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 linear algebra, better theorems and more insight emerge if complex numbers are investigated along with real numbers. Thus we will begin by introducing the complex numbers and their basic properties.

We will generalize the examples of a plane and of ordinary space to $\mathbf{R}^n$ and $\mathbf{C}^n$, which we then will generalize to the notion of a vector space. As we will see, a vector space is a set with operations of addition and scalar multiplication that satisfy natural algebraic properties.

Then our next topic will be subspaces, which play a role for vector spaces analogous to the role played by subsets for sets. Finally, we will look at sums of subspaces (analogous to unions of subsets) and direct sums of subspaces (analogous to unions of disjoint sets).

Pierre Louis Dumesnil, Nils Forsberg

René Descartes explaining his work to Queen Christina of Sweden. Vector spaces are a generalization of the description of a plane using two coordinates, as published by Descartes in 1637.

1A $\mathbf{R}^n$ and $\mathbf{C}^n$

Complex Numbers

You should already be familiar with basic properties of the set $\mathbf{R}$ of real numbers. Complex numbers were invented so that we can take square roots of negative numbers. The idea is to assume we have a square root of -1, denoted by i, that obeys the usual rules of arithmetic. Here are the formal definitions.

1.1 definition: *complex numbers,* $\mathbf{C}$

- 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 $\mathbf{C}$:

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

- *Addition* and *multiplication* on $\mathbf{C}$ are defined by

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

here $a, b, c, d \in \mathbf{R}$.

If $a \in \mathbf{R}$, we identify $a + 0i$ with the real number a. Thus we think of $\mathbf{R}$ as a subset of $\mathbf{C}$. We usually write $0 + bi$ as just bi, and we usually write $0 + 1i$ as just i.

To motivate the definition of complex multiplication given above, pretend that we knew that $i^2 = -1$ and then use the usual rules of arithmetic to derive the formula above for the product of two complex numbers. Then use that formula to verify that we indeed have

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

$$i^2 = -1.$$

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 (as given by 1.3). The next example illustrates this procedure.

1.2 example: *complex arithmetic*

The product $(2 + 3i)(4 + 5i)$ can be evaluated by applying the distributive and commutative properties from 1.3:

$$\begin{aligned}(2 + 3i)(4 + 5i) &= 2 \cdot (4 + 5i) + (3i)(4 + 5i)\\ &= 2 \cdot 4 + 2 \cdot 5i + 3i \cdot 4 + (3i)(5i)\\ &= 8 + 10i + 12i - 15\\ &= -7 + 22i.\end{aligned}$$

Our first result states that complex addition and complex multiplication have the familiar properties that we expect.

1.3 *properties of complex arithmetic*

commutativity

$\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbf{C}$.

associativity

$(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ and $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

identities

$\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$ for all $\lambda \in \mathbf{C}$.

additive inverse

For every $\alpha \in \mathbf{C}$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha + \beta = 0$.

multiplicative inverse

For every $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha\beta = 1$.

distributive property

$\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$.

The properties above are proved using the familiar properties of real numbers and the definitions of complex addition and multiplication. The next example shows how commutativity of complex multiplication is proved. Proofs of the other properties above are left as exercises.

1.4 example: *commutativity of complex multiplication*

To show that $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbf{C}$, suppose

$$\alpha = a + bi \quad \text{and} \quad \beta = c + di,$$

where $a, b, c, d \in \mathbf{R}$. Then the definition of multiplication of complex numbers shows that

$$\begin{aligned} \alpha\beta &= (a + bi)(c + di) \\ &= (ac - bd) + (ad + bc)i \end{aligned}$$

and

$$\begin{aligned} \beta\alpha &= (c + di)(a + bi) \\ &= (ca - db) + (cb + da)i. \end{aligned}$$

The equations above and the commutativity of multiplication and addition of real numbers show that $\alpha\beta = \beta\alpha$.

Next, we define the additive and multiplicative inverses of complex numbers, and then use those inverses to define subtraction and division operations with complex numbers.

1.5 definition: $-\alpha$, *subtraction*, $1/\alpha$, *division*

Suppose $\alpha, \beta \in \mathbf{C}$.

