CHAPTER 2

Vector Spaces

We now begin our treatment of the principal subject matter of this text. We shall see that all of linear algebra is essentially a study of various transformation properties defined on a vector space, and hence it is only natural that we carefully define vector spaces. This chapter therefore presents a fairly rigorous development of (finite-dimensional) vector spaces, and a discussion of their most important fundamental properties. Basically, the general definition of a vector space is simply an axiomatization of the elementary properties of ordinary three-dimensional Euclidean space.

2.1 DEFINITIONS

A nonempty set V is said to be a **vector space** over a field $\mathcal{F}$ if: (i) there exists an operation called **addition** that associates to each pair $x, y \in V$ a new vector $x + y \in V$ called the **sum** of x and y; (ii) there exists an operation called **scalar multiplication** that associates to each $a \in \mathcal{F}$ and $x \in V$ a new vector $ax \in V$ called the **product** of a and x; (iii) these operations satisfy the following axioms:

(V1) $x + y = y + x$ for all $x, y \in V$.
(V2) $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$.
(V3) There exists an element $0 \in V$ such that $0 + x = x$ for all $x \in V$.

(V4) For all $x \in V$ there exists an element $-x \in V$ such that $x + (-x) = 0$.
(V5) $a(x + y) = ax + ay$ for all $x, y \in V$ and all $a \in \mathcal{F}$.
(V6) $(a + b)x = ax + bx$ for all $x \in V$ and all $a, b \in \mathcal{F}$.
(V7) $a(bx) = (ab)x$ for all $x \in V$ and all $a, b \in \mathcal{F}$.
(V8) $1x = x$ for all $x \in V$ where 1 is the (multiplicative) identity in $\mathcal{F}$.

Note that (V1) – (V4) simply require that V be an additive abelian group. The members of V are called **vectors**, and the members of $\mathcal{F}$ are called scalars. The vector $0 \in V$ is called the **zero vector**, and the vector $-x$ is called the **negative** of the vector x.

We mention only in passing that if we replace the field $\mathcal{F}$ by an arbitrary ring R, then we obtain what is called an **R-module** (or simply a **module** over R). If R is a ring with unit element, then the module is called a **unital** R-module. In fact, this is the only kind of module that is usually considered in treatments of linear algebra. We shall not discuss modules in this text, although the interested reader can learn something about them from several of the books listed in the bibliography.

Throughout this chapter V will always denote a vector space, and the corresponding field $\mathcal{F}$ will be understood even if it is not explicitly mentioned. If $\mathcal{F}$ is the real field $\mathbb{R}$, then we obtain a **real vector space** while if $\mathcal{F}$ is the complex field $\mathbb{C}$, then we obtain a **complex vector space**. It may be easiest for the reader to first think in terms of these spaces rather than the more abstract general case.

Example 2.1 Probably the best known example of a vector space is the set $\mathcal{F}^n = \mathcal{F} \times \cdots \times \mathcal{F}$ of all n-tuples $(a_1, \dots, a_n)$ where each $a_i \in \mathcal{F}$. To make $\mathcal{F}^n$ into a vector space, we define the sum of two elements $(a_1, \dots, a_n) \in \mathcal{F}^n$ and $(b_1, \dots, b_n) \in \mathcal{F}^n$ by

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n)$$

and scalar multiplication for any $k \in \mathcal{F}$ by

$$k(a_1, \dots, a_n) = (ka_1, \dots, ka_n) .$$

If $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$, then we say that $A = B$ if and only if $a_i = b_i$ for each $i = 1, \dots, n$. Defining $0 = (0, \dots, 0)$ and $-A = (-a_1, \dots, -a_n)$ as the identity and inverse elements respectively of $\mathcal{F}^n$, the reader should have no trouble verifying properties (V1) – (V8).

The most common examples of the space $\mathcal{F}^n$ come from considering the fields $\mathbb{R}$ and $\mathbb{C}$. For instance, the space $\mathbb{R}^3$ is (with the Pythagorean notion of distance defined on it) just the ordinary three-dimensional Euclidean space (x, y, z) of elementary physics and geometry.

We shall soon see that any finite-dimensional vector space V over a field $\mathcal{F}$ is essentially the same as the space $\mathcal{F}^n$. In particular, we will prove that V is isomorphic to $\mathcal{F}^n$ for some positive integer n. ∕∕

Example 2.2 Another very useful vector space is the space $\mathcal{F}[x]$ of all polynomials in the indeterminate x over the field $\mathcal{F}$ (polynomials will be defined carefully in Chapter 6). In other words, every element in $\mathcal{F}[x]$ is a polynomial of the form $a_0 + a_1x + \cdots + a_nx^n$ where each $a_i \in \mathcal{F}$ and n is any positive integer (called the degree of the polynomial). Addition and scalar multiplication are defined in the obvious way by

$$\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i = \sum_{i=0}^{n} (a_i + b_i) x^i$$

and

$$c\sum_{i=0}^{n} a_i x^i = \sum_{i=0}^{n} (ca_i) x^i .$$

(If we wish to add together two polynomials $\sum_{i=0}^{n} a_i x^i$ and $\sum_{i=0}^{m} b_i x^i$ where $m > n$, then we simply define $a_i = 0$ for $i = n + 1, \ldots, m$.)

Since we have not yet defined the multiplication of vectors, we ignore the fact that polynomials can be multiplied together. It should be clear that $\mathcal{F}[x]$ does indeed form a vector space. ∕∕

Example 2.3 We can also view the field $\mathbb{C}$ as a vector space over $\mathbb{R}$. In fact, we may generally consider the set of n-tuples $(x_1, \ldots, x_n)$, where each $x_i \in \mathbb{C}$, to be a vector space over $\mathbb{R}$ by defining addition and scalar multiplication (by *real* numbers) as in Example 2.1. We thus obtain a real vector space that is quite distinct from the space $\mathbb{C}^n$. ∕∕

We now prove several useful properties of vector spaces that are analogous to the properties given in Theorem 1.7 for rings.

Theorem 2.1 Let V be a vector space over $\mathcal{F}$. Then for all x, y, z ∈ V and every $a \in \mathcal{F}$ we have

(a) $x + y = z + y$ implies $x = z$.
(b) $ax = 0$ if and only if $a = 0$ or $x = 0$.
(c) $-(ax) = (-a)x = a(-x)$.

Proof We first remark that there is a certain amount of sloppiness in our notation since the symbol 0 is used both as an element of V and as an element of $\mathcal{F}$. However, there should never be any confusion as to which of these sets 0 lies in, and we will continue with this common practice.

(a) If x + y = z + y, then

$$(x + y) + (-y) = (z + y) + (-y)$$

implies

$$x + (y + (-y)) = z + (y + (-y))$$

which implies x + 0 = z + 0 and hence x = z. This is frequently called the (right) **cancellation law**. It is also clear that x + y = x + z implies y = z (left cancellation). (This is just a special case of the general result proved for groups in Section 1.1.)

(b) If a = 0, then

$$0x = (0 + 0)x = 0x + 0x \ .$$

But 0x = 0 + 0x so that 0 + 0x = 0x + 0x, and hence (a) implies 0 = 0x. If x = 0, then

$$a0 = a(0 + 0) = a0 + a0 \ .$$

But a0 = 0 + a0 so that 0 + a0 = a0 + a0, and again we have 0 = a0. Conversely, assume that ax = 0. If a ≠ 0 then a^{-1} exists, and hence

$$x = 1x = (a^{-1}a)x = a^{-1}(ax) = a^{-1}0 = 0$$

by the previous paragraph.

(c) By (V4) we have ax + (−(ax)) = 0, whereas by (b) and (V6), we have

$$0 = 0x = (a + (-a))x = ax + (-a)x \ .$$

Hence ax + (−(ax)) = ax + (−a)x implies −(ax) = (−a)x by (a). Similarly, 0 = x + (−x) so that

$$0 = a0 = a(x + (-x)) = ax + a(-x) \ .$$

Then 0 = ax + (−(ax)) = ax + a(−x) implies −(ax) = a(−x). ∎

In view of this theorem, it makes sense to define **subtraction** in V by

$$x - y = x + (-y) \ .$$

It should then be clear that a vector space will also have the properties we expect, such as a(x − y) = ax − ay, and −(x − y) = −x + y.

If we take an arbitrary subset of vectors in a vector space then, in general, this subset will not be a vector space itself. The reason for this is that in general, even the addition of two vectors in the subset will not result in a vector that is again a member of the subset. Because of this, we make the following definition. Suppose V is a vector space over $\mathcal{F}$ and $W \subset V$. Then if $x, y \in W$ and $c \in \mathcal{F}$ implies $x + y \in W$ and $cx \in W$, we say that W is a **subspace** of V. Indeed, if $c = 0$ then $0 = 0x \in W$ so that $0 \in W$, and similarly $-x = (-1)x \in W$ so that $-x \in W$ also. It is now easy to see that W obeys (V1) – (V8) if V does. It should also be clear that an equivalent way to define a subspace is to require that $cx + y \in W$ for all $x, y \in W$ and all $c \in \mathcal{F}$.

If W is a subspace of V and $W \neq V$, then W is called a **proper** subspace of V. In particular, $W = \{0\}$ is a subspace of V, but it is not very interesting, and hence from now on we assume that any proper subspace contains more than simply the zero vector. (One sometimes refers to $\{0\}$ and V as **trivial** subspaces of V.)

Example 2.4 Consider the elementary Euclidean space $\mathbb{R}^3$ consisting of all triples (x, y, z) of scalars. If we restrict our consideration to those vectors of the form (x, y, 0), then we obtain a subspace of $\mathbb{R}^3$. In fact, this subspace is essentially just the space $\mathbb{R}^2$ which we think of as the usual xy-plane. We leave it as a simple exercise for the reader to show that this does indeed define a subspace of $\mathbb{R}^3$. Note that any other plane parallel to the xy–plane is *not* a subspace. ∫∫

Example 2.5 Let V be a vector space over $\mathcal{F}$, and let $S = \{x_1, \ldots, x_n\}$ be any n vectors in V. Given any set of scalars $\{a_1, \ldots, a_n\}$, the vector

$$\sum_{i=1}^{n} a_i x_i = a_1 x_1 + \cdots + a_n x_n$$

is called a **linear combination** of the n vectors $x_i \in S$, and the set $\mathcal{S}$ of all such linear combinations of elements in S is called the subspace **spanned** (or **generated**) by S. Indeed, if $A = \sum_{i=1}^{n} a_i x_i$ and $B = \sum_{i=1}^{n} b_i x_i$ are vectors in $\mathcal{S}$ and $c \in \mathcal{F}$, then both

$$A + B = \sum_{i=1}^{n} (a_i + b_i) x_i$$

and

$$cA = \sum_{i=1}^{n} (ca_i) x_i$$

are vectors in $\mathcal{S}$. Hence $\mathcal{S}$ is a subspace of V. $\mathcal{S}$ is sometimes called the **linear span** of S, and we say that S **spans** $\mathcal{S}$. ∫∫

In view of this example, we might ask whether or not *every* vector space is in fact the linear span of some set of vectors in the space. In the next section we shall show that this leads naturally to the concept of the dimension of a vector space.

Exercises

1. Verify axioms (V1) – (V8) for the space $\mathcal{F}^n$.

2. Let S be any set, and consider the collection V of all mappings f of S into a field $\mathcal{F}$. For any f, g ∈ V and $\alpha \in \mathcal{F}$, we define $(f + g)(x) = f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$ for every x ∈ S. Show that V together with these operations defines a vector space over $\mathcal{F}$.

3. Consider the two element set {x, y} with addition and scalar multiplication by $c \in \mathcal{F}$ defined by

$$x + x = x \qquad x + y = y + x = y \qquad y + y = x \qquad cx = x \qquad cy = x.$$

   Does this define a vector space over $\mathcal{F}$?

4. Let V be a vector space over $\mathcal{F}$. Show that if x ∈ V and $a, b \in \mathcal{F}$ with $a \neq b$, then $ax = bx$ implies that $x = 0$.

5. Let (V, +, •) be a real vector space with the addition operation denoted by + and the scalar multiplication operation denoted by •. Let $v_0 \in V$ be fixed. We define a new addition operation ⊕ by $x \oplus y = x + y + v_0$, and a new scalar multiplication operation ⊗ by $\alpha \otimes x = \alpha \bullet x + (\alpha - 1) \bullet v_0$. Show that (V, ⊕ , ⊗) defines a real vector space.

6. Let F[ℝ] denote the space of all real-valued functions defined on ℝ with addition and scalar multiplication defined as in Exercise 1.2. In other words, f ∈ F[ℝ] means $f: \mathbb{R} \rightarrow \mathbb{R}$.
   (a) Let C[ℝ] denote the set of all continuous real-valued functions defined on ℝ. Show that C[ℝ] is a subspace of F[ℝ].
   (b) Repeat part (a) with the set D[ℝ] of all such differentiable functions.

