CHAPTER 1

Vector Spaces

Linear algebra is the study of linear maps on finite-dimensional vector spaces. Eventually we will learn what all these terms mean. In this chapter we will define vector spaces and discuss their elementary properties.

In some areas of mathematics, including linear algebra, better theorems and more insight emerge if complex numbers are investigated along with real numbers. Thus we begin by introducing the complex numbers and their basic properties.

Complex Numbers

The symbol i was first used to denote $\sqrt{-1}$ by the Swiss mathematician Leonhard Euler in 1777.

You should already be familiar with the basic properties of the set **R** of real numbers. Complex numbers were invented so that we can take square roots of negative numbers. The key idea is to assume we have a square root of -1, denoted i, and manipulate it using the usual rules of arithmetic. Formally, a ***complex number*** is an ordered pair (a, b), where $a, b \in \mathbf{R}$, but we will write this as $a + bi$. The set of all complex numbers is denoted by **C**:

$$\mathbf{C} = \{a + bi : a, b \in \mathbf{R}\}.$$

If $a \in \mathbf{R}$, we identify $a + 0i$ with the real number a. Thus we can think of **R** as a subset of **C**.

Addition and multiplication on **C** are defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$
$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i;$$

here $a, b, c, d \in \mathbf{R}$. Using multiplication as defined above, you should verify that $i^2 = -1$. Do not memorize the formula for the product of two complex numbers; you can always rederive it by recalling that $i^2 = -1$ and then using the usual rules of arithmetic.

You should verify, using the familiar properties of the real numbers, that addition and multiplication on **C** satisfy the following properties:

commutativity
$w + z = z + w$ and $wz = zw$ for all $w, z \in \mathbf{C}$;

associativity
$(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$ and $(z_1 z_2)z_3 = z_1(z_2 z_3)$ for all $z_1, z_2, z_3 \in \mathbf{C}$;

identities
$z + 0 = z$ and $z1 = z$ for all $z \in \mathbf{C}$;

additive inverse
for every $z \in \mathbf{C}$, there exists a unique $w \in \mathbf{C}$ such that $z + w = 0$;

multiplicative inverse
for every $z \in \mathbf{C}$ with $z \neq 0$, there exists a unique $w \in \mathbf{C}$ such that $zw = 1$;

distributive property
$\lambda(w+z) = \lambda w + \lambda z$ for all $\lambda, w, z \in \mathbf{C}$.

For $z \in \mathbf{C}$, we let $-z$ denote the additive inverse of z. Thus $-z$ is the unique complex number such that

$$z + (-z) = 0.$$

Subtraction on $\mathbf{C}$ is defined by

$$w - z = w + (-z)$$

for $w, z \in \mathbf{C}$.

For $z \in \mathbf{C}$ with $z \neq 0$, we let $1/z$ denote the multiplicative inverse of z. Thus $1/z$ is the unique complex number such that

$$z(1/z) = 1.$$

Division on $\mathbf{C}$ is defined by

$$w/z = w(1/z)$$

for $w, z \in \mathbf{C}$ with $z \neq 0$.

So that we can conveniently make definitions and prove theorems that apply to both real and complex numbers, we adopt the following notation:

Throughout this book, **F** stands for either **R** or **C**.

The letter **F** *is used because* **R** *and* **C** *are examples of what are called* ***fields****. In this book we will not need to deal with fields other than* **R** *or* **C**. *Many of the definitions, theorems, and proofs in linear algebra that work for both* **R** *and* **C** *also work without change if an arbitrary field replaces* **R** *or* **C**.

Thus if we prove a theorem involving **F**, we will know that it holds when **F** is replaced with **R** and when **F** is replaced with **C**. Elements of **F** are called ***scalars***. The word "scalar", which means number, is often used when we want to emphasize that an object is a number, as opposed to a vector (vectors will be defined soon).

For $z \in \mathbf{F}$ and m a positive integer, we define z^m to denote the product of z with itself m times:

$$z^m = \underbrace{z \cdot \dots \cdot z}_{m \text{ times}}.$$

Clearly $(z^m)^n = z^{mn}$ and $(wz)^m = w^m z^m$ for all $w, z \in \mathbf{F}$ and all positive integers m, n.

Definition of Vector Space

Before defining what a vector space is, let's look at two important examples. The vector space $\mathbf{R}^2$, which you can think of as a plane, consists of all ordered pairs of real numbers:

$$\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}.$$

The vector space $\mathbf{R}^3$, which you can think of as ordinary space, consists of all ordered triples of real numbers:

$$\mathbf{R}^3 = \{(x, y, z) : x, y, z \in \mathbf{R}\}.$$

To generalize $\mathbf{R}^2$ and $\mathbf{R}^3$ to higher dimensions, we first need to discuss the concept of lists. Suppose n is a nonnegative integer. A ***list*** of ***length*** n is an ordered collection of n objects (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses. A list of length n looks like this:

*Many mathematicians call a list of length n an n-**tuple**.*

$$(x_1, \dots, x_n).$$

Thus a list of length 2 is an ordered pair and a list of length 3 is an ordered triple. For $j \in \{1, \dots, n\}$, we say that x_j is the j^{th} ***coordinate*** of the list above. Thus x_1 is called the first coordinate, x_2 is called the second coordinate, and so on.

