollary 2.6.8, it is a multiple of $\mathbf{\Delta}$. So, $\mathbf{\Delta}_A = \alpha\mathbf{\Delta}$. Furthermore, it is independent of the nonzero determinant function chosen, because if $\mathbf{\Delta}'$ is another nonzero determinant function, then again by Corollary 2.6.8, $\mathbf{\Delta}' = \lambda\mathbf{\Delta}$, and

$$\mathbf{\Delta}'_A = \lambda\mathbf{\Delta}_A = \lambda\alpha\mathbf{\Delta} = \alpha\mathbf{\Delta}'.$$

determinant of an operator defined

This means that α is determined only by $\mathbf{A}$, independent of the nonzero determinant function and the basis chosen.

Definition 2.6.10 Let $\mathbf{A} \in \operatorname{End}(\mathcal{V})$. Let $\mathbf{\Delta}$ be a nonzero determinant function in $\mathcal{V}$, and let $\mathbf{\Delta}_A$ be as in Eq. (2.31). Then

$$\mathbf{\Delta}_A = \det\mathbf{A} \cdot \mathbf{\Delta} \tag{2.32}$$

defines the **determinant** of $\mathbf{A}$.

Using Eq. (2.32), we have the following theorem whose proof is left as Problem 2.38:

Theorem 2.6.11 *The determinant of a linear operator* $\mathbf{A}$ *has the following properties*:

1. *If* $\mathbf{A} = \lambda\mathbf{1}$, *then* $\det\mathbf{A} = \lambda^N$.
2. $\mathbf{A}$ *is invertible iff* $\det\mathbf{A} \neq 0$.
3. $\det(\mathbf{A} \circ \mathbf{B}) = \det\mathbf{A}\det\mathbf{B}$.

2.6.2 Classical Adjoint

Let $\mathcal{V}$ be an N-dimensional vector space, $\mathbf{\Delta}$ a determinant function in $\mathcal{V}$, and $\mathbf{A} \in \operatorname{End}(\mathcal{V})$. For $|v\rangle, |v_i\rangle \in \mathcal{V}$, define $\varPhi : \mathcal{V}^N \to \operatorname{End}(\mathcal{V})$ by

$$\begin{aligned}\varPhi\bigl(|v_1\rangle, \ldots, |v_N\rangle\bigr)|v\rangle \\ = \sum_{j=1}^{N}(-1)^{j-1}\mathbf{\Delta}\bigl(|v\rangle, \mathbf{A}|v_1\rangle, \ldots, \widehat{\mathbf{A}|v_j\rangle}, \ldots, \mathbf{A}|v_N\rangle\bigr) \cdot |v_j\rangle.\end{aligned}$$

Clearly $\varPhi$ is skew-symmetric. Therefore, by Proposition 2.6.7, there is a unique linear operator—call it $\operatorname{ad}(\mathbf{A})$—such that

$$\varPhi\bigl(|v_1\rangle, \ldots, |v_N\rangle\bigr) = \mathbf{\Delta}\bigl(|v_1\rangle, \ldots, |v_N\rangle\bigr) \cdot \operatorname{ad}(\mathbf{A}),$$

i.e.,

$$\begin{aligned}\sum_{j=1}^{N}(-1)^{j-1}\mathbf{\Delta}\bigl(|v\rangle, \mathbf{A}|v_1\rangle, \ldots, \widehat{\mathbf{A}|v_j\rangle}, \ldots, \mathbf{A}|v_N\rangle\bigr) \cdot |v_j\rangle \\ = \mathbf{\Delta}\bigl(|v_1\rangle, \ldots, |v_N\rangle\bigr) \cdot \operatorname{ad}(\mathbf{A})|v\rangle.\end{aligned} \tag{2.33}$$

This equation shows that ad($\mathbf{A}$) is independent of the determinant function chosen, and is called the **classical adjoint** of $\mathbf{A}$.

classical adjoint of an operator

Proposition 2.6.12 *The classical adjoint satisfies the following relations*:

$$\mathrm{ad}(\mathbf{A}) \circ \mathbf{A} = \det \mathbf{A} \cdot \mathbf{1} = \mathbf{A} \circ \mathrm{ad}(\mathbf{A}) \tag{2.34}$$

where $\mathbf{1}$ *is the unit operator.*

Proof Replace $|v\rangle$ with $\mathbf{A}|v\rangle$ in Eq. (2.33) to obtain

$$\sum_{j=1}^{N} (-1)^{j-1} \boldsymbol{\Delta}\big(\mathbf{A}|v\rangle, \mathbf{A}|v_1\rangle, \ldots, \widehat{\mathbf{A}|v_j\rangle}, \ldots, \mathbf{A}|v_N\rangle\big) \cdot |v_j\rangle$$
$$= \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big)\, \mathrm{ad}(\mathbf{A}) \circ \mathbf{A}|v\rangle.$$