7. Referring to the previous exercise, let $D^n[\mathbb{R}]$ denote the set of all n-times differentiable functions from ℝ to ℝ. Consider the subset V of $D^n[\mathbb{R}]$ given by the set of all functions that satisfy the differential equation

$$f^{(n)}(x)+a_{n-1}f^{(n-1)}(x)+a_{n-2}f^{(n-2)}(x)+\cdots+a_1f^{(1)}(x)+a_0f(x)=0$$

where $f^{(i)}(x)$ denotes the i*th* derivative of f(x) and a_i is a fixed real constant. Show that V is a vector space.

8. Let $V = \mathbb{R}^3$. In each of the following cases, determine whether or not the subset W is a subspace of V:
(a) $W = \{(x, y, 0): x, y \in \mathbb{R}\}$ (see Example 2.4).
(b) $W = \{(x, y, z) \in \mathbb{R}^3: z \geq 0\}$.
(c) $W = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 \leq 1\}$.
(d) $W = \{(x, y, z) \in \mathbb{R}^3: x + y + z = 0\}$.
(e) $W = \{(x, y, z) \in \mathbb{R}^3: x, y, z \in \mathbb{Q}\}$.
(f) $W = \{(x, y, z) \in \mathbb{R}^3 - \{0, 0, 0\}\}$.

9. Let S be a nonempty subset of a vector space V. In Example 2.5 we showed that the linear span $\mathcal{S}$ of S is a subspace of V. Show that if W is any other subspace of V containing S, then $\mathcal{S} \subset W$.

10. (a) Determine whether or not the intersection $\cap_{i=1}^{n} W_i$ of a finite number of subspaces W_i of a vector space V is a subspace of V.
(b) Determine whether or not the union $\cup_{i=1}^{n} W_i$ of a finite number of subspaces W_i of a space V is a subspace of V.

11. Let W_1 and W_2 be subspaces of a space V such that $W_1 \cup W_2$ is also a subspace of V. Show that one of the W_i is subset of the other.

12. Let W_1 and W_2 be subspaces of a vector space V. If for every $v \in V$ we have $v = w_1 + w_2$ where $w_i \in W_i$, then we write $V = W_1 + W_2$ and say that V is the **sum** of the subspaces W_i. If $V = W_1 + W_2$ and $W_1 \cap W_2 = \{0\}$, show that every $v \in V$ has a *unique* representation $v = w_1 + w_2$ with $w_i \in W_i$.

13. Let V be the set of all (infinite) real sequences. In other words, any $v \in V$ is of the form $(x_1, x_2, x_3, \ldots)$ where each $x_i \in \mathbb{R}$. If we define the addition and scalar multiplication of distinct sequences componentwise exactly as in Example 2.1, then it should be clear that V is a vector space over $\mathbb{R}$. Determine whether or not each of the following subsets of V in fact forms a subspace of V:
(a) All sequences containing only a finite number of nonzero terms.

(b) All sequences of the form $\{x_1, x_2, \ldots, x_N, 0, 0, \ldots\}$ where N is fixed.

(c) All **decreasing sequences**, i.e., sequences where $x_{k+1} \le x_k$ for each $k = 1, 2, \ldots$.

(d) All convergent sequences, i.e., sequences for which $\lim_{k\to\infty} x_k$ exists.

14. For which value of k will the vector $v = (1, -2, k) \in \mathbb{R}^3$ be a linear combination of the vectors $x_1 = (3, 0, -2)$ and $x_2 = (2, -1, -5)$?

15. Write the vector $v = (1, -2, 5)$ as a linear combination of the vectors $x_1 = (1, 1, 1)$, $x_2 = (1, 2, 3)$ and $x_3 = (2, -1, 1)$.

2.2 LINEAR INDEPENDENCE AND BASES

Let $x_1, \ldots, x_n$ be vectors in a vector space V. We say that these vectors are **linearly dependent** if there exist scalars $a_1, \ldots, a_n \in \mathcal{F}$, not all equal to 0, such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i = 0.$$

The vectors x_i are said to be **linearly independent** if they are not linearly dependent. From these definitions, it follows that any set containing a linearly dependent subset must be linearly dependent, and any subset of a linearly independent set is necessarily linearly independent.

It is important to realize that a set of vectors may be linearly dependent with respect to one field, but independent with respect to another. For example, the set $\mathbb{C}$ of all complex numbers is itself a vector space over either the field of real numbers or over the field of complex numbers. However, the set $\{x_1 = 1, x_2 = i\}$ is linearly independent if $\mathcal{F} = \mathbb{R}$, but linearly dependent if $\mathcal{F} = \mathbb{C}$ since $ix_1 + (-1)x_2 = 0$. We will always assume that a linear combination is taken with respect to the same field that V is defined over.

As a means of simplifying our notation, we will frequently leave off the limits of a sum when there is no possibility of ambiguity. Thus, if we are considering the set $\{x_1, \ldots, x_n\}$, then a linear combination of the x_i will often be written as $\sum a_ix_i$ rather than $\sum_{i=1}^{n} a_ix_i$. In addition, we will often denote a collection $\{x_1, \ldots, x_n\}$ of vectors simply by $\{x_i\}$.

Example 2.6 Consider the three vectors in $\mathbb{R}^3$ given by

$$e_1 = (1, 0, 0)$$
$$e_2 = (0, 1, 0)$$
$$e_3 = (0, 0, 1).$$

Using the definitions of addition and scalar multiplication given in Example 2.1, it is easy to see that these three vectors are linearly independent. This is because the zero vector in $\mathbb{R}^3$ is given by (0, 0, 0), and hence

$$a_1 e_1 + a_2 e_2 + a_3 e_3 = (a_1, a_2, a_3) = (0, 0, 0)$$

implies that $a_1 = a_2 = a_3 = 0$.

On the other hand, the vectors

$$x_1 = (1, 0, 0)$$
$$x_2 = (0, 1, 2)$$
$$x_3 = (1, 3, 6)$$

are linearly dependent since $x_3 = x_1 + 3x_2$. ∕∕

Theorem 2.2 A finite set S of vectors in a space V is linearly dependent if and only if one vector in the set is a linear combination of the others. In other words, S is linearly dependent if one vector in S is in the subspace spanned by the remaining vectors in S.

Proof If $S = \{x_1, \dots, x_n\}$ is a linearly dependent subset of V, then

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$$

for some set of scalars $a_1, \dots, a_n \in \mathcal{F}$ not all equal to 0. Suppose, to be specific, that $a_1 \neq 0$. Then we may write

$$x_1 = -(a_2/a_1)x_2 - \cdots - (a_n/a_1)x_n$$

which shows that x_1 is a linear combination of $x_2, \dots, x_n$.

Conversely, if $x_1 = \Sigma_{i\neq1}a_ix_i$ then

$$x_1 + (-a_2)x_2 + \cdots + (-a_n)x_n = 0$$

which shows that the collection $\{x_1, \dots, x_n\}$ is linearly dependent. ∎

It is important to realize that no linearly independent set of vectors can contain the zero vector. To see this, note that if $S = \{x_1, \ldots, x_n\}$ and $x_1 = 0$, then $ax_1 + 0x_2 + \cdots + 0x_n = 0$ for all $a \in \mathcal{F}$, and hence by definition, S is a linearly dependent set.

Theorem 2.3 Let $S = \{x_1, \ldots, x_n\} \subset V$ be a linearly independent set, and let $\mathcal{S}$ be the linear span of S. Then every $v \in \mathcal{S}$ has a unique representation

$$v = \sum_{i=1}^{n} a_i x_i$$

where each $a_i \in \mathcal{F}$.

Proof By definition of $\mathcal{S}$, we can always write $v = \sum a_i x_i$. As to uniqueness, it must be shown that if we also have $v = \sum b_i x_i$, then it follows that $b_i = a_i$ for every $i = 1, \ldots, n$. But this is easy since $\sum a_i x_i = \sum b_i x_i$ implies $\sum (a_i - b_i) x_i = 0$, and hence $a_i - b_i = 0$ (since $\{x_i\}$ is linearly independent). Therefore $a_i = b_i$ for each $i = 1, \ldots, n$. ∎

If S is a finite subset of a vector space V such that $V = \mathcal{S}$ (the linear span of S), then we say that V is **finite-dimensional**. However, we must define what is meant in general by the dimension of V. If $S \subset V$ is a linearly independent set of vectors with the property that $V = \mathcal{S}$, then we say that S is a **basis** for V. In other words, a **basis** for V is a linearly independent set that spans V. We shall see that the number of elements in a basis is what is meant by the **dimension** of V. But before we can state this precisely, we must be sure that such a number is well-defined. In other words, we must show that any basis has the same number of elements. We prove this (see the corollary to Theorem 2.6) in several steps.

Theorem 2.4 Let $\mathcal{S}$ be the linear span of $S = \{x_1, \ldots, x_n\} \subset V$. If $k \le n$ and $\{x_1, \ldots, x_k\}$ is linearly independent, then there exists a linearly independent subset of S of the form $\{x_1, \ldots, x_k, x_{i_1}, \ldots, x_{i_\alpha}\}$ whose linear span also equals $\mathcal{S}$.

Proof If $k = n$ there is nothing left to prove, so we assume that $k < n$. Since $x_1, \ldots, x_k$ are linearly independent, we let x_j (where $j > k$) be the first vector in S that is a linear combination of the preceding $x_1, \ldots, x_{j-1}$. If no such j exists, then take $(i_1, \ldots, i_\alpha) = (k + 1, \ldots, n)$. Then the set of $n - 1$ vectors $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ has a linear span that must be contained in $\mathcal{S}$ (since this set is just a subset of S). However, if v is any vector in $\mathcal{S}$, we can write $v = \sum_{i=1}^{n} a_i x_i$ where x_j is just a linear combination of the first $j - 1$ vectors. In other words, v is a linear combination of $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ and hence these n - 1 vectors also span $\mathcal{S}$.

We now continue this process by picking out the first vector in this set of n - 1 vectors that is a linear combination of the preceding vectors. An identical argument shows that the linear span of this set of n - 2 vectors must also be $\mathcal{S}$. It is clear that we will eventually obtain a set $\{x_1, \ldots, x_k, x_{i_1}, \ldots, x_{i_\alpha}\}$ whose linear span is still $\mathcal{S}$, but in which no vector is a linear combination of the preceding ones. This means that the set must be linearly independent (Theorem 2.2). ∎

Corollary 1 If V is a finite-dimensional vector space such that the set S = $\{x_1, \ldots, x_m\} \subset V$ spans V, then some subset of S is a basis for V.

Proof By Theorem 2.4, S contains a linearly independent subset that also spans V. But this is precisely the requirement that S contain a basis for V. ∎

Corollary 2 Let V be a finite-dimensional vector space and let $\{x_1, \ldots, x_n\}$ be a basis for V. Then any element $v \in V$ has a unique representation of the form

$$v = \sum_{i=1}^{n} a_i x_i$$

where each $a_i \in \mathcal{F}$.

Proof Since $\{x_i\}$ is linearly independent and spans V, Theorem 2.3 shows us that any $v \in V$ may be written in the form $v = \sum_{i=1}^{n} a_i x_i$ where each $a_i \in \mathcal{F}$ is unique (for this particular basis). ∎

It is important to realize that Corollary 1 asserts the existence of a finite basis in any finite-dimensional vector space, but says nothing about the uniqueness of this basis. In fact, there are an infinite number of possible bases for any such space. However, by Corollary 2, once a particular basis has been chosen, then any vector has a unique expansion in terms of this basis.

Example 2.7 Returning to the space $\mathcal{F}^n$, we see that any $(a_1, \ldots, a_n) \in \mathcal{F}^n$ can be written as the linear combination

$$a_1(1, 0, \ldots, 0) + a_2(0, 1, 0, \ldots, 0) + \cdots + a_n(0, \ldots, 0, 1) \;.$$

This means that the n vectors

$$
\begin{aligned}
e_1 &= (1, 0, 0, \dots, 0)\\
e_2 &= (0, 1, 0, \dots, 0)\\
&\vdots\\
e_n &= (0, 0, 0, \dots, 1)
\end{aligned}
$$

span $\mathcal{F}^n$. They are also linearly independent since $\sum a_i e_i = (a_1, \dots, a_n) = 0$ if and only if $a_i = 0$ for all $i = 1, \dots, n$. The set $\{e_i\}$ is extremely useful, and will be referred to as the **standard basis** for $\mathcal{F}^n$. ∕∕

This example leads us to make the following generalization. By an **ordered basis** for a finite-dimensional space V, we mean a finite sequence of vectors that is linearly independent and spans V. If the sequence $x_1, \dots, x_n$ is an ordered basis for V, then the set $\{x_1, \dots, x_n\}$ is a basis for V. In other words, the set $\{x_1, \dots, x_n\}$ gives rise to n! different ordered bases. Since there is usually nothing lost in assuming that a basis is ordered, we shall continue to assume that $\{x_1, \dots, x_n\}$ denotes an ordered basis unless otherwise noted.

Given any (ordered) basis $\{x_1, \dots, x_n\}$ for V, we know that any $v \in V$ has a unique representation $v = \sum_{i=1}^n a_i x_i$. We call the scalars $a_1, \dots, a_n$ the **coordinates** of v relative to the (ordered) basis $\{x_1, \dots, x_n\}$. In particular, we call a_i the *ith* coordinate of v. Moreover, we now proceed to show that these coordinates define an isomorphism between V and $\mathcal{F}^n$.