Sometimes we will use the word ***list*** without specifying its length. Remember, however, that by definition each list has a finite length that is a nonnegative integer, so that an object that looks like

$$(x_1, x_2, \dots),$$

which might be said to have infinite length, is not a list. A list of length 0 looks like this: $()$. We consider such an object to be a list so that some of our theorems will not have trivial exceptions.

Two lists are equal if and only if they have the same length and the same coordinates in the same order. In other words, $(x_1, \dots, x_m)$ equals $(y_1, \dots, y_n)$ if and only if $m = n$ and $x_1 = y_1, \dots, x_m = y_m$.

Lists differ from sets in two ways: in lists, order matters and repetitions are allowed, whereas in sets, order and repetitions are irrelevant. For example, the lists $(3, 5)$ and $(5, 3)$ are not equal, but the sets $\{3, 5\}$ and $\{5, 3\}$ are equal. The lists $(4, 4)$ and $(4, 4, 4)$ are not equal (they do not have the same length), though the sets $\{4, 4\}$ and $\{4, 4, 4\}$ both equal the set $\{4\}$.

To define the higher-dimensional analogues of $\mathbf{R}^2$ and $\mathbf{R}^3$, we will simply replace $\mathbf{R}$ with $\mathbf{F}$ (which equals $\mathbf{R}$ or $\mathbf{C}$) and replace the 2 or 3 with an arbitrary positive integer. Specifically, fix a positive integer n for the rest of this section. We define $\mathbf{F}^n$ to be the set of all lists of length n consisting of elements of $\mathbf{F}$:

$$\mathbf{F}^n = \{(x_1, \dots, x_n) : x_j \in \mathbf{F} \text{ for } j = 1, \dots, n\}.$$

For example, if $\mathbf{F} = \mathbf{R}$ and n equals 2 or 3, then this definition of $\mathbf{F}^n$ agrees with our previous notions of $\mathbf{R}^2$ and $\mathbf{R}^3$. As another example, $\mathbf{C}^4$ is the set of all lists of four complex numbers:

$$\mathbf{C}^4 = \{(z_1, z_2, z_3, z_4) : z_1, z_2, z_3, z_4 \in \mathbf{C}\}.$$

If $n \geq 4$, we cannot easily visualize $\mathbf{R}^n$ as a physical object. The same problem arises if we work with complex numbers: $\mathbf{C}^1$ can be thought of as a plane, but for $n \geq 2$, the human brain cannot provide geometric models of $\mathbf{C}^n$. However, even if n is large, we can perform algebraic manipulations in $\mathbf{F}^n$ as easily as in $\mathbf{R}^2$ or $\mathbf{R}^3$. For example, addition is defined on $\mathbf{F}^n$ by adding corresponding coordinates:

For an amusing account of how $\mathbf{R}^3$ would be perceived by a creature living in $\mathbf{R}^2$, read ***Flatland: A Romance of Many Dimensions****, by Edwin A. Abbott. This novel, published in 1884, can help creatures living in three-dimensional space, such as ourselves, imagine a physical space of four or more dimensions.*

1.1 $$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n).$$

Often the mathematics of $\mathbf{F}^n$ becomes cleaner if we use a single entity to denote an list of n numbers, without explicitly writing the coordinates. Thus the commutative property of addition on $\mathbf{F}^n$ should be expressed as

$$x + y = y + x$$

for all $x, y \in \mathbf{F}^n$, rather than the more cumbersome

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (y_1, \dots, y_n) + (x_1, \dots, x_n)$$

for all $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbf{F}$ (even though the latter formulation is needed to prove commutativity). If a single letter is used to denote an element of $\mathbf{F}^n$, then the same letter, with appropriate subscripts, is often used when coordinates must be displayed. For example, if $x \in \mathbf{F}^n$, then letting x equal $(x_1, \dots, x_n)$ is good notation. Even better, work with just x and avoid explicit coordinates, if possible.

We let 0 denote the list of length n all of whose coordinates are 0:

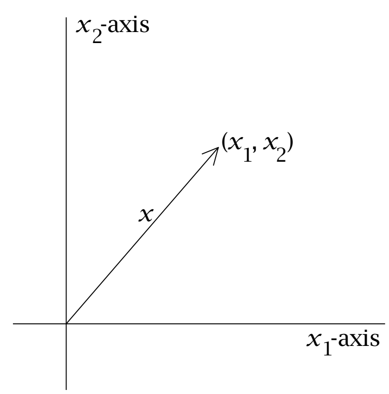
$$0 = (0, \dots, 0).$$

Note that we are using the symbol 0 in two different ways—on the left side of the equation above, 0 denotes a list of length n, whereas on the right side, each 0 denotes a number. This potentially confusing practice actually causes no problems because the context always makes clear what is intended. For example, consider the statement that 0 is an additive identity for $\mathbf{F}^n$:

$$x + 0 = x$$

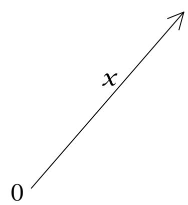
for all $x \in \mathbf{F}^n$. Here 0 must be a list because we have not defined the sum of an element of $\mathbf{F}^n$ (namely, x) and the number 0.