Then, the left-hand side can be written as

$$\mathrm{LHS} = \det \mathbf{A} \cdot \sum_{j=1}^{N} (-1)^{j-1} \boldsymbol{\Delta}\big(|v\rangle, |v_1\rangle, \ldots, \widehat{|v_j\rangle}, \ldots, |v_N\rangle\big) \cdot |v_j\rangle$$
$$= \det \mathbf{A} \cdot \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big) \cdot |v\rangle,$$

where the last equality follows from Proposition 2.6.9. Noting that $|v\rangle$ is arbitrary, the first equality of the proposition follows.

To obtain the second equality, apply $\mathbf{A}$ to (2.33). Then by Proposition 2.6.9, the left-hand side becomes

$$\mathrm{LHS} = \sum_{j=1}^{N} (-1)^{j-1} \boldsymbol{\Delta}\big(|v\rangle, \mathbf{A}|v_1\rangle, \ldots, \widehat{\mathbf{A}|v_j\rangle}, \ldots, \mathbf{A}|v_N\rangle\big) \cdot \mathbf{A}|v_j\rangle$$
$$= \boldsymbol{\Delta}\big(\mathbf{A}|v_1\rangle, \ldots, \mathbf{A}|v_N\rangle\big) \cdot |v\rangle = \det \mathbf{A} \cdot \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big) \cdot |v\rangle,$$

and the right-hand side becomes

$$\mathrm{RHS} = \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big) \cdot \mathbf{A} \circ \mathrm{ad}(\mathbf{A})|v\rangle.$$

Since the two sides hold for arbitrary $|v\rangle$, the second equality of the proposition follows. □

Corollary 2.6.13 *If* $\det \mathbf{A} \neq 0$, *then* $\mathbf{A}$ *is invertible and*

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \cdot \mathrm{ad}(\mathbf{A}).$$

2.7 Problems

2.1 Let $\mathbb{R}^+$ denote the set of positive real numbers. Define the "sum" of two elements of $\mathbb{R}^+$ to be their usual product, and define scalar multiplication by elements of $\mathbb{R}$ as being given by $r \cdot p = p^r$ where $r \in \mathbb{R}$ and $p \in \mathbb{R}^+$. With these operations, show that $\mathbb{R}^+$ is a vector space over $\mathbb{R}$.

2.2 Show that the intersection of two subspaces is also a subspace.

2.3 For each of the following sub*sets* of $\mathbb{R}^3$ determine whether it is a sub*space* of $\mathbb{R}^3$:

(a) $\{(x, y, z) \in \mathbb{R}^3 | x + y - 2z = 0\}$;
(b) $\{(x, y, z) \in \mathbb{R}^3 | x + y - 2z = 3\}$;
(c) $\{(x, y, z) \in \mathbb{R}^3 | xyz = 0\}$.

2.4 Prove that the components of a vector in a given basis are unique.

2.5 Show that the following vectors form a basis for $\mathbb{C}^n$ (or $\mathbb{R}^n$).

$$|a_1\rangle = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}, \quad |a_2\rangle = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad \ldots, \quad |a_n\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

2.6 Prove Theorem 2.1.6.

2.7 Let $\mathcal{W}$ be a subspace of $\mathbb{R}^5$ defined by

$$\mathcal{W} = \left\{(x_1, \ldots, x_5) \in \mathbb{R}^5 \mid x_1 = 3x_2 + x_3, x_2 = x_5, \text{ and } x_4 = 2x_3\right\}.$$

Find a basis for $\mathcal{W}$.

2.8 Let $\mathcal{U}_1$ and $\mathcal{U}_2$ be subspaces of $\mathcal{V}$. Show that

(a) $\dim(\mathcal{U}_1 + \mathcal{U}_2) = \dim \mathcal{U}_1 + \dim \mathcal{U}_2 - \dim(\mathcal{U}_1 \cap \mathcal{U}_2)$. Hint: Let $\{|a_i\rangle\}_{i=1}^m$ be a basis of $\mathcal{U}_1 \cap \mathcal{U}_2$. Extend this to $\{\{|a_i\rangle\}_{i=1}^m, \{|b_i\rangle\}_{i=1}^k\}$, a basis for $\mathcal{U}_1$, and to $\{\{|a_i\rangle\}_{i=1}^m, \{|c_i\rangle\}_{i=1}^l\}$, a basis for $\mathcal{U}_2$. Now show that $\{\{|a_i\rangle\}_{i=1}^m, \{|b_i\rangle\}_{i=1}^k, \{|c_i\rangle\}_{i=1}^l\}$ is a basis for $\mathcal{U}_1 + \mathcal{U}_2$.
(b) If $\mathcal{U}_1 + \mathcal{U}_2 = \mathcal{V}$ and $\dim \mathcal{U}_1 + \dim \mathcal{U}_2 = \dim \mathcal{V}$, then $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$.
(c) If $\dim \mathcal{U}_1 + \dim \mathcal{U}_2 > \dim \mathcal{V}$, then $\mathcal{U}_1 \cap \mathcal{U}_2 \neq \{0\}$.