- Let $-\alpha$ denote the additive inverse of α. Thus $-\alpha$ is the unique complex number such that
$$\alpha + (-\alpha) = 0.$$
- *Subtraction* on $\mathbf{C}$ is defined by
$$\beta - \alpha = \beta + (-\alpha).$$
- For $\alpha \neq 0$, let $1/\alpha$ and $\frac{1}{\alpha}$ denote the multiplicative inverse of α. Thus $1/\alpha$ is the unique complex number such that
$$\alpha(1/\alpha) = 1.$$
- For $\alpha \neq 0$, *division* by α is defined by
$$\beta/\alpha = \beta(1/\alpha).$$

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

1.6 notation: $\mathbf{F}$

Throughout this book, $\mathbf{F}$ stands for either $\mathbf{R}$ or $\mathbf{C}$.

Thus if we prove a theorem involving $\mathbf{F}$, we will know that it holds when $\mathbf{F}$ is replaced with $\mathbf{R}$ and when $\mathbf{F}$ is replaced with $\mathbf{C}$.

The letter $\mathbf{F}$ is used because $\mathbf{R}$ and $\mathbf{C}$ are examples of what are called ***fields***.

Elements of $\mathbf{F}$ are called *scalars*. The word "scalar" (which is just a fancy word for "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 $\alpha \in \mathbf{F}$ and m a positive integer, we define α^m to denote the product of α with itself m times:
$$\alpha^m = \underbrace{\alpha \cdots \alpha}_{m \text{ times}}.$$
This definition implies that
$$\left(\alpha^m\right)^n = \alpha^{mn} \quad \text{and} \quad (\alpha\beta)^m = \alpha^m \beta^m$$
for all $\alpha, \beta \in \mathbf{F}$ and all positive integers m, n.

Lists

Before defining $\mathbf{R}^n$ and $\mathbf{C}^n$, we look at two important examples.

1.7 example: *$\mathbf{R}^2$ and $\mathbf{R}^3$*

- The set $\mathbf{R}^2$, which you can think of as a plane, is the set of all ordered pairs of real numbers:
$$\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}.$$
- The set $\mathbf{R}^3$, which you can think of as ordinary space, is the set 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.

1.8 definition: *list, length*

- Suppose n is a nonnegative integer. A *list* of *length* n is an ordered collection of n elements (which might be numbers, other lists, or more abstract objects).
- Two lists are equal if and only if they have the same length and the same elements in the same order.

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

Lists are often written as elements separated by commas and surrounded by parentheses. Thus a list of length two is an ordered pair that might be written as (a, b). A list of length three is an ordered triple that might be written as (x, y, z). A list of length n might look like this:
$$(z_1, \ldots, z_n).$$

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. Thus an object that looks like $(x_1, x_2, \ldots)$, 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.

Lists differ from finite sets in two ways: in lists, order matters and repetitions have meaning; in sets, order and repetitions are irrelevant.

1.9 example: *lists versus sets*

- 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), although the sets $\{4, 4\}$ and $\{4, 4, 4\}$ both equal the set $\{4\}$.

$\mathbf{F}^n$

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.

1.10 notation: n

Fix a positive integer n for the rest of this chapter.

1.11 definition: $\mathbf{F}^n$, *coordinate*

$\mathbf{F}^n$ is the set of all lists of length n of elements of $\mathbf{F}$:

$$\mathbf{F}^n = \{(x_1, \ldots, x_n) : x_k \in \mathbf{F} \text{ for } k = 1, \ldots, n\}.$$

For $(x_1, \ldots, x_n) \in \mathbf{F}^n$ and $k \in \{1, \ldots, n\}$, we say that x_k is the k^{th} *coordinate* of $(x_1, \ldots, x_n)$.

If $\mathbf{F} = \mathbf{R}$ and n equals 2 or 3, then the definition above of $\mathbf{F}^n$ agrees with our previous notions of $\mathbf{R}^2$ and $\mathbf{R}^3$.

1.12 example: $\mathbf{C}^4$

$\mathbf{C}^4$ 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 visualize $\mathbf{R}^n$ as a physical object. Similarly, $\mathbf{C}^1$ can be thought of as a plane, but for $n \geq 2$, the human brain cannot provide a full image 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 in $\mathbf{F}^n$ is defined as follows.

*Read **Flatland: A Romance of Many Dimensions**, by Edwin A. Abbott, for an amusing account of how $\mathbf{R}^3$ would be perceived by creatures living in $\mathbf{R}^2$. This novel, published in 1884, may help you imagine a physical space of four or more dimensions.*

1.13 definition: *addition in* $\mathbf{F}^n$

Addition in $\mathbf{F}^n$ is defined by adding corresponding coordinates:

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

Often the mathematics of $\mathbf{F}^n$ becomes cleaner if we use a single letter to denote a list of n numbers, without explicitly writing the coordinates. For example, the next result is stated with x and y in $\mathbf{F}^n$ even though the proof requires the more cumbersome notation of $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$.