A picture can often aid our intuition. We will draw pictures depicting $\mathbf{R}^2$ because we can easily sketch this space on two-dimensional surfaces such as paper and blackboards. A typical element of $\mathbf{R}^2$ is a point $x = (x_1, x_2)$. Sometimes we think of x not as a point but as an arrow starting at the origin and ending at (x_1, x_2), as in the picture below. When we think of x as an arrow, we refer to it as a ***vector***.



Elements of $\mathbf{R}^2$ *can be thought of as points or as vectors.*

The coordinate axes and the explicit coordinates unnecessarily clutter the picture above, and often you will gain better understanding by dispensing with them and just thinking of the vector, as in the next picture.

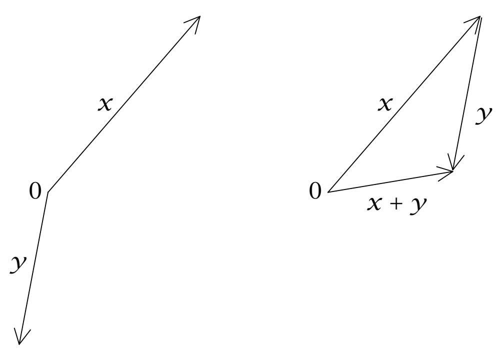


A vector

Whenever we use pictures in $\mathbf{R}^2$ or use the somewhat vague language of points and vectors, remember that these are just aids to our understanding, not substitutes for the actual mathematics that we will develop. Though we cannot draw good pictures in high-dimensional spaces, the elements of these spaces are as rigorously defined as elements of $\mathbf{R}^2$. For example, $(2, -3, 17, \pi, \sqrt{2})$ is an element of $\mathbf{R}^5$, and we may casually refer to it as a point in $\mathbf{R}^5$ or a vector in $\mathbf{R}^5$ without worrying about whether the geometry of $\mathbf{R}^5$ has any physical meaning.

Mathematical models of the economy often have thousands of variables, say $x_1, \dots, x_{5000}$, which means that we must operate in $\mathbf{R}^{5000}$. Such a space cannot be dealt with geometrically, but the algebraic approach works well. That's why our subject is called linear ***algebra****.*

Recall that we defined the sum of two elements of $\mathbf{F}^n$ to be the element of $\mathbf{F}^n$ obtained by adding corresponding coordinates; see 1.1. In the special case of $\mathbf{R}^2$, addition has a simple geometric interpretation. Suppose we have two vectors x and y in $\mathbf{R}^2$ that we want to add, as in the left side of the picture below. Move the vector y parallel to itself so that its initial point coincides with the end point of the vector x. The sum $x + y$ then equals the vector whose initial point equals the initial point of x and whose end point equals the end point of the moved vector y, as in the right side of the picture below.



The sum of two vectors

Our treatment of the vector y in the picture above illustrates a standard philosophy when we think of vectors in $\mathbf{R}^2$ as arrows: we can move an arrow parallel to itself (not changing its length or direction) and still think of it as the same vector.

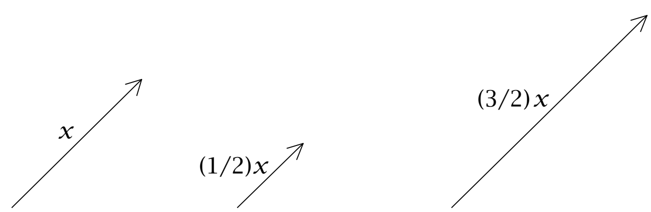
Having dealt with addition in $\mathbf{F}^n$, we now turn to multiplication. We could define a multiplication on $\mathbf{F}^n$ in a similar fashion, starting with two elements of $\mathbf{F}^n$ and getting another element of $\mathbf{F}^n$ by multiplying corresponding coordinates. Experience shows that this definition is not useful for our purposes. Another type of multiplication, called scalar multiplication, will be central to our subject. Specifically, we need to define what it means to multiply an element of $\mathbf{F}^n$ by an element of $\mathbf{F}$. We make the obvious definition, performing the multiplication in each coordinate:

$$a(x_1, \dots, x_n) = (ax_1, \dots, ax_n);$$

here $a \in \mathbf{F}$ and $(x_1, \dots, x_n) \in \mathbf{F}^n$.