Since a vector space is also an (additive abelian) group, it is reasonable that we make the following definition. Let V and W be vector spaces over $\mathcal{F}$. We say that a mapping $\phi: V \to W$ is a **vector space homomorphism** (or, as we shall call it later, a **linear transformation**) if

$$\phi(x + y) = \phi(x) + \phi(y)$$

and

$$\phi(ax) = a\phi(x)$$

for all $x, y \in V$ and $a \in \mathcal{F}$. This agrees with our previous definition for groups, except that now we must take into account the multiplication by scalars. If ϕ is injective, then we say that ϕ is an **isomorphism**, and if ϕ is bijective, that V and W are **isomorphic**.

As before, we define the kernel of ϕ to be the set

$$\operatorname{Ker} \phi = \{x \in V: \phi(x) = 0 \in W\} .$$

If $x, y \in \operatorname{Ker} \phi$ and $c \in F$ we have

$$\phi(x + y) = \phi(x) + \phi(y) = 0$$

and

$$\phi(cx) = c\phi(x) = c0 = 0 .$$

This shows that both x + y and cx are in Ker ϕ, and hence Ker ϕ is a subspace of V. Note also that if a = 0 and $x \in V$ then

$$\phi(0) = \phi(ax) = a\phi(x) = 0 .$$

Alternatively, we could also note that

$$\phi(x) = \phi(x + 0) = \phi(x) + \phi(0)$$

and hence $\phi(0) = 0$. Finally, we see that

$$0 = \phi(0) = \phi(x + (-x)) = \phi(x) + \phi(-x)$$

and therefore

$$\phi(-x) = -\phi(x) .$$

Our next result is essentially the content of Theorem 1.6 and its corollary.

Theorem 2.5 Let ϕ: V → W be a vector space homomorphism. Then ϕ is an isomorphism if and only if Ker $\phi = \{0\}$.

Proof If ϕ is injective, then the fact that $\phi(0) = 0$ implies that we must have Ker $\phi = \{0\}$. Conversely, if Ker $\phi = \{0\}$ and $\phi(x) = \phi(y)$, then

$$0 = \phi(x) - \phi(y) = \phi(x - y)$$

implies that $x - y = 0$, or $x = y$. ∎

Now let us return to the above notion of an ordered basis. For any finite-dimensional vector space V over $\mathcal{F}$ and any (ordered) basis $\{x_1, \ldots, x_n\}$, we define a mapping $\phi: V \to \mathcal{F}^n$ by

$$\phi(v) = \phi\left(\sum_{i=1}^{n} a_i x_i\right) = (a_1, \ldots, a_n)$$

for each

$$v = \sum_{i=1}^{n} a_i x_i \in V.$$

Since

$$\begin{aligned}\phi(\Sigma a_i x_i + \Sigma b_i x_i) &= \phi(\Sigma(a_i + b_i)x_i)\\ &= (a_1 + b_1, \ldots, a_n + b_n)\\ &= (a_1, \ldots, a_n) + (b_1, \ldots, b_n)\\ &= \phi(\Sigma a_i x_i) + \phi(\Sigma b_i x_i)\end{aligned}$$

and

$$\begin{aligned}\phi(kv) &= \phi(k\Sigma a_i x_i) = \phi(\Sigma(ka_i)x_i) = (ka_1, \ldots, ka_n) = k(a_1, \ldots, a_n)\\ &= k\phi(v)\end{aligned}$$

we see that ϕ is a vector space homomorphism. Because the coordinates of any vector are unique for a fixed basis, we see that this mapping is indeed well-defined and one-to-one. (Alternatively, the identity element in the space $\mathcal{F}^n$ is $(0, \ldots, 0)$, and the only vector that maps into this is the zero vector in V. Hence $\text{Ker}\,\phi = \{0\}$ and ϕ is an isomorphism.) It is clear that ϕ is surjective since, given any ordered set of scalars $a_1, \ldots, a_n \in \mathcal{F}$, we can define the vector $v = \Sigma a_i x_i \in V$. Therefore we have shown that V and $\mathcal{F}^n$ are isomorphic for some n, where n is the number of vectors in an ordered basis for V.

If V has a basis consisting of n elements, is it possible to find another basis consisting of $m \neq n$ elements? Intuitively we guess not, for if this were true then V would be isomorphic to $\mathcal{F}^m$ as well as to $\mathcal{F}^n$, which implies that $\mathcal{F}^m$ is isomorphic to $\mathcal{F}^n$ for $m \neq n$. That this is not possible should be obvious by simply considering the projection of a point in $\mathbb{R}^3$ down onto the plane $\mathbb{R}^2$. Any point in $\mathbb{R}^2$ is thus the image of an entire vertical line in $\mathbb{R}^3$, and hence this projection can not possibly be an isomorphism. Nevertheless, we proceed to prove this in detail beginning with our next theorem.

Theorem 2.6 Let $\{x_1, \ldots, x_n\}$ be a basis for V, and let $\{y_1, \ldots, y_m\}$ be linearly independent vectors in V. Then $m \leq n$.

Proof Since $\{x_1, \ldots, x_n\}$ spans V, we may write each y_i as a linear combination of the x_j. In particular, choosing y_m, it follows that the set

$$\{y_m, x_1, \ldots, x_n\}$$

is linearly dependent (Theorem 2.2) and spans V (since the x_k already do so). Hence there must be a proper subset $\{y_m, x_{i_1}, \ldots, x_{i_r}\}$ with $r \leq n - 1$ that forms a basis for V (Theorem 2.4). Now this set spans V so that y_{m-1} is a linear combination of this set, and hence

$$\{y_{m-1}, y_m, x_{i_1}, \ldots, x_{i_r}\}$$

is linearly dependent and spans V. By Theorem 2.4 again, we can find a set $\{y_{m-1}, y_m, x_{j_1}, \ldots, x_{j_s}\}$ with $s \le n - 2$ that is also a basis for V. Continuing our process, we eventually obtain the set

$$\{y_2, \ldots, y_m, x_\alpha, x_\beta, \ldots\}$$

which spans V and must contain at least one of the x_k (since y_1 is not a linear combination of the set $\{y_2, \ldots, y_m\}$ by hypothesis). This set was constructed by adding $m - 1$ vectors y_i to the original set of n vectors x_k, and deleting at least $m - 1$ of the x_k along the way. However, we still have at least one of the x_k in our set, and hence it follows that $m - 1 \le n - 1$ or $m \le n$. ∎

Corollary Any two bases for a finite-dimensional vector space must consist of the same number of elements.

Proof Let $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ be bases for V. Since the y_i are linearly independent, Theorem 2.6 says that $m \le n$. On the other hand, the x_j are linearly independent so that $n \le m$. Therefore we must have $n = m$. ∎

We now return to the proof that $\mathcal{F}^m$ is isomorphic to $\mathcal{F}^n$ if and only if $m = n$. Let us first show that an isomorphism maps a basis to a basis.

Theorem 2.7 Let $\phi: V \to W$ be an isomorphism of finite-dimensional vector spaces. Then a set of vectors $\{\phi(v_1), \ldots, \phi(v_n)\}$ is linearly dependent in W if and only if the set $\{v_1, \ldots, v_n\}$ is linearly dependent in V.

Proof If the set $\{v_1, \ldots, v_n\}$ is linearly dependent, then for some set of scalars $\{a_1, \ldots, a_n\}$ not all equal to 0 we have $\sum_{i=1}^n a_i v_i = 0$. Applying ϕ to both sides of this equation yields

$$0 = \phi(0) = \phi(\sum a_i v_i) = \sum \phi(a_i v_i) = \sum a_i \phi(v_i) .$$

But since not all of the a_i are 0, this means that $\{\phi(v_i)\}$ must be linearly dependent.

Conversely, if $\phi(v_1), \ldots, \phi(v_n)$ are linearly dependent, then there exists a set of scalars $b_1, \ldots, b_n$ not all 0 such that $\sum b_i \phi(v_i) = 0$. But this means

$$0 = \sum b_i \phi(v_i) = \sum \phi(b_i v_i) = \phi(\sum b_i v_i)$$

which implies that $\sum b_i v_i = 0$ (since Ker $\phi = \{0\}$). This shows that the set $\{v_i\}$ is linearly dependent. ∎

Corollary If $\phi\colon V \to W$ is an isomorphism of finite-dimensional vector spaces, then $\{\phi(x_i)\} = \{\phi(x_1), \dots, \phi(x_n)\}$ is a basis for W if and only if $\{x_i\} = \{x_1, \dots, x_n\}$ is a basis for V.

Proof Since ϕ is an isomorphism, for any vector $w \in W$ there exists a unique $v \in V$ such that $\phi(v) = w$. If $\{x_i\}$ is a basis for V, then $v = \sum_{i=1}^n a_i x_i$ and

$$w = \phi(v) = \phi(\textstyle\sum a_i x_i) = \sum a_i \phi(x_i) \ .$$

Hence the $\phi(x_i)$ span W, and they are linearly independent by Theorem 2.7.

On the other hand, if $\{\phi(x_i)\}$ is a basis for W, then there exist scalars $\{b_i\}$ such that for any $v \in V$ we have

$$\phi(v) = w = \textstyle\sum b_i \phi(x_i) = \phi(\sum b_i x_i) \ .$$

Since ϕ is an isomorphism, this implies that $v = \sum b_i x_i$, and hence $\{x_i\}$ spans V. The fact that it is linearly independent follows from Theorem 2.7. This shows that $\{x_i\}$ is a basis for V. ∎

Theorem 2.8 $\mathcal{F}^n$ is isomorphic to $\mathcal{F}^m$ if and only if $n = m$.

Proof If $n = m$ the result is obvious. Now assume that $\mathcal{F}^n$ and $\mathcal{F}^m$ are isomorphic. We have seen in Example 2.7 that the standard basis of $\mathcal{F}^n$ consists of n vectors. Since an isomorphism carries one basis onto another (corollary to Theorem 2.7), any space isomorphic to $\mathcal{F}^n$ must have a basis consisting of n vectors. Hence by the corollary to Theorem 2.6 we must have $m = n$. ∎

Corollary If V is a finite-dimensional vector space over $\mathcal{F}$, then V is isomorphic to $\mathcal{F}^n$ for a unique integer n.

Proof It was shown following Theorem 2.5 that V is isomorphic to $\mathcal{F}^n$ for some integer n, and Theorem 2.8 shows that n must be unique. ∎

The corollary to Theorem 2.6 shows us that the number of elements in any basis for a finite-dimensional vector space is fixed. We call this unique number n the **dimension** of V over $\mathcal{F}$, and we write dim $V = n$. Our next result agrees with our intuition, and is quite useful in proving other theorems.

Theorem 2.9 Every subspace W of a finite-dimensional vector space V is finite-dimensional, and dim W ≤ dim V.

Proof We must show that W has a basis, and that this basis contains at most n = dim V elements. If W = {0}, then dim W = 0 ≤ n and we are done. If W contains some $x_1 \neq 0$, then let $W_1 \subset W$ be the subspace spanned by x_1. If $W = W_1$, then dim W = 1 and we are done. If $W \neq W_1$, then there exists some $x_2 \in W$ with $x_2 \notin W_1$, and we let W_2 be the subspace spanned by $\{x_1, x_2\}$. Again, if $W = W_2$, then dim W = 2. If $W \neq W_2$, then choose some $x_3 \in W$ with $x_3 \notin W_2$ and continue this procedure. However, by Theorem 2.6, there can be at most n linearly independent vectors in V, and hence dim W ≤ n. ∎

Note that the zero subspace is spanned by the vector 0, but {0} is not linearly independent so it can not form a basis. Therefore the zero subspace is *defined* to have dimension zero.

Finally, let us show that any set of linearly independent vectors may be extended to form a complete basis.

Theorem 2.10 Let V be finite-dimensional and $S = \{x_1, \dots, x_m\}$ any set of m linearly independent vectors in V. Then there exists a set $\{x_{m+1}, \dots, x_{m+r}\}$ of vectors in V such that $\{x_1, \dots, x_{m+r}\}$ is a basis for V.

Proof Since V is finite-dimensional, it has a basis $\{v_1, \dots, v_n\}$. Then the set $\{x_1, \dots, x_m, v_1, \dots, v_n\}$ spans V so, by Theorem 2.4, we can choose a subset $\{x_1, \dots, x_m, v_{i_1}, \dots, v_{i_r}\}$ of linearly independent vectors that span V. Letting $v_{i_1} = x_{m+1}, \dots, v_{i_r} = x_{m+r}$ proves the theorem. ∎

Exercises

1. Determine whether or not the three vectors $x_1 = (2, -1, 0)$, $x_2 = (1, -1, 1)$ and $x_3 = (0, 2, 3)$ form a basis for $\mathbb{R}^3$.

2. In each of the following, show that the given set of vectors is linearly independent, and decide whether or not it forms a basis for the indicated space:
   (a) $\{(1, 1), (1, -1)\}$ in $\mathbb{R}^2$.
   (b) $\{(2, 0, 1), (1, 2, 0), (0, 1, 0)\}$ in $\mathbb{R}^3$.
   (c) $\{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\}$ in $\mathbb{R}^4$.