2.9 Show that the vectors defined in Eq. (2.5) span $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$.

2.10 Show that the inner product of any vector with $|0\rangle$ is zero.

2.11 Find a_0, b_0, b_1, c_0, c_1, and c_2 such that the polynomials a_0, $b_0 + b_1 t$, and $c_0 + c_1 t + c_2 t^2$ are mutually orthonormal in the interval [0, 1]. The inner product is as defined for polynomials in Example 2.2.3 with $w(t) = 1$.

2.12 Given the linearly independent vectors $x(t) = t^n$, for $n = 0, 1, 2, \ldots$ in $\mathcal{P}^c[t]$, use the Gram–Schmidt process to find the orthonormal polynomials $e_0(t)$, $e_1(t)$, and $e_2(t)$

(a) when the inner product is defined as $\langle x | y\rangle = \int_{-1}^{1} x^*(t) y(t)\, dt$.

(b) when the inner product is defined with a nontrivial weight function:

$$\langle x|y\rangle = \int_{-\infty}^{\infty} e^{-t^2} x^*(t) y(t)\,dt.$$

Hint: Use the following result:

$$\int_{-\infty}^{\infty} e^{-t^2} t^n\,dt = \begin{cases} \sqrt{\pi} & \text{if } n = 0, \\ 0 & \text{if } n \text{ is odd}, \\ \sqrt{\pi}\,\frac{1\cdot 3\cdot 5\cdots(n-1)}{2^{n/2}} & \text{if } n \text{ is even}. \end{cases}$$

2.13 (a) Use the Gram–Schmidt process to find an orthonormal set of vectors out of $(1, -1, 1)$, $(-1, 0, 1)$, and $(2, -1, 2)$.

(b) Are these three vectors linearly independent? If not, find a zero linear combination of them by using part (a).

2.14 (a) Use the Gram–Schmidt process to find an orthonormal set of vectors out of $(1, -1, 2)$, $(-2, 1, -1)$, and $(-1, -1, 4)$.

(b) Are these three vectors linearly independent? If not, find a zero linear combination of them by using part (a).

2.15 Show that

$$\int_{-\infty}^{\infty} \left(t^{10} - t^6 + 5t^4 - 5\right) e^{-t^4}\,dt \le \sqrt{\int_{-\infty}^{\infty} \left(t^4 - 1\right)^2 e^{-t^4}\,dt}\sqrt{\int_{-\infty}^{\infty} \left(t^6 + 5\right)^2 e^{-t^4}\,dt}.$$

Hint: Define an appropriate inner product and use the Schwarz inequality.

2.16 Show that

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(x^5 - x^3 + 2x^2 - 2\right)\left(y^5 - y^3 + 2y^2 - 2\right) e^{-(x^4+y^4)} \le \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(x^4 - 2x^2 + 1\right)\left(y^6 + 4y^3 + 4\right) e^{-(x^4+y^4)}.$$

Hint: Define an appropriate inner product and use the Schwarz inequality.

2.17 Show that for any set of n complex numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$, we have

$$|\alpha_1 + \alpha_2 + \cdots + \alpha_n|^2 \le n\left(|\alpha_1|^2 + |\alpha_2|^2 + \cdots + |\alpha_n|^2\right).$$

Hint: Apply the Schwarz inequality to $(1, 1, \ldots, 1)$ and $(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

2.18 Using the Schwarz inequality show that if $\{\alpha_i\}_{i=1}^{\infty}$ and $\{\beta_i\}_{i=1}^{\infty}$ are in $\mathbb{C}^\infty$, then $\sum_{i=1}^{\infty} \alpha_i^* \beta_i$ is convergent.

2.19 Show that $\mathbf{T} : \mathbb{R}^2 \to \mathbb{R}^3$ given by $\mathbf{T}(x, y) = (x^2 + y^2, x + y, 2x - y)$ is not a linear mapping.

2.20 Verify that all the transformations of Example 2.3.5 are linear.

2.21 Let π be the permutation that takes $(1, 2, 3)$ to $(3, 1, 2)$. Find

$$\mathbf{A}_\pi |e_i\rangle, \quad i = 1, 2, 3,$$

where $\{|e_i\rangle\}_{i=1}^3$ is the standard basis of $\mathbb{R}^3$ (or $\mathbb{C}^3$), and $\mathbf{A}_\pi$ is as defined in Example 2.3.5.