1.14 *commutativity of addition in* $\mathbf{F}^n$

If $x, y \in \mathbf{F}^n$, then $x + y = y + x$.

Proof Suppose $x = (x_1, \dots, x_n) \in \mathbf{F}^n$ and $y = (y_1, \dots, y_n) \in \mathbf{F}^n$. Then

$$\begin{aligned} x + y &= (x_1, \dots, x_n) + (y_1, \dots, y_n) \\ &= (x_1 + y_1, \dots, x_n + y_n) \\ &= (y_1 + x_1, \dots, y_n + x_n) \\ &= (y_1, \dots, y_n) + (x_1, \dots, x_n) \\ &= y + x, \end{aligned}$$

where the second and fourth equalities above hold because of the definition of addition in $\mathbf{F}^n$ and the third equality holds because of the usual commutativity of addition in $\mathbf{F}$. ■

The symbol ■ means "end of proof".

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, as shown in the proof above. Even better, work with just x and avoid explicit coordinates when possible.

1.15 notation: 0

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

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

Here we are using the symbol 0 in two different ways—on the left side of the equation above, the symbol 0 denotes a list of length n, which is an element of $\mathbf{F}^n$, whereas on the right side, each 0 denotes a number. This potentially confusing practice actually causes no problems because the context should always make clear which 0 is intended.

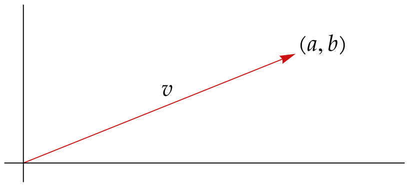
1.16 example: *context determines which* 0 *is intended*

Consider the statement that 0 is an additive identity for $\mathbf{F}^n$:

$$x + 0 = x \quad \text{for all } x \in \mathbf{F}^n.$$

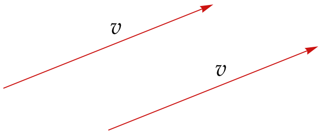
Here the 0 above is the list defined in 1.15, not the number 0, because we have not defined the sum of an element of $\mathbf{F}^n$ (namely, x) and the number 0.

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



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

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. With that viewpoint, you will often gain better understanding by dispensing with the coordinate axes and the explicit coordinates and just thinking of the vector, as shown in the figure here. The two arrows shown here have the same length and same direction, so we think of them as the same vector.



A vector.

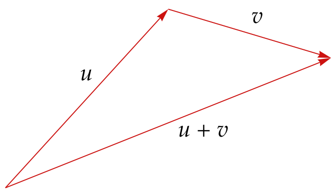
Whenever we use pictures in $\mathbf{R}^2$ or use the somewhat vague language of points and vectors, remember that these are just aids to our understanding, not substitutes for the actual mathematics that we will develop. Although we cannot draw good pictures in high-dimensional spaces, the elements of these spaces are as rigorously defined as elements of $\mathbf{R}^2$.

*Mathematical models of the economy can have thousands of variables, say $x_1, \ldots, x_{5000}$, which means that we must work in $\mathbf{R}^{5000}$. Such a space cannot be dealt with geometrically. However, the algebraic approach works well. Thus our subject is called **linear algebra**.*

For example, $\left(2, -3, 17, \pi, \sqrt{2}\right)$ 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.

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 1.13. As we will now see, addition has a simple geometric interpretation in the special case of $\mathbf{R}^2$.

Suppose we have two vectors u and v in $\mathbf{R}^2$ that we want to add. Move the vector v parallel to itself so that its initial point coincides with the end point of the vector u, as shown here. The sum $u + v$ then equals the vector whose initial point equals the initial point of u and whose end point equals the end point of the vector v, as shown here.



The sum of two vectors.

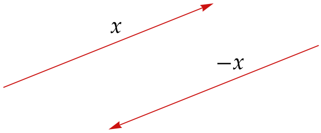
In the next definition, the 0 on the right side of the displayed equation is the list $0 \in \mathbf{F}^n$.

1.17 definition: *additive inverse in* $\mathbf{F}^n$, $-x$

For $x \in \mathbf{F}^n$, the *additive inverse* of x, denoted by $-x$, is the vector $-x \in \mathbf{F}^n$ such that

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

Thus if $x = (x_1, \dots, x_n)$, then $-x = (-x_1, \dots, -x_n)$.



A vector and its additive inverse.