3. Extend each of the following sets to a basis for the given space:
(a) $\{(1, 1, 0), (2, -2, 0)\}$ in $\mathbb{R}^3$.
(b) $\{(1, 0, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1)\}$ in $\mathbb{R}^4$.
(c) $\{(1, 1, 0, 0), (1, -1, 0, 0), (1, 0, 1, 0)\}$ in $\mathbb{R}^4$.

4. Show that the vectors $u = (1 + i, 2i)$, $v = (1, 1 + i) \in \mathbb{C}^2$ are linearly dependent over $\mathbb{C}$, but linearly independent over $\mathbb{R}$.

5. Find the coordinates of the vector $(3, 1, -4) \in \mathbb{R}^3$ relative to the basis $x_1 = (1, 1, 1)$, $x_2 = (0, 1, 1)$ and $x_3 = (0, 0, 1)$.

6. Let $\mathbb{R}_3[x]$ be the space of all real polynomials of degree ≤ 3. Determine whether or not each of the following sets of polynomials is linearly independent:
(a) $\{x^3 - 3x^2 + 5x + 1, x^3 - x^2 + 8x + 2, 2x^3 - 4x^2 + 9x + 5\}$.
(b) $\{x^3 + 4x^2 - 2x + 3, x^3 + 6x^2 - x + 4, 3x^3 + 8x^2 - 8x + 7\}$.

7. Let V be a finite-dimensional space, and let W be any subspace of V. Show that there exists a subspace W′ of V such that $W \cap W' = \{0\}$ and $V = W + W'$ (see Exercise 1.12 for the definition of $W + W'$).

8. Let $\phi: V \rightarrow W$ be a homomorphism of two vector spaces V and W.
(a) Show that ϕ maps any subspace of V onto a subspace of W.
(b) Let S′ be a subspace of W, and define the set $S = \{x \in V: \phi(x) \in S'\}$. Show that S is a subspace of V.

9. Let V be finite-dimensional, and assume that $\phi: V \rightarrow V$ is a surjective homomorphism. Prove that ϕ is in fact an isomorphism of V onto V.

10. Let V have basis $x_1, x_2, \ldots, x_n$, and let $v_1, v_2, \ldots, v_n$ be any n elements in V. Define a mapping $\phi: V \rightarrow V$ by

$$\phi\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i v_i$$

where each $a_i \in \mathcal{F}$.
(a) Show that ϕ is a surjective homomorphism.
(b) When is ϕ an isomorphism?

2.3 DIRECT SUMS

We now present some useful ways of constructing a new vector space from several given spaces. The reader is advised to think carefully about these concepts, as they will become quite important later in this book. We also repeat our earlier remark that all of the vector spaces that we are discussing are considered to be defined over the same field $\mathcal{F}$.

Let A and B be subspaces of a finite-dimensional vector space V. Then we may define the **sum** of A and B to be the set A + B given by

$$A + B = \{a + b : a \in A \text{ and } b \in B\} .$$

It is important to note that A and B must both be subspaces of the same space V, or else the addition of $a \in A$ to $b \in B$ is not defined. In fact, since A and B are subspaces of V, it is easy to show that A + B is also subspace of V. Indeed, given any $a_1 + b_1$ and $a_2 + b_2$ in A + B and any $k \in \mathcal{F}$ we see that

$$(a_1 + b_1) + (a_2 + b_2) = (a_1 + a_2) + (b_1 + b_2) \in A + B$$

and

$$k(a_1 + b_1) = ka_1 + kb_1 \in A + B$$

as required. This definition can clearly be extended by induction to any finite collection $\{A_i\}$ of subspaces.

In addition to the sum of the subspaces A and B, we may define their **intersection** $A \cap B$ by

$$A \cap B = \{x \in V : x \in A \text{ and } x \in B\} .$$

Since A and B are subspaces, we see that for any $x, y \in A \cap B$ we have both $x + y \in A$ and $x + y \in B$ so that $x + y \in A \cap B$, and if $x \in A \cap B$ then $kx \in A$ and $kx \in B$ so that $kx \in A \cap B$. Since $0 \in A \cap B$, we then see that $A \cap B$ is a nonempty subspace of V. This can also be extended to any finite collection of subspaces of V.

Our next theorem shows that the dimension of the sum of A and B is just the sum of the dimensions of A and B minus the dimension of their intersection.

Theorem 2.11 If A and B are subspaces of a finite-dimensional space V, then

$$\dim(A + B) = \dim A + \dim B - \dim(A \cap B) .$$

Proof Since A + B and A ∩ B are subspaces of V, it follows that both A + B and A ∩ B are finite-dimensional (Theorem 2.9). We thus let dim A = m, dim B = n and dim A ∩ B = r.

Let $\{u_1, \dots, u_r\}$ be a basis for A ∩ B. By Theorem 2.10 there exists a set $\{v_1, \dots, v_{m-r}\}$ of linearly independent vectors in V such that

$$\{u_1, \dots, u_r, v_1, \dots, v_{m-r}\}$$

is a basis for A. Similarly, we have a basis

$$\{u_1, \dots, u_r, w_1, \dots, w_{n-r}\}$$

for B. It is clear that the set

$$\{u_1, \dots, u_r, v_1, \dots, v_{m-r}, w_1, \dots, w_{n-r}\}$$

spans A + B since any a + b ∈ A + B (with a ∈ A and b ∈ B) can be written as a linear combination of these r + (m - r) + (n - r) = m + n - r vectors. To prove that they form a basis for A + B, we need only show that these m + n - r vectors are linearly independent.

Suppose we have sets of scalars $\{a_i\}$, $\{b_j\}$ and $\{c_k\}$ such that

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m-r} b_j v_j + \sum_{k=1}^{n-r} c_k w_k = 0$$

Then

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m-r} b_j v_j = -\sum_{k=1}^{n-r} c_k w_k .$$

Since the left side of this equation is an element of A while the right side is an element of B, their equality implies that they both belong to A ∩ B, and hence

$$-\sum_{k=1}^{n-r} c_k w_k = \sum_{i=1}^{r} d_i u_i$$

for some set of scalars $\{d_i\}$. But $\{u_1, \dots, u_r, w_1, \dots, w_{n-r}\}$ forms a basis for B and hence they are linearly independent. Therefore, writing the above equation as

$$\sum_{i=1}^{r} d_i u_i + \sum_{k=1}^{n-r} c_k w_k = 0$$

implies that

$$d_1 = \cdots = d_r = c_1 = \cdots = c_{n-r} = 0 .$$

We are now left with

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m-r} b_j v_j = 0.$$

But $\{u_1, \ldots, u_r, v_1, \ldots, v_{m-r}\}$ is also linearly independent so that

$$a_1 = \cdots = a_r = b_1 = \cdots = b_{m-r} = 0\ .$$

This proves that $\{u_1, \ldots, u_r, v_1, \ldots, v_{m-r}, w_1, \ldots, w_{n-r}\}$ is linearly independent as claimed. The proof is completed by simply noting that we have shown

$$\dim(A + B) = m + n - r = \dim A + \dim B - \dim(A \cap B)\ . \ ∎$$

We now consider a particularly important special case of the sum. If A and B are subspaces of V such that $A \cap B = \{0\}$ and $V = A + B$, then we say that V is the **internal direct sum** of A and B. A completely equivalent way of defining the internal direct sum is given in the following theorem.

Theorem 2.12 Let A and B be subspaces of a finite-dimensional vector space V. Then V is the internal direct sum of A and B if and only if every $v \in V$ can be *uniquely* written in the form $v = a + b$ where $a \in A$ and $b \in B$.

Proof Let us first assume that V is the internal direct sum of A and B. In other words, $V = A + B$ and $A \cap B = \{0\}$. Then by definition, for any $v \in V$ we have $v = a + b$ for some $a \in A$ and $b \in B$. Suppose we also have $v = a' + b'$ where $a' \in A$ and $b' \in B$. Then $a + b = a' + b'$ so that $a - a' = b' - b$. But note that $a - a' \in A$ and $b' - b \in B$, and hence the fact that $A \cap B = \{0\}$ implies that $a - a' = b' - b = 0$. Therefore $a = a'$ and $b = b'$ so that the expression for v is unique.

Conversely, suppose that every $v \in V$ may be written uniquely in the form $v = a + b$ with $a \in A$ and $b \in B$. This means that $V = A + B$, and we must still show that $A \cap B = \{0\}$. In particular, if $v \in A \cap B$ we may write $v = v + 0$ with $v \in A$ and $0 \in B$, or alternatively, we may write $v = 0 + v$ with $0 \in A$ and $v \in B$. But we are assuming that the expression for v is unique, and hence we must have $v = 0$ (since the contributions from A and B must be the same in both cases). Thus $A \cap B = \{0\}$ and the sum is direct. ∎

We emphasize that the internal direct sum is defined for two subspaces A and B of a given space V. As we stated above, this is because the addition of two vectors from distinct spaces is not defined. In spite of this, we now proceed to show that it is nevertheless possible to define the sum of two distinct vector spaces.

Let A and B be distinct vector spaces (over the same field $\mathcal{F}$, of course). While the sum of a vector in A and a vector in B makes no sense, we may relate these two spaces by considering the Cartesian product A × B defined as (see Section 0.1)

$$A \times B = \{(a, b): a \in A \text{ and } b \in B\} .$$

Using the ordered pairs (a, b), it is now easy to turn A × B into a vector space by making the following definitions (see Example 2.1).

First, we say that two elements (a, b) and (a′, b′) of A × B are equal if and only if a = a′ and b = b′. Next, we define addition and scalar multiplication in the obvious manner by

$$(a, b) + (a', b') = (a + a', b + b')$$

and

$$k(a, b) = (ka, kb) .$$

We leave it as an exercise for the reader to show that with these definitions, the set A × B defines a vector space V over $\mathcal{F}$. This vector space is called the **external direct sum** of the spaces A and B, and is denoted by A ⊕ B.

While the external direct sum was defined for arbitrary spaces A and B, there is no reason why this definition can not be applied to two subspaces of a larger space V. We now show that in such a case, the internal and external direct sums are isomorphic.

Theorem 2.13 If V is the internal direct sum of A and B, then V is isomorphic to the external direct sum A ⊕ B.

Proof If V is the internal direct sum of A and B, then any v ∈ V may be written uniquely in the form v = a + b. This uniqueness allows us to define the mapping $\phi: V \to A \oplus B$ by

$$\phi(v) = \phi(a + b) = (a, b) .$$

Since for any v = a + b and v′ = a′ + b′, and for any scalar k we have

$$\phi(v + v') = (a + a', b + b') = (a, b) + (a', b') = \phi(v) + \phi(v')$$

and

$$\phi(kv) = (ka, kb) = k(a, b) = k\phi(v)$$

it follows that ϕ is a vector space homomorphism. It is clear that ϕ is surjective, since for any (a, b) ∈ A ⊕ B we have $\phi(v) = (a, b)$ where v = a + b ∈ V. Finally, if $\phi(v) = (0, 0)$ then we must have a = b = 0 = v and hence Ker ϕ =

{0}. This shows that ϕ is also injective (Theorem 2.5). In other words, we have shown that V is isomorphic to $A \oplus B$. ∎

Because of this theorem, we shall henceforth refer only to the **direct sum** of A and B, and denote this sum by $A \oplus B$. It follows trivially from Theorem 2.11 that

$$\dim(A \oplus B) = \dim A + \dim B \ .$$

Example 2.8 Consider the ordinary Euclidean three-space $V = \mathbb{R}^3$. Note that any $v \in \mathbb{R}^3$ may be written as

$$(v_1, v_2, v_3) = (v_1, v_2, 0) + (0, 0, v_3)$$

which is just the sum of a vector in the xy-plane and a vector on the z-axis. It should also be clear that the only vector in the intersection of the xy-plane with the z-axis is the zero vector. In other words, defining the space A to be the xy-plane $\mathbb{R}^2$ and the space B to be the z-axis $\mathbb{R}^1$, we see that $V = A \oplus B$ or $\mathbb{R}^3 = \mathbb{R}^2 \oplus \mathbb{R}^1$.

On the other hand, if we try to write $\mathbb{R}^3$ as the direct sum of the xy-plane A with say, the yz-plane B, then the intersection condition is violated since $A \cap B$ is just the entire y-axis. In this case, any vector lying on the y-axis can be specified in terms of its components in either the xy-plane or in the yz-plane. //

In many of our later applications we shall need to take the direct sum of several vector spaces. While it should be obvious that this follows simply by induction from the above case, we go through the details nevertheless. We say that a vector space V is the **direct sum** of the subspaces $W_1, \dots, W_r$ if the following properties are true:

(a) $W_i \neq \{0\}$ for each $i = 1, \dots, r$;
(b) $W_i \cap (W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r) = \{0\}$ for $i = 1, \dots, r$;
(c) $V = W_1 + \cdots + W_r$.

If V is the direct sum of the W_i, then we write $V = W_1 \oplus \cdots \oplus W_r$. The generalization of Theorem 2.12 is the following.