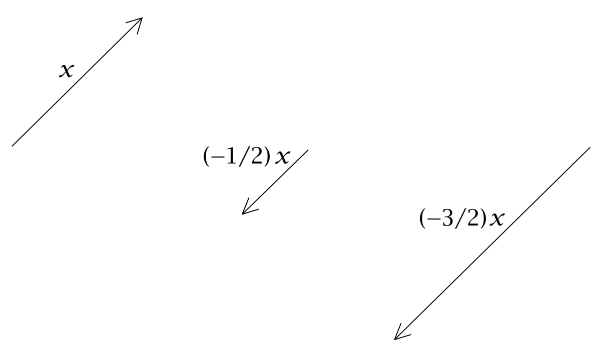
In scalar multiplication, we multiply together a scalar and a vector, getting a vector. You may be familiar with the dot product in $\mathbf{R}^2$ or $\mathbf{R}^3$, in which we multiply together two vectors and obtain a scalar. Generalizations of the dot product will become important when we study inner products in Chapter 6. You may also be familiar with the cross product in $\mathbf{R}^3$, in which we multiply together two vectors and obtain another vector. No useful generalization of this type of multiplication exists in higher dimensions.

Scalar multiplication has a nice geometric interpretation in $\mathbf{R}^2$. If a is a positive number and x is a vector in $\mathbf{R}^2$, then ax is the vector that points in the same direction as x and whose length is a times the length of x. In other words, to get ax, we shrink or stretch x by a factor of a, depending upon whether $a < 1$ or $a > 1$. The next picture illustrates this point.



Multiplication by positive scalars

If a is a negative number and x is a vector in $\mathbf{R}^2$, then ax is the vector that points in the opposite direction as x and whose length is $|a|$ times the length of x, as illustrated in the next picture.



Multiplication by negative scalars

The motivation for the definition of a vector space comes from the important properties possessed by addition and scalar multiplication on $\mathbf{F}^n$. Specifically, addition on $\mathbf{F}^n$ is commutative and associative and has an identity, namely, 0. Every element has an additive inverse. Scalar multiplication on $\mathbf{F}^n$ is associative, and scalar multiplication by 1 acts as a multiplicative identity should. Finally, addition and scalar multiplication on $\mathbf{F}^n$ are connected by distributive properties.

We will define a vector space to be a set V along with an addition and a scalar multiplication on V that satisfy the properties discussed in the previous paragraph. By an ***addition*** on V we mean a function that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$. By a ***scalar multiplication*** on V we mean a function that assigns an element $av \in V$ to each $a \in \mathbf{F}$ and each $v \in V$.

Now we are ready to give the formal definition of a vector space. A ***vector space*** is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:

commutativity

$u + v = v + u$ for all $u, v \in V$;

associativity

$(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbf{F}$;

additive identity

there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$;

additive inverse

for every $v \in V$, there exists $w \in V$ such that $v + w = 0$;

multiplicative identity

$1v = v$ for all $v \in V$;

distributive properties

$a(u + v) = au + av$ and $(a + b)u = au + bu$ for all $a, b \in \mathbf{F}$ and all $u, v \in V$.

The scalar multiplication in a vector space depends upon $\mathbf{F}$. Thus when we need to be precise, we will say that V is a vector space over $\mathbf{F}$ instead of saying simply that V is a vector space. For example, $\mathbf{R}^n$ is a vector space over $\mathbf{R}$, and $\mathbf{C}^n$ is a vector space over $\mathbf{C}$. Frequently, a vector space over $\mathbf{R}$ is called a ***real vector space*** and a vector space over

C is called a ***complex vector space***. Usually the choice of **F** is either obvious from the context or irrelevant, and thus we often assume that **F** is lurking in the background without specifically mentioning it.

Elements of a vector space are called ***vectors*** or ***points***. This geometric language sometimes aids our intuition.

Not surprisingly, $\mathbf{F}^n$ is a vector space over **F**, as you should verify. Of course, this example motivated our definition of vector space.

The simplest vector space contains only one point. In other words, $\{0\}$ is a vector space, though not a very interesting one.

For another example, consider $\mathbf{F}^\infty$, which is defined to be the set of all sequences of elements of **F**:

$$\mathbf{F}^\infty = \{(x_1, x_2, \dots) : x_j \in \mathbf{F} \text{ for } j = 1, 2, \dots\}.$$

Addition and scalar multiplication on $\mathbf{F}^\infty$ are defined as expected:

$$(x_1, x_2, \dots) + (y_1, y_2, \dots) = (x_1 + y_1, x_2 + y_2, \dots),$$
$$a(x_1, x_2, \dots) = (ax_1, ax_2, \dots).$$

With these definitions, $\mathbf{F}^\infty$ becomes a vector space over **F**, as you should verify. The additive identity in this vector space is the sequence consisting of all 0's.