The additive inverse of a vector in $\mathbf{R}^2$ is the vector with the same length but pointing in the opposite direction. The figure here illustrates this way of thinking about the additive inverse in $\mathbf{R}^2$. As you can see, the vector labeled $-x$ has the same length as the vector labeled x but points in the opposite direction.

Having dealt with addition in $\mathbf{F}^n$, we now turn to multiplication. We could define a multiplication in $\mathbf{F}^n$ 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 $\mathbf{F}$.

1.18 definition: *scalar multiplication in* $\mathbf{F}^n$

The *product* of a number λ and a vector in $\mathbf{F}^n$ is computed by multiplying each coordinate of the vector by λ:

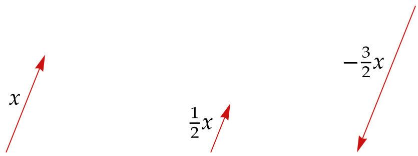
$$\lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n);$$

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

Scalar multiplication in $\mathbf{F}^n$ *multiplies together a scalar and a vector, getting a vector. In contrast, the dot product in* $\mathbf{R}^2$ *or* $\mathbf{R}^3$ *multiplies together two vectors and gets a scalar. Generalizations of the dot product will become important in Chapter 6.*

Scalar multiplication has a nice geometric interpretation in $\mathbf{R}^2$. If $\lambda > 0$ and $x \in \mathbf{R}^2$, then λx is the vector that points in the same direction as x and whose length is λ times the length of x. In other words, to get λx, we shrink or stretch x by a factor of λ, depending on whether $\lambda < 1$ or $\lambda > 1$.

If $\lambda < 0$ and $x \in \mathbf{R}^2$, then λx is the vector that points in the direction opposite to that of x and whose length is $|\lambda|$ times the length of x, as shown here.



Scalar multiplication.

Digression on Fields

A *field* is a set containing at least two distinct elements called 0 and 1, along with operations of addition and multiplication satisfying all properties listed in 1.3. Thus **R** and **C** are fields, as is the set of rational numbers along with the usual operations of addition and multiplication. Another example of a field is the set $\{0, 1\}$ with the usual operations of addition and multiplication except that $1 + 1$ is defined to equal 0.

In this book we will not deal with fields other than **R** and **C**. However, many of the definitions, theorems, and proofs in linear algebra that work for the fields **R** and **C** also work without change for arbitrary fields. If you prefer to do so, throughout much of this book (except for Chapters 6 and 7, which deal with inner product spaces) you can think of **F** as denoting an arbitrary field instead of **R** or **C**. For results (except in the inner product chapters) that have as a hypothesis that **F** is **C**, you can probably replace that hypothesis with the hypothesis that **F** is an algebraically closed field, which means that every nonconstant polynomial with coefficients in **F** has a zero. A few results, such as Exercise 13 in Section 1C, require the hypothesis on **F** that $1 + 1 \neq 0$.

Exercises 1A

1 Show that $\alpha + \beta = \beta + \alpha$ for all $\alpha, \beta \in \mathbf{C}$.

2 Show that $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

3 Show that $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

4 Show that $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$.

5 Show that for every $\alpha \in \mathbf{C}$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha + \beta = 0$.

6 Show that for every $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha\beta = 1$.

7 Show that

$$\frac{-1 + \sqrt{3}\,i}{2}$$

is a cube root of 1 (meaning that its cube equals 1).

8 Find two distinct square roots of i.

9 Find $x \in \mathbf{R}^4$ such that

$$(4, -3, 1, 7) + 2x = (5, 9, -6, 8).$$

10 Explain why there does not exist $\lambda \in \mathbf{C}$ such that

$$\lambda(2 - 3i, 5 + 4i, -6 + 7i) = (12 - 5i, 7 + 22i, -32 - 9i).$$

11 Show that $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbf{F}^n$.

12 Show that $(ab)x = a(bx)$ for all $x \in \mathbf{F}^n$ and all $a, b \in \mathbf{F}$.

13 Show that $1x = x$ for all $x \in \mathbf{F}^n$.

14 Show that $\lambda(x + y) = \lambda x + \lambda y$ for all $\lambda \in \mathbf{F}$ and all $x, y \in \mathbf{F}^n$.

15 Show that $(a + b)x = ax + bx$ for all $a, b \in \mathbf{F}$ and all $x \in \mathbf{F}^n$.

"Can you do addition?" the White Queen asked. "What's one and one and one and one and one and one and one and one and one and one?"
"I don't know," said Alice. "I lost count."

—*Through the Looking Glass*, Lewis Carroll