Theorem 2.14 If $W_1, \dots, W_r$ are subspaces of V, then

$$V = W_1 \oplus \cdots \oplus W_r$$

if and only if every $v \in V$ has a unique representation of the form

$$v = v_1 + \cdots + v_r$$

where $v_i \in W_i$ for each $i = 1, \ldots, r$.

Proof First assume that V is the direct sum of $W_1, \ldots, W_r$. Given any $v \in V$, part (c) in the definition of direct sum tells us that we have

$$v = v_1 + \cdots + v_r$$

where $v_i \in W_i$ for each $i = 1, \ldots, r$. If we also have another representation

$$v = v'_1 + \cdots + v'_r$$

with $v'_i \in W_i$, then

$$v_1 + \cdots + v_r = v'_1 + \cdots + v'_r$$

so that for any $i = 1, \ldots, r$ we have

$$v'_i - v_i = (v_1 - v'_1) + \cdots + (v_{i-1} - v'_{i-1}) + (v_{i+1} - v'_{i+1}) + \cdots + (v_r - v'_r).$$

Since $v'_i - v_i \in W_i$ and the right hand side of this equation is an element of $W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r$, we see that part (b) of the definition requires that $v'_i - v_i = 0$, and hence $v'_i = v_i$. This proves the uniqueness of the representation.

Conversely, assume that each $v \in V$ has a unique representation of the form $v = v_1 + \cdots + v_r$ where $v_i \in W_i$ for each $i = 1, \ldots, r$. Since part (c) of the definition of direct sum is automatically satisfied, we must show that part (b) is also satisfied. Suppose

$$v_1 \in W_1 \cap (W_2 + \cdots + W_r) .$$

Since

$$v_1 \in W_2 + \cdots + W_r$$

we must also have

$$v_1 = v_2 + \cdots + v_r$$

for some $v_2 \in W_2, \ldots, v_r \in W_r$. But then

$$0 = -v_1 + v_2 + \cdots + v_r$$

and

$$0 = 0 + \cdots + 0$$

are two representations of the vector 0, and hence the uniqueness of the representations implies that $v_i = 0$ for each $i = 1, \ldots, r$. In particular, the case $i = 1$ means that

$$W_1 \cap (W_2 + \cdots + W_r) = \{0\} \;.$$

A similar argument applies to $W_i \cap (W_2 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r)$ for any $i = 1, \ldots, r$. This proves part (b) in the definition of direct sum. ∎

If $V = W_1 \oplus \cdots \oplus W_r$, then it seems reasonable that we should be able to form a basis for V by adding up the bases of the subspaces W_i . This is indeed the case as we now show.

Theorem 2.15 Let $W_1, \ldots, W_r$ be subspaces of V, and for each $i = 1, \ldots, r$ let W_i have basis $B_i = \{w_{i1}, \ldots, w_{in_i}\}$. Then V is the direct sum of the W_i if and only if the union of bases

$$B = \cup_{i=1}^r B_i = \{w_{11}, \ldots, w_{1n_1}, \ldots, w_{r1}, \ldots, w_{rn_r}\}$$

is a basis for V.

Proof Suppose that B is a basis for V. Then for any $v \in V$ we may write

$$\begin{aligned} v &= (a_{11}w_{11} + \cdots + a_{1n_1}w_{1n_1}) + \cdots + (a_{r1}w_{r1} + \cdots + a_{rn_r}w_{rn_r}) \\ &= w_1 + \cdots + w_r \end{aligned}$$

where

$$w_i = a_{i1}w_{i1} + \cdots + a_{in_i}w_{in_i} \in W_i$$

and $a_{ij} \in \mathcal{F}$. Now let

$$v = w'_1 + \cdots + w'_r$$

be any other expansion of v, where each $w'_i \in W_i$. Using the fact that B_i is a basis for W_i we have

$$w'_i = b_{i1}w_{i1} + \cdots + b_{in_i}w_{in_i}$$

for some set of scalars b_{ij}. This means that we may also write

$$v = (b_{11}w_{11} + \cdots + b_{1n_1}w_{1n_1}) + \cdots + (b_{r1}w_{r1} + \cdots + b_{rn_r}w_{rn_r}) \;.$$

However, since B is a basis for V, we may equate the coefficients of w_{ij} in these two expressions for v to obtain $a_{ij} = b_{ij}$ for all i, j. We have thus proved that the representation of v is unique, and hence Theorem 2.14 tells us that V is the direct sum of the W_i.

Now suppose that V is the direct sum of the W_i. This means that any $v \in V$ may be expressed in the unique form $v = w_1 + \cdots + w_r$ where $w_i \in W_i$ for each I = 1, . . . , r. Given that $B_i = \{w_{i\,1}, \ldots, w_{i\,n_i}\}$ is a basis for W_i, we must show that $B = \cup B_i$ is a basis for V. We first note that each $w_i \in W_i$ may be expanded in terms of the members of B_i , and therefore $\cup B_i$ clearly spans V. It remains to show that the elements of B are linearly independent. We first write

$$(c_{11}w_{11} + \cdots + c_{1n_1}w_{1n_1}) + \cdots + (c_{r\,1}w_{r\,1} + \cdots + c_{r\,n_r}w_{r\,n_r}) = 0$$

and note that

$$c_{i\,1}w_{i\,1} + \cdots + c_{in_i}w_{in_i} \in W_i\ .$$

Using the fact that $0 + \cdots + 0 = 0$ (where each $0 \in W_i$) along with the uniqueness of the representation in any direct sum, we see that for each i = 1, . . . , r we must have

$$c_{i1}w_{i1} + \cdots + c_{in_i}w_{in_i} = 0\ .$$

However, since B_i is a basis for W_i, this means that $c_{i\,j} = 0$ for every i and j, and hence the elements of $B = \cup B_i$ are linearly independent. ∎

Corollary If $V = W_1 \oplus \cdots \oplus W_r$, then

$$\dim V = \sum_{i=1}^{r} \dim W_i.$$

Proof Obvious from Theorem 2.15. This also follows by induction from Theorem 2.11. ∎

Exercises

1. Let W_1 and W_2 be subspaces of $\mathbb{R}^3$ defined by $W_1 = \{(x, y, z): x = y = z\}$ and $W_2 = \{(x, y, z): x = 0\}$. Show that $\mathbb{R}^3 = W_1 \oplus W_2$.

2. Let W_1 be any subspace of a finite-dimensional space V. Prove that there exists a subspace W_2 of V such that $V = W_1 \oplus W_2$.

3. Let W_1, W_2 and W_3 be subspaces of a vector space V. Show that

$$(W_1 \cap W_2) + (W_1 \cap W_3) \subset W_1 \cap (W_2 + W_3)\ .$$

Give an example in $V = \mathbb{R}^2$ for which equality does not hold.

4. Let $V = F[\mathbb{R}]$ be as in Exercise 2.1.6. Let W_+ and W_- be the subsets of V defined by $W_+ = \{f \in V: f(-x) = f(x)\}$ and $W_- = \{f \in V: f(-x) = -f(x)\}$. In other words, W_+ is the subset of all even functions, and W_- is the subset of all odd functions.
   (a) Show that W_+ and W_- are subspaces of V.
   (b) Show that $V = W_+ \oplus W_-$.

5. Let W_1 and W_2 be subspaces of a vector space V.
   (a) Show that $W_1 \subset W_1 + W_2$ and $W_2 \subset W_1 + W_2$.
   (b) Prove that $W_1 + W_2$ is the smallest subspace of V that contains both W_1 and W_2. In other words, if $\mathcal{S}(W_1, W_2)$ denotes the linear span of W_1 and W_2, show that $W_1 + W_2 = \mathcal{S}(W_1, W_2)$. [*Hint*: Show that $W_1 + W_2 \subset \mathcal{S}(W_1, W_2)$ and $\mathcal{S}(W_1, W_2) \subset W_1 + W_2$.]

6. Let V be a finite-dimensional vector space. For any $x \in V$, we define $\mathcal{F}x = \{ax: a \in \mathcal{F}\}$. Prove that $\{x_1, x_2, \ldots, x_n\}$ is a basis for V if and only if $V = \mathcal{F}x_1 \oplus \mathcal{F}x_2 \oplus \cdots \oplus \mathcal{F}x_n$.

7. If A and B are vector spaces, show that A + B is the span of $A \cup B$.

2.4 INNER PRODUCT SPACES

Before proceeding with the general theory of inner products, let us briefly review what the reader should already know from more elementary courses. It is assumed that the reader is familiar with vectors in $\mathbb{R}^3$, and we show that for any $\vec{a}, \vec{b} \in \mathbb{R}^3$ the **scalar product** (also called the "**dot product**") $\vec{a} \bullet \vec{b}$ may be written as either

$$\vec{a} \bullet \vec{b} = \sum_{i=1}^{3} a_i b_i$$

where $\{a_i\}$ and $\{b_i\}$ are the coordinates of $\vec{a}$ and $\vec{b}$ relative to the standard basis for $\mathbb{R}^3$ (see Example 2.7), or as

$$\vec{a} \bullet \vec{b} = \|\vec{a}\|\,\|\vec{b}\| \cos\theta$$

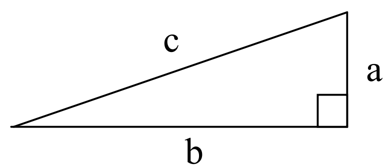
where $\theta = \angle(\vec{a}, \vec{b})$ and

$$\|\vec{a}\|^2 = \sum_{i=1}^{3} a_i^2$$

with a similar equation for $\|\vec{b}\|$. The symbol $\|\;\|$ is just the vector space generalization of the absolute value of numbers, and will be defined carefully below (see Example 2.9). For now, just think of $\|\vec{a}\|$ as meaning the length of the vector $\vec{a}$ in $\mathbb{R}^3$.

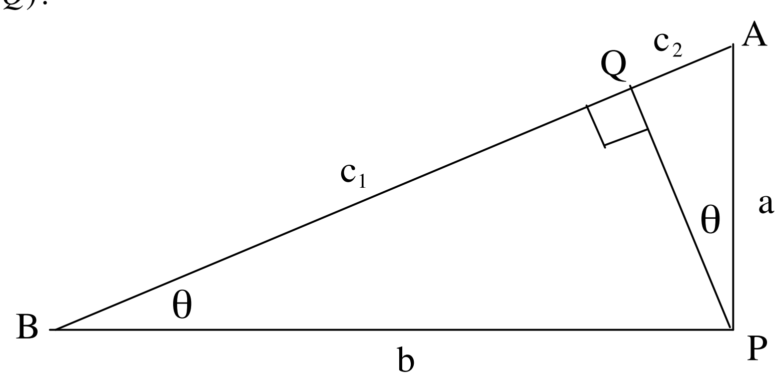
Just for fun, for the sake of completeness, and to show exactly what these equations depend on, we prove this as a series of simple lemmas. Our first lemma is known as the Pythagorean theorem.

Lemma 2.1 Given a right triangle with sides a, b, and c as shown,



we have $c^2 = a^2 + b^2$.

Proof Draw the line PQ perpendicular to the hypotenuse c = AB. Note that we can now write c as the sum of the two parts c_1 and c_2. First observe that the triangle ABP is similar to triangle APQ because they are both right triangles and they have the angle at A in common (so they must have their third angle the same). If we let this third angle be $\theta = \angle(ABP)$, then we also have $\theta = \angle(APQ)$.



Note that the three triangles ABP, APQ and PBQ are all similar, and hence we have (remember $c = c_1 + c_2$)

$$\frac{c_1}{b} = \frac{b}{c} \quad \text{and} \quad \frac{c_2}{a} = \frac{a}{c}\,.$$

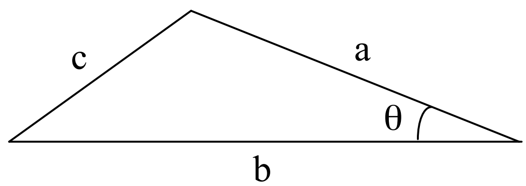
Therefore

$$c = c_1 + c_2 = \frac{a^2 + b^2}{c}$$

from which the lemma follows immediately. ∎

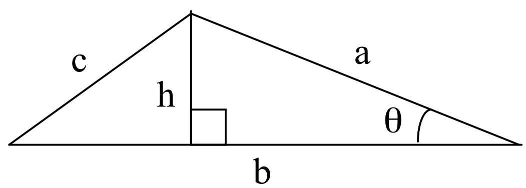
Our next lemma is known as the law of cosines. This law, together with Lemma 2.1, shows that for any triangle T with sides $a \le b \le c$, it is true that $a^2 + b^2 = c^2$ if and only if T is a right triangle.

Lemma 2.2 For any triangle as shown,



we have $c^2 = a^2 + b^2 - 2ab \cos \theta$.