Our next example of a vector space involves polynomials. A function $p\colon \mathbf{F} \to \mathbf{F}$ is called a ***polynomial*** with coefficients in **F** if there exist $a_0, \dots, a_m \in \mathbf{F}$ such that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

Though $\mathbf{F}^n$ is our crucial example of a vector space, not all vector spaces consist of lists. For example, the elements of $\mathcal{P}(\mathbf{F})$ consist of functions on $\mathbf{F}$, not lists. In general, a vector space is an abstract entity whose elements might be lists, functions, or weird objects.

for all $z \in \mathbf{F}$. We define $\mathcal{P}(\mathbf{F})$ to be the set of all polynomials with coefficients in **F**. Addition on $\mathcal{P}(\mathbf{F})$ is defined as you would expect: if $p, q \in \mathcal{P}(\mathbf{F})$, then $p + q$ is the polynomial defined by

$$(p + q)(z) = p(z) + q(z)$$

for $z \in \mathbf{F}$. For example, if p is the polynomial defined by $p(z) = 2z + z^3$ and q is the polynomial defined by $q(z) = 7 + 4z$, then $p + q$ is the polynomial defined by $(p + q)(z) = 7 + 6z + z^3$. Scalar multiplication on $\mathcal{P}(\mathbf{F})$ also has the obvious definition: if $a \in \mathbf{F}$ and $p \in \mathcal{P}(\mathbf{F})$, then ap is the polynomial defined by

$$(ap)(z) = ap(z)$$

for $z \in \mathbf{F}$. With these definitions of addition and scalar multiplication, $\mathcal{P}(\mathbf{F})$ is a vector space, as you should verify. The additive identity in this vector space is the polynomial all of whose coefficients equal 0.

Soon we will see further examples of vector spaces, but first we need to develop some of the elementary properties of vector spaces.

Properties of Vector Spaces

The definition of a vector space requires that it have an additive identity. The proposition below states that this identity is unique.

1.2 Proposition: *A vector space has a unique additive identity.*

PROOF: Suppose 0 and $0'$ are both additive identities for some vector space V. Then

$$0' = 0' + 0 = 0,$$

where the first equality holds because 0 is an additive identity and the second equality holds because $0'$ is an additive identity. Thus $0' = 0$, proving that V has only one additive identity. ■

The symbol ■ means "end of the proof".

Each element v in a vector space has an additive inverse, an element w in the vector space such that $v + w = 0$. The next proposition shows that each element in a vector space has only one additive inverse.

1.3 Proposition: *Every element in a vector space has a unique additive inverse.*

PROOF: Suppose V is a vector space. Let $v \in V$. Suppose that w and w' are additive inverses of v. Then

$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w'.$$

Thus $w = w'$, as desired. ■

Because additive inverses are unique, we can let $-v$ denote the additive inverse of a vector v. We define $w - v$ to mean $w + (-v)$.

Almost all the results in this book will involve some vector space. To avoid being distracted by having to restate frequently something such as "Assume that V is a vector space", we now make the necessary declaration once and for all:

Let's agree that for the rest of the book V will denote a vector space over $\mathbf{F}$.

Because of associativity, we can dispense with parentheses when dealing with additions involving more than two elements in a vector space. For example, we can write $u+v+w$ without parentheses because the two possible interpretations of that expression, namely, $(u+v)+w$ and $u+(v+w)$, are equal. We first use this familiar convention of not using parentheses in the next proof. In the next proposition, 0 denotes a scalar (the number $0 \in \mathbf{F}$) on the left side of the equation and a vector (the additive identity of V) on the right side of the equation.

Note that 1.4 and 1.5 assert something about scalar multiplication and the additive identity of V. The only part of the definition of a vector space that connects scalar multiplication and vector addition is the distributive property. Thus the distributive property must be used in the proofs.

1.4 Proposition: $0v = 0$ *for every* $v \in V$.

PROOF: For $v \in V$, we have

$$0v = (0+0)v = 0v + 0v.$$

Adding the additive inverse of $0v$ to both sides of the equation above gives $0 = 0v$, as desired. ■

In the next proposition, 0 denotes the additive identity of V. Though their proofs are similar, 1.4 and 1.5 are not identical. More precisely, 1.4 states that the product of the scalar 0 and any vector equals the vector 0, whereas 1.5 states that the product of any scalar and the vector 0 equals the vector 0.

1.5 Proposition: $a0 = 0$ *for every* $a \in \mathbf{F}$.

PROOF: For $a \in \mathbf{F}$, we have

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

Adding the additive inverse of $a0$ to both sides of the equation above gives $0 = a0$, as desired. ■

Now we show that if an element of V is multiplied by the scalar -1, then the result is the additive inverse of the element of V.

1.6 Proposition: $(-1)v = -v$ *for every* $v \in V$.

PROOF: For $v \in V$, we have

$$v + (-1)v = 1v + (-1)v = \big(1 + (-1)\big)v = 0v = 0.$$

This equation says that $(-1)v$, when added to v, gives 0. Thus $(-1)v$ must be the additive inverse of v, as desired. ■

Subspaces

A subset U of V is called a ***subspace*** of V if U is also a vector space (using the same addition and scalar multiplication as on V). For example,

$$\{(x_1, x_2, 0) : x_1, x_2 \in \mathbf{F}\}$$

is a subspace of $\mathbf{F}^3$.