2.22 Show that if $\mathbf{T} \in \mathcal{L}(\mathbb{C}, \mathbb{C})$, then there exists $\alpha \in \mathbb{C}$ such that $\mathbf{T}|a\rangle = \alpha|a\rangle$ for all $|a\rangle \in \mathbb{C}$.

2.23 Show that if $\{|a_i\rangle\}_{i=1}^n$ spans $\mathcal{V}$ and $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then $\{\mathbf{T}|a_i\rangle\}_{i=1}^n$ spans $\mathbf{T}(\mathcal{V})$. In particular, if $\mathbf{T}$ is surjective, then $\{\mathbf{T}|a_i\rangle\}_{i=1}^n$ spans $\mathcal{W}$.

2.24 Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f\big(\alpha|a\rangle\big) = \alpha f\big(|a\rangle\big) \quad \forall\alpha \in \mathbb{R} \text{ and } |a\rangle \in \mathbb{R}^2$$

but f is not linear. Hint: Consider a homogeneous function of degree 1.

2.25 Show that the following transformations are linear:

(a) $\mathcal{V}$ is $\mathbb{C}$ over the reals and $\mathbf{C}|z\rangle = |z^*\rangle$. Is $\mathbf{C}$ linear if instead of real numbers, complex numbers are used as scalars?
(b) $\mathcal{V}$ is $\mathcal{P}^c[t]$ and $\mathbf{T}|x(t)\rangle = |x(t+1)\rangle - |x(t)\rangle$.

2.26 Verify that the kernel of a transformation $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ is a subspace of $\mathcal{V}$, and that $\mathbf{T}(\mathcal{V})$ is a subspace of $\mathcal{W}$.

2.27 Let $\mathcal{V}$ and $\mathcal{W}$ be finite dimensional vector spaces. Show that if $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is surjective, then $\dim W \le \dim V$.

2.28 Suppose that $\mathcal{V}$ is finite dimensional and $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is not zero. Prove that there exists a subspace $\mathcal{U}$ of $\mathcal{V}$ such that $\ker\mathbf{T} \cap \mathcal{U} = \{0\}$ and $\mathbf{T}(\mathcal{V}) = \mathbf{T}(\mathcal{U})$.

2.29 Using Theorem 2.3.11, prove Theorem 2.3.18.

2.30 Using Theorem 2.3.11, prove Theorem 2.3.19.

2.31 Let $B_V = \{|a_i\rangle\}_{i=1}^N$ be a basis for $\mathcal{V}$ and $B_W = \{|b_i\rangle\}_{i=1}^N$ a basis for $\mathcal{W}$. Define the linear transformation $\mathbf{T}|a_i\rangle = |b_i\rangle$, $i = 1, 2, \ldots, N$. Now prove Theorem 2.3.20 by showing that $\mathbf{T}$ is an isomorphism.

2.32 Show that $(\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A}$ for the adjoint given in Definition 2.4.3.

2.33 Show that $\mathcal{W}^0$ is a subspace of $\mathcal{V}^*$ and

$$\dim\mathcal{V} = \dim\mathcal{W} + \dim\mathcal{W}^0.$$

2.34 Show that every vector in the N-dimensional vector space $\mathcal{V}^*$ has $N-1$ linearly independent annihilators. Stated differently, show that a linear functional maps $N-1$ linearly independent vectors to zero.

2.35 Show that $\mathbf{T}$ and $\mathbf{T}^*$ have the same rank. In particular, show that if $\mathbf{T}$ is injective, then $\mathbf{T}^*$ is surjective. Hint: Use the dimension theorem for $\mathbf{T}$ and $\mathbf{T}^*$ and Eq. (2.25).

2.36 Prove Theorem 2.6.3.

2.37 Prove Proposition 2.6.9. Hint: First show that you get zero on both sides if $\{|v_k\rangle\}_{k=1}^N$ are linearly dependent. Next assume their linear independence and choose them as a basis, write $|v\rangle$ in terms of them, and note that

$$\boldsymbol{\Delta}\big(|v\rangle, |v_1\rangle, \ldots, \widehat{|v_j\rangle}, \ldots, |v_N\rangle\big) = 0$$

unless $i = j$.

2.38 Prove Theorem 2.6.11. Hint: For the second part of the theorem, use the fact that an invertible $\mathbf{A}$ maps linearly independent sets of vectors onto linearly independent sets.

http://www.springer.com/978-3-319-01194-3

Mathematical Physics
A Modern Introduction to Its Foundations
Hassani, S.
2013, XXXI, 1205 p. 160 illus., Hardcover
ISBN: 978-3-319-01194-3