Proof Draw a perpendicular to side b as shown:



By the Pythagorean theorem we have

$$\begin{aligned} c^2 &= h^2 + (b - a\cos\theta)^2 \\ &= (a\sin\theta)^2 + (b - a\cos\theta)^2 \\ &= a^2\sin^2\theta + b^2 - 2ab\cos\theta + a^2\cos^2\theta \\ &= a^2 + b^2 - 2ab\cos\theta \end{aligned}$$

where we used $\sin^2\theta + \cos^2\theta = 1$ which follows directly from Lemma 2.1 with $a = c(\sin\theta)$ and $b = c(\cos\theta)$. ∎

We now *define* the scalar product $\vec{a} \cdot \vec{b}$ for any $\vec{a}, \vec{b} \in \mathbb{R}^3$ by

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{3} a_i b_i = \vec{b} \cdot \vec{a}$$

where $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$. It is easy to see that

$$\vec{a} \bullet (\vec{b} + \vec{c}) = \sum_{i=1}^{3} a_i(b_i + c_i) = \sum_{i=1}^{3} (a_ib_i + a_ic_i) = \vec{a} \bullet \vec{b} + \vec{a} \bullet \vec{c}$$

and similarly, it is easy to show that

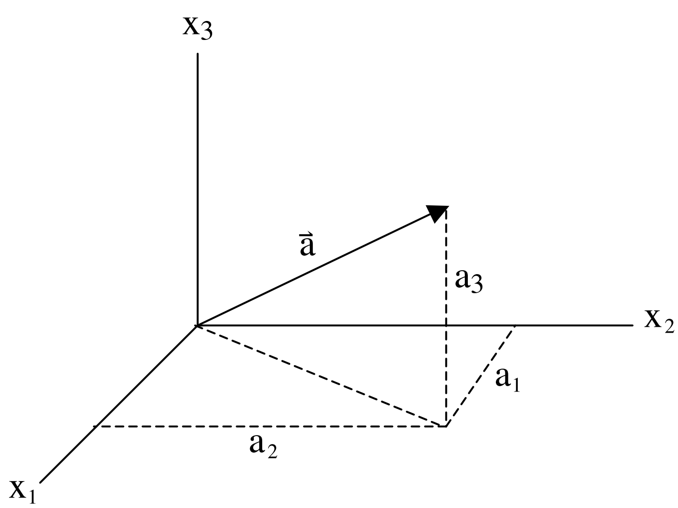
$$(\vec{a} + \vec{b}) \bullet \vec{c} = \vec{a} \bullet \vec{c} + \vec{b} \bullet \vec{c}$$

and

$$(k\vec{a}) \bullet \vec{b} = k(\vec{a} \bullet \vec{b})$$

where $k \in \mathbb{R}$. From the figure below, we see that the Pythagorean theorem also shows us that

$$\| \vec{a} \|^2 = \sum_{i=1}^{3} a_ia_i = \vec{a} \bullet \vec{a} \ .$$



This is the justification for writing $\| \vec{a} \|$ to mean the length of the vector $\vec{a} \in \mathbb{R}^3$.

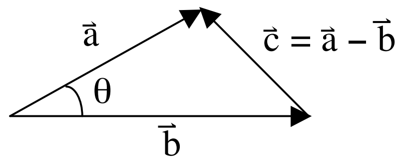
Noting that any two vectors (with a common origin) in $\mathbb{R}^3$ lie in a plane, we have the following well-known formula for the dot product.

Lemma 2.3 For any $\vec{a}, \vec{b} \in \mathbb{R}^3$ we have

$$\vec{a} \bullet \vec{b} = ab \cos \theta$$

where $a = \| \vec{a} \|$, $b = \| \vec{b} \|$ and $\theta = \angle(\vec{a}, \vec{b})$.

Proof Draw the vectors $\vec{a}$ and $\vec{b}$ along with their difference $\vec{c} = \vec{a} - \vec{b}$:



By the law of cosines we have $c^2 = a^2 + b^2 - 2ab \cos\theta$, while on the other hand

$$c^2 = \|\vec{a} - \vec{b}\|^2 = (\vec{a} - \vec{b}) \bullet (\vec{a} - \vec{b}) = a^2 + b^2 - 2\,\vec{a} \bullet \vec{b}\ .$$

Therefore we see that $\vec{a} \bullet \vec{b} = ab \cos\theta$. ∎

The main reason that we went through all of this is to motivate the generalization to arbitrary vector spaces. For example, if $u, v \in \mathbb{R}^n$, then to say that

$$u \bullet v = \sum_{i=1}^{n} u_i v_i$$

makes sense, whereas to say that $u \bullet v = \|u\|\,\|v\| \cos\theta$ leaves one wondering just what the "angle" θ means in higher dimensions. In fact, this will be used to *define* the angle θ.

We now proceed to define a general scalar (or inner) product $\langle u, v\rangle$ of vectors $u, v \in V$. Throughout this section, we let V be a vector space over either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. By way of motivation, we will want the inner product $\langle\ ,\ \rangle$ applied to a single vector $v \in V$ to yield the length (or norm) of v, so that $\|v\|^2 = \langle v, v\rangle$. But $\|v\|$ must be a real number even if the field we are working with is $\mathbb{C}$. Noting that for any complex number $z \in \mathbb{C}$ we have $|z|^2 = zz^*$, we are led to make the following definition.

Let V be a vector space over $\mathcal{F}$ (where $\mathcal{F}$ is either $\mathbb{R}$ or $\mathbb{C}$). By an **inner product** on V (sometimes called the **Hermitian inner product**), we mean a mapping $\langle\ ,\ \rangle\colon V \times V \to \mathcal{F}$ such that for all $u, v, w \in V$ and $a, b \in \mathcal{F}$ we have

(IP1) $\langle au + bv, w\rangle = a^*\langle u, w\rangle + b^*\langle v, w\rangle$;
(IP2) $\langle u, v\rangle = \langle v, u\rangle^*$;
(IP3) $\langle u, u\rangle \ge 0$ and $\langle u, u\rangle = 0$ if and only if $u = 0$.

Using these properties, we also see that

$$\begin{aligned}\langle u, av + bw\rangle &= \langle av + bw, u\rangle^* \\ &= (a^*\langle v, u\rangle + b^*\langle w, u\rangle)^* \\ &= a\langle u, v\rangle + b\langle u, w\rangle\end{aligned}$$

and hence, for the sake of reference, we call this

(IP1′) $\langle u, av + bw\rangle = a\langle u, v\rangle + b\langle u, w\rangle$.

(The reader should be aware that instead of $\langle au, v\rangle = a^*\langle u, v\rangle$, many authors define $\langle au, v\rangle = a\langle u, v\rangle$ and $\langle u, av\rangle = a^*\langle u, v\rangle$. This is particularly true in mathematics texts, whereas we have chosen the convention used by most physics texts. Of course, this has no effect on any of our results.)

A space V together with an inner product is called an **inner product space**. If V is an inner product space over the field $\mathbb{C}$, then V is called a **complex** inner product space, and if the field is $\mathbb{R}$, then V is called a **real** inner product space. A complex inner product space is frequently called a **unitary space**, and a real inner product space is frequently called a **Euclidean space**. Note that in the case of a real space, the complex conjugates in (IP1) and (IP2) are superfluous.

By (IP2) we have $\langle u, u\rangle \in \mathbb{R}$ so that we may define the **length** (or **norm**) of u to be the nonnegative real number

$$\|u\| = \langle u, u\rangle^{1/2} .$$

If $\|u\| = 1$, then u is said to be a **unit vector**. If $\|v\| \neq 0$, then we can normalize v by setting $u = v/\|v\|$. One sometimes writes $\hat{v}$ to mean the unit vector in the direction of v, i.e., $v = \|v\|\,\hat{v}$.

Example 2.9 Let $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ be vectors in $\mathbb{C}^n$. We define

$$\langle X, Y\rangle = \sum_{i=1}^{n} x_i{}^* y_i$$

and leave it to the reader to show that this satisfies (IP1) – (IP3). In the case of the space $\mathbb{R}^n$, we have $\langle X, Y\rangle = X\bullet Y = \sum x_i y_i$. This inner product is called the **standard inner product** in $\mathbb{C}^n$ (or $\mathbb{R}^n$).

We also see that if $X, Y \in \mathbb{R}^n$ then

$$\|X - Y\|^2 = \langle X - Y, X - Y\rangle = \sum_{i=1}^{n} (x_i - y_i)^2 .$$

Thus $\|X - Y\|$ is indeed just the distance between the points $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ that we would expect by applying the Pythagorean theorem to points in $\mathbb{R}^n$. In particular, $\|X\|$ is simply the length of the vector X. ∕∕

It is now easy to see why we defined the inner product as we did. For example, consider simply the space $\mathbb{C}^3$. Then with respect to the standard inner product on $\mathbb{C}^3$, the vector $X = (1, i, 0)$ will have norm $\|X\|^2 = \langle X, X\rangle = 1 + 1 + 0 = 2$, while if we had used the expression corresponding to the standard inner product on $\mathbb{R}^3$, we would have found $\|X\|^2 = 1 - 1 + 0 = 0$ even though $X \neq 0$.

Example 2.10 Let V be the vector space of continuous complex-valued functions defined on the real interval [a, b]. We may define an inner product on V by

$$\langle f, g\rangle = \int_a^b f^*(x)g(x)dx$$

for all $f, g \in V$. It should be obvious that this satisfies the three required properties of an inner product. ∕∕

We now prove the generalization of Theorem 0.7, an important result known as the **Cauchy-Schwartz inequality**.

Theorem 2.16 Let V be an inner product space. Then for any $u, v \in V$ we have

$$|\langle u, v\rangle| \le \|u\|\,\|v\| \;.$$

Proof If either u or v is zero the theorem is trivially true. We therefore assume that $u \neq 0$ and $v \neq 0$. Then, for any real number c, we have (using (IP2) and the fact that $|z|^2 = zz^*$)

$$\begin{aligned} 0 &\le \|v - c\langle u, v\rangle u\|^2 \\ &= \langle v - c\langle u, v\rangle u,\; v - c\langle u, v\rangle u\rangle \\ &= \langle v, v\rangle - c\langle u, v\rangle\langle v, u\rangle - c\langle u, v\rangle^*\langle u, v\rangle + c^2\langle u, v\rangle^*\langle u, v\rangle\langle u, u\rangle \\ &= \|v\|^2 - 2c|\langle u, v\rangle|^2 + c^2|\langle u, v\rangle|^2\|u\|^2 . \end{aligned}$$

Now let $c = 1/\|u\|^2$ to obtain

$$0 \le \|v\|^2 - |\langle u, v\rangle|^2/\|u\|^2$$

or

$$|\langle u, v\rangle|^2 \le \|u\|^2\,\|v\|^2 \;.$$

Taking the square root proves the theorem. ∎

Theorem 2.17 The norm in an inner product space V has the following properties for all u, v ∈ V and k ∈ $\mathcal{F}$:

(N1) $\|u\| \geq 0$ and $\|u\| = 0$ if and only if u = 0.
(N2) $\|ku\| = |k|\,\|u\|$.
(N3) $\|u + v\| \leq \|u\| + \|v\|$.

Proof Since $\|u\| = \langle u, u\rangle^{1/2}$, (N1) follows from (IP3). Next, we see that

$$\|ku\|^2 = \langle ku, ku\rangle = |k|^2\,\|u\|^2$$

and hence taking the square root yields (N2). Finally, using Theorem 2.16 and the fact that $z + z^* = 2\,\mathrm{Re}\,z \leq 2|z|$ for any $z \in \mathbb{C}$, we have

$$\begin{aligned}\|u+v\|^2 &= \langle u+v,\, u+v\rangle \\ &= \langle u,\, u\rangle + \langle u,\, v\rangle + \langle v,\, u\rangle + \langle v,\, v\rangle \\ &= \|u\|^2 + \langle u,\, v\rangle + \langle u,\, v\rangle^* + \|v\|^2 \\ &\leq \|u\|^2 + 2|\langle u,\, v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2 .\end{aligned}$$

Taking the square root yields (N3). ∎

We note that property (N3) is frequently called the **triangle inequality** because in two or three dimensions, it simply says that the sum of two sides of a triangle is greater than the third. Furthermore, we remark that properties (N1) – (N3) may be used to *define* a normed vector space. In other words, a **normed vector space** is defined to be a vector space V together with a mapping $\|\ \| : V \rightarrow \mathbb{R}$ that obeys properties (N1) – (N3). While a normed space V does not in general have an inner product defined on it, the existence of an inner product leads in a natural way (i.e., by Theorem 2.17) to the existence of a norm on V.