*Some mathematicians use the term **linear subspace**, which means the same as subspace.*

If U is a subset of V, then to check that U is a subspace of V we need only check that U satisfies the following:

additive identity

$0 \in U$

closed under addition

$u, v \in U$ implies $u + v \in U$;

closed under scalar multiplication

$a \in \mathbf{F}$ and $u \in U$ implies $au \in U$.

The first condition insures that the additive identity of V is in U. The second condition insures that addition makes sense on U. The third condition insures that scalar multiplication makes sense on U. To show that U is a vector space, the other parts of the definition of a vector space do not need to be checked because they are automatically satisfied. For example, the associative and commutative properties of addition automatically hold on U because they hold on the larger space V. As another example, if the third condition above holds and $u \in U$, then $-u$ (which equals $(-1)u$ by 1.6) is also in U, and hence every element of U has an additive inverse in U.

Clearly $\{0\}$ is the smallest subspace of V and V itself is the largest subspace of V. The empty set is not a subspace of V because a subspace must be a vector space and a vector space must contain at least one element, namely, an additive identity.

The three conditions above usually enable us to determine quickly whether a given subset of V is a subspace of V. For example, if $b \in \mathbf{F}$, then

$$\{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = 5x_4 + b\}$$

is a subspace of $\mathbf{F}^4$ if and only if $b = 0$, as you should verify. As another example, you should verify that

$$\{p \in \mathcal{P}(\mathbf{F}) : p(3) = 0\}$$

is a subspace of $\mathcal{P}(\mathbf{F})$.

The subspaces of $\mathbf{R}^2$ are precisely $\{0\}$, $\mathbf{R}^2$, and all lines in $\mathbf{R}^2$ through the origin. The subspaces of $\mathbf{R}^3$ are precisely $\{0\}$, $\mathbf{R}^3$, all lines in $\mathbf{R}^3$

through the origin, and all planes in $\mathbf{R}^3$ through the origin. To prove that all these objects are indeed subspaces is easy—the hard part is to show that they are the only subspaces of $\mathbf{R}^2$ or $\mathbf{R}^3$. That task will be easier after we introduce some additional tools in the next chapter.

Sums and Direct Sums

In later chapters, we will find that the notions of vector space sums and direct sums are useful. We define these concepts here.

When dealing with vector spaces, we are usually interested only in subspaces, as opposed to arbitrary subsets. The union of subspaces is rarely a subspace (see Exercise 9 in this chapter), which is why we usually work with sums rather than unions.

Suppose $U_1, \dots, U_m$ are subspaces of V. The ***sum*** of $U_1, \dots, U_m$, denoted $U_1 + \cdots + U_m$, is defined to be the set of all possible sums of elements of $U_1, \dots, U_m$. More precisely,

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \dots, u_m \in U_m\}.$$

You should verify that if $U_1, \dots, U_m$ are subspaces of V, then the sum $U_1 + \cdots + U_m$ is a subspace of V.

Let's look at some examples of sums of subspaces. Suppose U is the set of all elements of $\mathbf{F}^3$ whose second and third coordinates equal 0, and W is the set of all elements of $\mathbf{F}^3$ whose first and third coordinates equal 0:

$$U = \{(x, 0, 0) \in \mathbf{F}^3 : x \in \mathbf{F}\} \quad \text{and} \quad W = \{(0, y, 0) \in \mathbf{F}^3 : y \in \mathbf{F}\}.$$

Then

Sums of subspaces in the theory of vector spaces are analogous to unions of subsets in set theory. Given two subspaces of a vector space, the smallest subspace containing them is their sum. Analogously, given two subsets of a set, the smallest subset containing them is their union.

1.7 $$U + W = \{(x, y, 0) : x, y \in \mathbf{F}\},$$

as you should verify.

As another example, suppose U is as above and W is the set of all elements of $\mathbf{F}^3$ whose first and second coordinates equal each other and whose third coordinate equals 0:

$$W = \{(y, y, 0) \in \mathbf{F}^3 : y \in \mathbf{F}\}.$$

Then $U + W$ is also given by 1.7, as you should verify.

Suppose $U_1, \dots, U_m$ are subspaces of V. Clearly $U_1, \dots, U_m$ are all contained in $U_1 + \cdots + U_m$ (to see this, consider sums $u_1 + \cdots + u_m$ where all except one of the u's are 0). Conversely, any subspace of V containing $U_1, \dots, U_m$ must contain $U_1 + \cdots + U_m$ (because subspaces must contain all finite sums of their elements). Thus $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$.

Suppose $U_1, \ldots, U_m$ are subspaces of V such that $V = U_1 + \cdots + U_m$. Thus every element of V can be written in the form

$$u_1 + \cdots + u_m,$$

where each $u_j \in U_j$. We will be especially interested in cases where each vector in V can be uniquely represented in the form above. This situation is so important that we give it a special name: direct sum. Specifically, we say that V is the ***direct sum*** of subspaces $U_1, \ldots, U_m$, written $V = U_1 \oplus \cdots \oplus U_m$, if each element of V can be written uniquely as a sum $u_1 + \cdots + u_m$, where each $u_j \in U_j$.

The symbol $\oplus$, consisting of a plus sign inside a circle, is used to denote direct sums as a reminder that we are dealing with a special type of sum of subspaces—each element in the direct sum can be represented only one way as a sum of elements from the specified subspaces.

Let's look at some examples of direct sums. Suppose U is the subspace of $\mathbf{F}^3$ consisting of those vectors whose last coordinate equals 0, and W is the subspace of $\mathbf{F}^3$ consisting of those vectors whose first two coordinates equal 0:

$$U = \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\} \quad \text{and} \quad W = \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\}.$$

Then $\mathbf{F}^3 = U \oplus W$, as you should verify.

As another example, suppose U_j is the subspace of $\mathbf{F}^n$ consisting of those vectors whose coordinates are all 0, except possibly in the j^{th} slot (for example, $U_2 = \{(0, x, 0, \ldots, 0) \in \mathbf{F}^n : x \in \mathbf{F}\}$). Then

$$\mathbf{F}^n = U_1 \oplus \cdots \oplus U_n,$$

as you should verify.

As a final example, consider the vector space $\mathcal{P}(\mathbf{F})$ of all polynomials with coefficients in $\mathbf{F}$. Let U_e denote the subspace of $\mathcal{P}(\mathbf{F})$ consisting of all polynomials p of the form

$$p(z) = a_0 + a_2 z^2 + \cdots + a_{2m} z^{2m},$$

and let U_o denote the subspace of $\mathcal{P}(\mathbf{F})$ consisting of all polynomials p of the form

$$p(z) = a_1 z + a_3 z^3 + \cdots + a_{2m+1} z^{2m+1};$$

here m is a nonnegative integer and $a_0, \ldots, a_{2m+1} \in \mathbf{F}$ (the notations U_e and U_o should remind you of even and odd powers of z). You should verify that

$$\mathcal{P}(\mathbf{F}) = U_e \oplus U_o.$$

Sometimes nonexamples add to our understanding as much as examples. Consider the following three subspaces of $\mathbf{F}^3$:

$$\begin{aligned}
U_1 &= \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\};\\
U_2 &= \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\};\\
U_3 &= \{(0, y, y) \in \mathbf{F}^3 : y \in \mathbf{F}\}.
\end{aligned}$$

Clearly $\mathbf{F}^3 = U_1 + U_2 + U_3$ because an arbitrary vector $(x, y, z) \in \mathbf{F}^3$ can be written as

$$(x, y, z) = (x, y, 0) + (0, 0, z) + (0, 0, 0),$$

where the first vector on the right side is in U_1, the second vector is in U_2, and the third vector is in U_3. However, $\mathbf{F}^3$ does not equal the direct sum of U_1, U_2, U_3 because the vector $(0, 0, 0)$ can be written in two different ways as a sum $u_1+u_2+u_3$, with each $u_j \in U_j$. Specifically, we have

$$(0, 0, 0) = (0, 1, 0) + (0, 0, 1) + (0, -1, -1)$$

and, of course,

$$(0, 0, 0) = (0, 0, 0) + (0, 0, 0) + (0, 0, 0),$$

where the first vector on the right side of each equation above is in U_1, the second vector is in U_2, and the third vector is in U_3.

In the example above, we showed that something is not a direct sum by showing that 0 does not have a unique representation as a sum of appropriate vectors. The definition of direct sum requires that every vector in the space have a unique representation as an appropriate sum. Suppose we have a collection of subspaces whose sum equals the whole space. The next proposition shows that when deciding whether this collection of subspaces is a direct sum, we need only consider whether 0 can be uniquely written as an appropriate sum.

1.8 Proposition: *Suppose that $U_1, \dots, U_n$ are subspaces of V. Then $V = U_1 \oplus \cdots \oplus U_n$ if and only if both the following conditions hold:*

(a) $V = U_1 + \cdots + U_n$;

(b) *the only way to write 0 as a sum $u_1 + \cdots + u_n$, where each $u_j \in U_j$, is by taking all the u_j's equal to 0.*

PROOF: First suppose that $V = U_1 \oplus \cdots \oplus U_n$. Clearly (a) holds (because of how sum and direct sum are defined). To prove (b), suppose that $u_1 \in U_1, \dots, u_n \in U_n$ and

$$0 = u_1 + \cdots + u_n.$$

Then each u_j must be 0 (this follows from the uniqueness part of the definition of direct sum because $0 = 0 + \cdots + 0$ and $0 \in U_1, \dots, 0 \in U_n$), proving (b).