Example 2.11 Let us prove a simple but useful result dealing with the norm in any normed space V. From the properties of the norm, we see that for any u, v ∈ V we have

$$\|u\| = \|u - v + v\| \leq \|u - v\| + \|v\|$$

and

$$\|v\| = \|v - u + u\| \leq \|u - v\| + \|u\|.$$

Rearranging each of these yields

$$\|u\| - \|v\| \le \|u - v\|$$

and

$$\|v\| - \|u\| \le \|u - v\|.$$

This shows that

$$\big|\, \|u\| - \|v\| \,\big| \le \|u - v\|. \quad /\!/$$

Example 2.12 Consider the space V of Example 2.10 and the associated inner product ⟨f, g⟩. Applying Theorem 2.16 we have

$$\left|\int_a^b f^*(x)g(x)dx\right| \le \left\{\int_a^b |f(x)|^2\, dx\right\}^{1/2} \left\{\int_a^b |g(x)|^2\, dx\right\}^{1/2}$$

and applying Theorem 2.17 we see that

$$\left\{\int_a^b |f(x)+g(x)|^2\, dx\right\}^{1/2} \le \left\{\int_a^b |f(x)|^2\, dx\right\}^{1/2} + \left\{\int_a^b |g(x)|^2\, dx\right\}^{1/2}.$$

The reader might try and prove either of these directly from the definition of the integral if he or she wants to gain an appreciation of the power of the axiomatic approach to inner products. ∕∕

Finally, let us finish our generalization of Lemmas 2.1 – 2.3. If we repeat the proof of Lemma 2.3 using the inner product and norm notations, we find that for any u, v ∈ $\mathbb{R}^3$ we have ⟨u, v⟩ = ‖u‖ ‖v‖ cos θ. Now let V be any real vector space. We define the **angle** θ between two nonzero vectors u, v ∈ V by

$$\cos\theta = \frac{\langle u,\, v\rangle}{\|u\|\|v\|}.$$

Note that |cos θ| ≤ 1 by Theorem 2.16 so that this definition makes sense. We say that u is **orthogonal** (or **perpendicular**) to v if ⟨u, v⟩ = 0. If u and v are orthogonal, we often write this as u ⊥ v. From the basic properties of the inner product, it then follows that ⟨v, u⟩ = ⟨u, v⟩* = 0* = 0 so that v is orthogonal to u also. Thus u ⊥ v if and only if cos θ = 0. While cos θ is only defined in a real vector space, our definition of orthogonality is valid in any space V over $\mathcal{F}$.

Exercises

1. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be vectors in $\mathbb{R}^2$, and define the mapping $\langle\ ,\ \rangle: \mathbb{R}^2 \rightarrow \mathbb{R}$ by $\langle x, y\rangle = x_1y_1 - x_1y_2 - x_2y_1 + 3x_2y_2$. Show that this defines an inner product on $\mathbb{R}^2$.

2. Let $x = (3, 4) \in \mathbb{R}^2$, and evaluate $\|x\|$ with respect to the norm induced by:
   (a) The standard inner product on $\mathbb{R}^2$.
   (b) The inner product defined in the previous exercise.

3. Let V be an inner product space, and let $x, y \in V$.
   (a) Prove the **parallelogram law**:

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \ .$$

   (The geometric meaning of this equation is that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of the sides.)
   (b) Prove the **Pythagorean theorem**:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \quad \text{if } x \perp y \ .$$

4. Find a unit vector orthogonal to the vectors $x = (1, 1, 2)$ and $y = (0, 1, 3)$ in $\mathbb{R}^3$.

5. Let $u = (z_1, z_2)$ and $v = (w_1, w_2)$ be vectors in $\mathbb{C}^2$, and define the mapping $\langle\ ,\ \rangle: \mathbb{C}^2 \rightarrow \mathbb{R}$ by

$$\langle u, v\rangle = z_1w_1{}^* + (1 + i)z_1w_2{}^* + (1 - i)z_2w_1{}^* + 3z_2w_2{}^* \ .$$

   Show that this defines an inner product on $\mathbb{C}^2$.

6. Let $u = (1 - 2i, 2 + 3i) \in \mathbb{C}^2$ and evaluate $\|u\|$ with respect to the norm induced by:
   (a) The standard norm on $\mathbb{C}^2$.
   (b) The inner product defined in the previous exercise.

7. Let V be an inner product space. Verify the following polar form identities:
   (a) If V is a real space and $x, y \in V$, then

$$\langle x, y\rangle = (1/4)(\ \|x + y\|^2 - \|x - y\|^2)\ .$$

(b) If V is a complex space and x, y ∈ V, then

$$\langle x, y\rangle = (1/4)(\ \|x + y\|^2 - \|x - y\|^2) + (i/4)(\ \|i\, x + y\|^2 - \|i x - y\|^2)$$

(If we were using instead the inner product defined by $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$, then the last two terms in this equation would read $\|x \pm i y\|$.)

8. Let V = C[0, 1] be the space of continuous real-valued functions defined on the interval [0, 1]. Define an inner product on C[0, 1] by

$$\langle f, g\rangle = \int_0^1 f(t)g(t)dt.$$

(a) Verify that this does indeed define an inner product on V.

(b) Evaluate $\|f\|$ where $f = t^2 - 2t + 3 \in V$.

9. Given a vector space V, we define a mapping d: V × V → ℝ by d(x, y) = $\|x - y\|$ for all x, y ∈ V. Show that:

(a) d(x, y) ≥ 0 and d(x, y) = 0 if and only if x = y.

(b) d(x, y) = d(y, x).

(c) d(x, z) ≤ d(x, y) + d(y, z) (triangle inequality).

The number d(x, y) is called the **distance** from x to y, and the mapping d is called a **metric** on V. Any arbitrary set S on which we have defined a function d: S × S → ℝ satisfying these three properties is called a **metric space**.

10. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for a complex space V, and let x ∈ V be arbitrary. Show that

(a) $x = \sum_{i=1}^{n} e_i \langle e_i, x\rangle$.

(b) $\|x\|^2 = \sum_{i=1}^{n} |\langle e_i, x\rangle|^2$.

11. Show equality holds in the Cauchy-Schwartz inequality if and only if one vector is proportional to the other.

2.5 ORTHOGONAL SETS

If a vector space V is equipped with an inner product, then we may define a subspace of V that will turn out to be extremely useful in a wide variety of applications. Let W be any subset of such a vector space V. (Note that W need not be a subspace of V.) We define the **orthogonal compliment** of W to be the set $W^\perp$ given by

$$W^\perp = \{v \in V: \langle v, w\rangle = 0 \text{ for all } w \in W\} \ .$$

Theorem 2.18 Let W be any subset of a vector space V. Then $W^\perp$ is a subspace of V.

Proof We first note that $0 \in W^\perp$ since for any $v \in V$ we have

$$\langle 0, v\rangle = \langle 0v, v\rangle = 0\langle v, v\rangle = 0 \ .$$

To finish the proof, we simply note that for any $u, v \in W^\perp$, for any scalars a, $b \in \mathcal{F}$, and for every $w \in W$ we have

$$\langle au + bv, w\rangle = a^*\langle u, w\rangle + b^*\langle v, w\rangle = a^*0 + b^*0 = 0$$

so that $au + bv \in W^\perp$. ∎

Consider the space $\mathbb{R}^3$ with the usual Cartesian coordinate system (x, y, z). If we let $W = \mathbb{R}^2$ be the xy-plane, then $W^\perp = \mathbb{R}^1$ is just the z-axis since the standard inner product on $\mathbb{R}^3$ shows that any $v \in \mathbb{R}^3$ of the form (0, 0, c) is orthogonal to any $w \in \mathbb{R}^3$ of the form (a, b, 0). Thus, in this case anyway, we see that $W \oplus W^\perp = \mathbb{R}^3$. We will shortly prove that $W \oplus W^\perp = V$ for any inner product space V and subspace $W \subset V$. Before we can do this however, we must first discuss orthonormal sets of vectors.

A set $\{v_i\}$ of nonzero vectors in a space V is said to be an **orthogonal set** (or to be **mutually orthogonal**) if $\langle v_i, v_j\rangle = 0$ for $i \neq j$. If in addition, each v_i is a unit vector, then the set $\{v_i\}$ is said to be an **orthonormal set** and we write

$$\langle v_i, v_j \rangle = \delta_{ij}$$

where the very useful symbol δ_{ij} (called the **Kronecker delta**) is defined by

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

Theorem 2.19 Any orthonormal set of vectors $\{v_i\}$ is linearly independent.

Proof If $\sum a_i v_i = 0$ for some set of scalars $\{a_i\}$, then

$$0 = \langle v_j, 0\rangle = \langle v_j, \Sigma a_i v_i\rangle = \Sigma_i a_i\langle v_j, v_i\rangle = \Sigma_i a_i\delta_{ij} = a_j$$

so that $a_j = 0$ for each j, and hence $\{v_i\}$ is linearly independent. ∎

Note that in the proof of Theorem 2.19 it was not really necessary that each v_i be a unit vector. Any orthogonal set would work just as well.

Theorem 2.20 If $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal set in V and if $w \in V$ is arbitrary, then the vector

$$u = w - \Sigma_i \langle v_i, w\rangle v_i$$

is orthogonal to each of the v_i.

Proof We simply compute $\langle v_j, u\rangle$:

$$\begin{aligned}\langle v_j, u\rangle &= \langle v_j, w - \Sigma_i \langle v_i, w\rangle v_i\rangle \\ &= \langle v_j, w\rangle - \Sigma_i \langle v_i, w\rangle\langle v_j, v_i\rangle \\ &= \langle v_j, w\rangle - \Sigma_i \langle v_i, w\rangle\delta_{ij} \\ &= \langle v_j, w\rangle - \langle v_j, w\rangle = 0 \ . \end{aligned}$$ ∎

The numbers $c_i = \langle v_i, w\rangle$ are frequently called the **Fourier coefficients** of w with respect to v_i. In fact, we leave it as an exercise for the reader to show that the expression $\|w - \Sigma_i a_i v_i\|$ achieves its minimum precisely when $a_i = c_i$ (see Exercise 2.5.4). Furthermore, we also leave it to the reader (see Exercise 2.5.5) to show that

$$\sum_{i=1}^{n} |c_i|^2 \le \|w\|^2$$

which is called **Bessel's inequality**.

As we remarked earlier, most mathematics texts write $\langle u, av\rangle = a^*\langle u, v\rangle$ rather than $\langle u, av\rangle = a\langle u, v\rangle$. In this case, Theorem 2.20 would be changed to read that the vector

$$u = w - \Sigma_i \langle w, v_i\rangle v_i$$

is orthogonal to each v_j.

Example 2.13 The simplest and best known example of an orthonormal set is the set $\{e_i\}$ of standard basis vectors in $\mathbb{R}^n$. Thus

$$\begin{aligned} e_1 &= (1, 0, 0, \dots, 0) \\ e_2 &= (0, 1, 0, \dots, 0) \\ &\vdots \\ e_n &= (0, 0, 0, \dots, 1) \end{aligned}$$

and clearly

$$\langle e_i, e_j \rangle = e_i \bullet e_j = \delta_{ij}$$

since for any $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, we have

$$\langle X, Y \rangle = X \bullet Y = \sum_{i=1}^{n} x_i y_i \ .$$

(It would perhaps be better to write the unit vectors as $\hat{e}_i$ rather than e_i, but this will generally not cause any confusion.) ∕∕

Example 2.14 Let V be the space of continuous complex-valued functions defined on the real interval $[-\pi, \pi]$. As in Example 2.10, we define

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f^*(x) g(x)\, dx$$

for all $f, g \in V$. We show that the set of functions

$$f_n = \left(\frac{1}{2\pi}\right)^{1/2} e^{inx}$$

for $n = 1, 2, \dots$ forms an orthonormal set.

If $m = n$, then

$$\langle f_m, f_n \rangle = \langle f_n, f_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} e^{inx}\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx = 1 \ .$$

If $m \neq n$, then we have

$$\begin{aligned} \langle f_m, f_n \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-imx} e^{inx}\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)x}\, dx \\ &= \frac{1}{2\pi} \frac{e^{i(n-m)x}}{i(n-m)} \bigg|_{-\pi}^{\pi} \\ &= \frac{\sin(n-m)\pi}{\pi(n-m)} = 0 \end{aligned}$$

since $\sin n\pi = 0$ for any integer n. (Note that we also used the fact that

$$\sin\theta = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right)$$

which follows from the Euler formula mentioned in Chapter 0.) Therefore, $\langle f_m, f_n\rangle = \delta_{mn}$. That the set $\{f_n\}$ is orthonormal is of great use in the theory of Fourier series. //

We now wish to show that every finite-dimensional vector space with an inner product has an orthonormal basis. The proof is based on the famous Gram-Schmidt orthogonalization process, the precise statement of which we present as a corollary following the proof.

Theorem 2.21 Let V be a finite-dimensional inner product space. Then there exists an orthonormal set of vectors that forms a basis for V.

Proof Let dim V = n and let $\{u_1, \dots, u_n\}$ be a basis for V. We will construct a new basis $\{w_1, \dots, w_n\}$ such that $\langle w_i, w_j\rangle = \delta_{ij}$. To begin, we choose

$$w_1 = u_1/\|u_1\|$$

so that

$$\begin{aligned}\|w_1\|^2 &= \langle w_1, w_1\rangle = \langle u_1/\|u_1\|, u_1/\|u_1\|\rangle = (1/\|u_1\|^2)\langle u_1, u_1\rangle \\ &= (1/\|u_1\|^2)\|u_1\|^2 = 1\end{aligned}$$

and hence w_1 is a unit vector. We now take u_2 and subtract off its "projection" along w_1. This will leave us with a new vector v_2 that is orthogonal to w_1. Thus, we define

$$v_2 = u_2 - \langle w_1, u_2\rangle w_1$$

so that

$$\langle w_1, v_2\rangle = \langle w_1, u_2\rangle - \langle w_1, u_2\rangle\langle w_1, w_1\rangle = 0$$

(this also follows from Theorem 2.20). If we let

$$w_2 = v_2/\|v_2\|$$

then $\{w_1, w_2\}$ is an orthonormal set (that $v_2 \neq 0$ will be shown below).

We now go to u_3 and subtract off its projection along w_1 and w_2. In other words, we define

$$v_3 = u_3 - \langle w_2, u_3\rangle w_2 - \langle w_1, u_3\rangle w_1$$

so that $\langle w_1, v_3\rangle = \langle w_2, v_3\rangle = 0$. Choosing

$$w_3 = v_3/\|v_3\|$$

we now have an orthonormal set $\{w_1, w_2, w_3\}$.

It is now clear that given an orthonormal set $\{w_1, \ldots, w_k\}$, we let

$$v_{k+1} = u_{k+1} - \sum_{i=1}^{k} \langle w_i, u_{k+1}\rangle w_i$$

so that v_{k+1} is orthogonal to $w_1, \ldots, w_k$ (Theorem 2.20), and hence we define

$$w_{k+1} = v_{k+1}/\|v_{k+1}\| \;.$$

It should now be obvious that we can construct an orthonormal set of n vectors from our original basis of n vectors. To finish the proof, we need only show that $w_1, \ldots, w_n$ are linearly independent.

To see this, note first that since u_1 and u_2 are linearly independent, w_1 and u_2 must also be linearly independent, and hence $v_2 \neq 0$ by definition of linear independence. Thus w_2 exists and $\{w_1, w_2\}$ is linearly independent by Theorem 2.19. Next, $\{w_1, w_2, u_3\}$ is linearly independent since w_1 and w_2 are in the linear span of u_1 and u_2. Hence $v_3 \neq 0$ so that w_3 exists, and Theorem 2.19 again shows that $\{w_1, w_2, w_3\}$ is linearly independent.

In general then, if $\{w_1, \ldots, w_k\}$ is linearly independent, it follows that $\{w_1, \ldots, w_k, u_{k+1}\}$ is also independent since $\{w_1, \ldots, w_k\}$ is in the linear span of $\{u_1, \ldots, u_k\}$. Hence $v_{k+1} \neq 0$ and w_{k+1} exists. Then $\{w_1, \ldots, w_{k+1}\}$ is linearly independent by Theorem 2.19. Thus $\{w_1, \ldots, w_n\}$ forms a basis for V, and $\langle w_i, w_j\rangle = \delta_{ij}$. ∎

Corollary (**Gram-Schmidt process**) Let $\{u_1, \ldots, u_n\}$ be a linearly independent set of vectors in an inner product space V. Then there exists a set of orthonormal vectors $w_1, \ldots, w_n \in V$ such that the linear span of $\{u_1, \ldots, u_k\}$ is equal to the linear span of $\{w_1, \ldots, w_k\}$ for each $k = 1, \ldots, n$.

Proof This corollary follows by a careful inspection of the proof of Theorem 2.21. ∎

We emphasize that the Gram-Schmidt algorithm (the "orthogonalization process" of Theorem 2.21) as such applies to any inner product space, and is not restricted to only finite-dimensional spaces (see Chapter 12).

We are now ready to prove our earlier assertion. Note that here we require W to be a subspace of V.

Theorem 2.22 Let W be a subspace of a finite-dimensional inner product space V. Then $V = W \oplus W^\perp$.

Proof By Theorem 2.9, W is finite-dimensional. Therefore, if we choose a basis $\{v_1, \ldots, v_k\}$ for W, it may be extended to a basis $\{v_1, \ldots, v_n\}$ for V (Theorem 2.10). Applying Theorem 2.21 to this basis, we construct a new orthonormal basis $\{u_1, \ldots, u_n\}$ for V where

$$u_r = \sum_{j=1}^{r} a_{rj} v_j$$

for $r = 1, \ldots, n$ and some coefficients a_{rj} (determined by the Gram-Schmidt process). In particular, we see that $u_1, \ldots, u_k$ are all in W, and hence they form an orthonormal basis for W.

Since $\{u_1, \ldots, u_n\}$ are orthonormal, it follows that $u_{k+1}, \ldots, u_n$ are in $W^\perp$ (since $\langle u_i, u_j \rangle = 0$ for all $i \le k$ and any $j = k + 1, \ldots, n$). Therefore, given any $x \in V$ we have

$$x = a_1u_1 + \cdots + a_nu_n$$

where

$$a_1u_1 + \cdots + a_ku_k \in W$$

and

$$a_{k+1}u_{k+1} + \cdots + a_nu_n \in W^\perp .$$

This means that $V = W + W^\perp$, and we must still show that $W \cap W^\perp = \{0\}$. But if $y \in W \cap W^\perp$, then $\langle y, y \rangle = 0$ since $y \in W^\perp$ implies that y is orthogonal to any vector in W, and in particular, $y \in W$. Hence $y = 0$ by (IP3), and it therefore follows that $W \cap W^\perp = \{0\}$. ∎

Corollary If V is finite-dimensional and W is a subspace of V, then $(W^\perp)^\perp = W$.

Proof Given any $w \in W$ we have $\langle w, v \rangle = 0$ for all $v \in W^\perp$. This implies that $w \in (W^\perp)^\perp$ and hence $W \subset (W^\perp)^\perp$. By Theorem 2.22, $V = W \oplus W^\perp$ and hence

$$\dim V = \dim W + \dim W^\perp$$

(Theorem 2.11). But $W^\perp$ is also a subspace of V, and hence $V = W^\perp \oplus (W^\perp)^\perp$ (Theorem 2.22) which implies

$$\dim V = \dim W^\perp + \dim(W^\perp)^\perp .$$

Therefore, comparing these last two equations shows that $\dim W = \dim(W^\perp)^\perp$. This result together with $W \subset (W^\perp)^\perp$ implies that $W = (W^\perp)^\perp$. ∎

Finally, note that if $\{e_i\}$ is an orthonormal basis for V, then any $x \in V$ may be written as $x = \sum_i x_i e_i$ where

$$\langle e_j, x\rangle = \langle e_j, \textstyle\sum_i x_i e_i\rangle = \textstyle\sum_i x_i \langle e_j, e_i\rangle = \textstyle\sum_i x_i \delta_{ij} = x_j \ .$$

Therefore we may write

$$x = \textstyle\sum_i \langle e_i, x\rangle e_i$$

which is a very useful expression.

Example 2.15 Consider the following basis vectors for $\mathbb{R}^3$:

$$u_1 = (3, 0, 4) \quad u_2 = (-1, 0, 7) \quad u_3 = (2, 9, 11) \ .$$

Let us apply the Gram-Schmidt process (with the standard inner product on $\mathbb{R}^3$) to obtain a new orthonormal basis for $\mathbb{R}^3$. Since $\|u_1\| = \sqrt{9+16} = 5$, we define

$$w_1 = u_1/5 = (3/5, 0, 4/5) \ .$$

Next, using $\langle w_1, u_2\rangle = -3/5 + 28/5 = 5$ we let

$$v_2 = (-1, 0, 7) - (3, 0, 4) = (-4, 0, 3) \ .$$

Since $\|v_2\| = 5$, we have

$$w_2 = (-4/5, 0, 3/5) \ .$$

Finally, using $\langle w_1, u_3\rangle = 10$ and $\langle w_2, u_3\rangle = 5$ we let

$$v_3 = (2, 9, 11) - (-4, 0, 3) - (6, 0, 8) = (0, 9, 0)$$

and hence, since $\|v_3\| = 9$, our third basis vector becomes

$$w_3 = (0, 1, 0) \ .$$

We leave it to the reader to show that $\{w_1, w_2, w_3\}$ does indeed form an orthonormal basis for $\mathbb{R}^3$. ∕∕

We will have much more to say about inner product spaces after we have treated linear transformations in detail. For the rest of this book, unless explicitly stated otherwise, all vector spaces will be assumed to be finite-dimensional. In addition, the specific scalar field $\mathcal{F}$ will generally not be mentioned, but it is to be understood that all scalars are elements of $\mathcal{F}$.

Exercises

1. Let W be a subset of a vector space V. Prove the following:
   (a) $0^\perp = V$ and $V^\perp = 0$.
   (b) $W \cap W^\perp = \{0\}$.
   (c) $W_1 \subset W_2$ implies $W_2{}^\perp \subset W_1{}^\perp$.

2. Let U and W be subspaces of a finite-dimensional inner product space V. Prove the following:
   (a) $(U + W)^\perp = U^\perp \cap W^\perp$.
   (b) $(U \cap W)^\perp = U^\perp + W^\perp$.

3. Let $\{e_1, \dots, e_n\}$ be an orthonormal basis for an arbitrary inner product space V. If $u = \sum_i u_i e_i$ and $v = \sum_i v_i e_i$ are any vectors in V, show that
$$\langle u, v \rangle = \sum_{i=1}^{n} u_i{}^* v_i$$
   (this is just the generalization of Example 2.9).

4. Suppose $\{e_1, \dots, e_n\}$ is an orthonormal set in a vector space V, and x is any element of V. Show that the expression
$$\left\| x - \sum_{k=1}^{n} a_k e_k \right\|$$
   achieves its minimum value when each of the scalars a_k is equal to the Fourier coefficient $c_k = \langle e_k, x \rangle$. [*Hint*: Using Theorem 2.20 and the Pythagorean theorem (see Exercise 2.4.3), add and subtract the term $\sum_{k=1}^{n} c_k e_k$ in the above expression to conclude that
$$\left\| x - \sum_{k=1}^{n} c_k e_k \right\|^2 \le \left\| x - \sum_{k=1}^{n} a_k e_k \right\|^2$$
   for any set of scalars a_k.]

5. Let $\{e_1, \dots, e_n\}$ be an orthonormal set in an inner product space V, and let $c_k = \langle e_k, x \rangle$ be the Fourier coefficient of $x \in V$ with respect to e_k. Prove **Bessel's inequality**:
$$\sum_{k=1}^{n} |c_k|^2 \le \|x\|^2 \ .$$

[*Hint*: Use the definition of the norm along with the obvious fact that $0 \le \|x - \sum_{k=1}^{n} c_k e_k\|^2$.]

6. Find an orthonormal basis (relative to the standard inner product) for the following subspaces:
   (a) The subspace W of $\mathbb{C}^3$ spanned by the vectors $u_1 = (1, i, 0)$ and $u_2 = (1, 2, 1 - i)$.
   (b) The subspace W of $\mathbb{R}^4$ spanned by $u_1 = (1, 1, 0, 0)$, $u_2 = (0, 1, 1, 0)$ and $u_3 = (0, 0, 1, 1)$.

7. Consider the space $\mathbb{R}^3$ with the standard inner product.
   (a) Convert the vectors $u_1 = (1, 0, 1)$, $u_2 = (1, 0, -1)$ and $u_3 = (0, 3, 4)$ to an orthonormal basis $\{e_1, e_2, e_3\}$ of $\mathbb{R}^3$.
   (b) Write the components of an arbitrary vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ in terms of the basis $\{e_i\}$.

8. Let V = C[0, 1] be the inner product space defined in Exercise 2.4.8. Find an orthonormal basis for V generated by the functions $\{1, x, x^2, x^3\}$.

9. Let V and W be isomorphic inner product spaces under the vector space homomorphism $\phi: V \to W$, and assume that ϕ has the additional property that

$$\|\phi(x_1) - \phi(x_2)\| = \|x_1 - x_2\| \ .$$

   Such a ϕ is called an **isometry**, and V and W are said to be **isometric** spaces. (We also note that the norm on the left side of this equation is in W, while the norm on the right side is in V. We shall rarely distinguish between norms in different spaces unless there is some possible ambiguity.) Let V have orthonormal basis $\{v_1, \ldots, v_n\}$ so that any $x \in V$ may be written as $x = \sum x_i v_i$. Prove that the mapping $\phi: V \to \mathbb{R}^n$ defined by $\phi(x) = (x_1, \ldots, x_n)$ is an isometry of V onto $\mathbb{R}^n$ (with the standard inner product).

10. Let $\{e_1, e_2, e_3\}$ be an orthonormal basis for $\mathbb{R}^3$, and let $\{u_1, u_2, u_3\}$ be three mutually orthonormal vectors in $\mathbb{R}^3$. Let $u_\lambda{}^i$ denote the *ith* component of u_λ with respect to the basis $\{e_i\}$. Prove the **completeness relation**

$$\sum_{\lambda=1}^{3} u_\lambda{}^i u_\lambda{}^j = \delta_{ij} \ .$$

11. Let W be a finite-dimensional subspace of a possibly infinite-dimensional inner product space V. Prove that $V = W \oplus W^\perp$. [*Hint*: Let $\{w_1, \ldots, w_k\}$ be an orthonormal basis for W, and for any $x \in V$ define

$$x_1 = \sum_{i=1}^{k} \langle w_i, x \rangle w_i$$

and $x_2 = x - x_1$. Show that $x_1 + x_2 \in W + W^\perp$, and that $W \cap W^\perp = \{0\}$.]