Now suppose that (a) and (b) hold. Let $v \in V$. By (a), we can write

$$v = u_1 + \cdots + u_n$$

for some $u_1 \in U_1, \dots, u_n \in U_n$. To show that this representation is unique, suppose that we also have

$$v = v_1 + \cdots + v_n,$$

where $v_1 \in U_1, \dots, v_n \in U_n$. Subtracting these two equations, we have

$$0 = (u_1 - v_1) + \cdots + (u_n - v_n).$$

Clearly $u_1 - v_1 \in U_1, \dots, u_n - v_n \in U_n$, so the equation above and (b) imply that each $u_j - v_j = 0$. Thus $u_1 = v_1, \dots, u_n = v_n$, as desired. ■

The next proposition gives a simple condition for testing which pairs of subspaces give a direct sum. Note that this proposition deals only with the case of two subspaces. When asking about a possible direct sum with more than two subspaces, it is not enough to test that any two of the subspaces intersect only at 0. To see this, consider the nonexample presented just before 1.8. In that nonexample, we had $\mathbf{F}^3 = U_1 + U_2 + U_3$, but $\mathbf{F}^3$ did not equal the direct sum of U_1, U_2, U_3. However, in that nonexample, we have $U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{0\}$ (as you should verify). The next proposition shows that with just two subspaces we get a nice necessary and sufficient condition for a direct sum.

Sums of subspaces are analogous to unions of subsets. Similarly, direct sums of subspaces are analogous to disjoint unions of subsets. No two subspaces of a vector space can be disjoint because both must contain 0. So disjointness is replaced, at least in the case of two subspaces, with the requirement that the intersection equals $\{0\}$.

1.9 Proposition: *Suppose that U and W are subspaces of V. Then $V = U \oplus W$ if and only if $V = U + W$ and $U \cap W = \{0\}$.*

PROOF: First suppose that $V = U \oplus W$. Then $V = U + W$ (by the definition of direct sum). Also, if $v \in U \cap W$, then $0 = v + (-v)$, where

$v \in U$ and $-v \in W$. By the unique representation of 0 as the sum of a vector in U and a vector in W, we must have $v = 0$. Thus $U \cap W = \{0\}$, completing the proof in one direction.

To prove the other direction, now suppose that $V = U + W$ and $U \cap W = \{0\}$. To prove that $V = U \oplus W$, suppose that

$$0 = u + w,$$

where $u \in U$ and $w \in W$. To complete the proof, we need only show that $u = w = 0$ (by 1.8). The equation above implies that $u = -w \in W$. Thus $u \in U \cap W$, and hence $u = 0$. This, along with equation above, implies that $w = 0$, completing the proof. ■

Exercises

1. Suppose a and b are real numbers, not both 0. Find real numbers c and d such that
$$1/(a+bi) = c + di.$$

2. Show that
$$\frac{-1+\sqrt{3}i}{2}$$
is a cube root of 1 (meaning that its cube equals 1).

3. Prove that $-(-v) = v$ for every $v \in V$.

4. Prove that if $a \in \mathbf{F}$, $v \in V$, and $av = 0$, then $a = 0$ or $v = 0$.

5. For each of the following subsets of $\mathbf{F}^3$, determine whether it is a subspace of $\mathbf{F}^3$:

   (a) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 0\}$;

   (b) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 4\}$;

   (c) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 x_2 x_3 = 0\}$;

   (d) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 = 5x_3\}$.

6. Give an example of a nonempty subset U of $\mathbf{R}^2$ such that U is closed under addition and under taking additive inverses (meaning $-u \in U$ whenever $u \in U$), but U is not a subspace of $\mathbf{R}^2$.

7. Give an example of a nonempty subset U of $\mathbf{R}^2$ such that U is closed under scalar multiplication, but U is not a subspace of $\mathbf{R}^2$.

8. Prove that the intersection of any collection of subspaces of V is a subspace of V.

9. Prove that the union of two subspaces of V is a subspace of V if and only if one of the subspaces is contained in the other.

10. Suppose that U is a subspace of V. What is $U + U$?

11. Is the operation of addition on the subspaces of V commutative? Associative? (In other words, if U_1, U_2, U_3 are subspaces of V, is $U_1 + U_2 = U_2 + U_1$? Is $(U_1 + U_2) + U_3 = U_1 + (U_2 + U_3)$?)

12. Does the operation of addition on the subspaces of V have an additive identity? Which subspaces have additive inverses?

13. Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that
$$U_1 + W = U_2 + W,$$
then $U_1 = U_2$.

14. Suppose U is the subspace of $\mathcal{P}(\mathbf{F})$ consisting of all polynomials p of the form
$$p(z) = az^2 + bz^5,$$
where $a, b \in \mathbf{F}$. Find a subspace W of $\mathcal{P}(\mathbf{F})$ such that $\mathcal{P}(\mathbf{F}) = U \oplus W$.

15. Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that
$$V = U_1 \oplus W \quad \text{and} \quad V = U_2 \oplus W,$$
then $U_1 = U_2$.