Shifrin
Adams

LINEAR ALGEBRA
A GEOMETRIC APPROACH

second edition

Theodore Shifrin
Malcolm R. Adams

LINEAR ALGEBRA
A GEOMETRIC APPROACH

AR ALGEBRA
TRIC APPROACH second edition
frin, Malcolm R. Adams, University of Georgia

on of *Linear Algebra: A* Second Edition introduces ector algebra to do a bit the first section and dot Other important features ude an emphasis on s and proof reasoning and the text that discuss logic, problem-solving strategies.

he *Bibliotheque* ne of the largest, most st modern libraries in the

plays on the artistic Although the lines of the y appear to focus in the d computer graphics are of basic linear algebra n learn more in the last

New to this Edition

- **Reorganization.** Topics are reorganized within each chapter based on reviewer and user feedback. Examples include treatment of the Change-of-Basis Theorem in Chapter 4 and geometric interpretation of the determinant in Chapter 5.
- **New sections.** Chapter 2 contains new sections covering linear transformations and elementary matrices making the introduction of linear transformations at once more detailed and more accessible.
- **Added examples.** Most notably, Chapter 1 includes several new proof reasoning and biological examples.
- **Expanded exercise sets.** Key concepts are reinforced through an increased number of exercises, giving students additional practice in computation.
- **Completely revised design.** The new contemporary design emphasizes the pedagogical features so students will find the text more accessible and easier to read.

Refer to the Preface for more information on what's new to this edition.

any
York, NY 10010
RG21 6XS, England

ARAGOZ

ISBN-13: 978-1-4292-1521-3
ISBN-10: 1-4292-1521-6
90000
9 781429 215213

case of proof exercises, solutions (some more detailed than others) at the back of the book. Resist as long as possible the temptation to refer to the solutions! Try to be sure you've worked the problem correctly before you glance at the answer. Be careful: Some solutions in the book are not complete, so it is your responsiblity to fill in the details. The problems that are marked with a sharp ($^\sharp$) are not necessarily particularly difficult, but they generally involve concepts and results to which we shall refer later in the text. Thus, if your instructor assigns them, you should make sure you understand how to do them. Occasional exercises are quite challenging, and we hope you will work hard on a few; we firmly believe that only by struggling with a real puzzler do we all progress as mathematicians.

Once again, we hope you will have fun as you embark on your voyage to learn linear algebra. Please let us know if there are parts of the book you find particularly enjoyable or troublesome.

CHAPTER 1

VECTORS AND MATRICES

Linear algebra provides a beautiful example of the interplay between two branches of mathematics: geometry and algebra. We begin this chapter with the geometric concepts and algebraic representations of points, lines, and planes in the more familiar setting of two and three dimensions ($\mathbb{R}^2$ and $\mathbb{R}^3$, respectively) and then generalize to the "n-dimensional" space $\mathbb{R}^n$. We come across two ways of describing (hyper)planes—either parametrically or as solutions of a Cartesian equation. Going back and forth between these two formulations will be a major theme of this text. The fundamental tool that is used in bridging these descriptions is Gaussian elimination, a standard algorithm used to solve systems of linear equations. As we shall see, it also has significant consequences in the theory of systems of equations. We close the chapter with a variety of applications, some not of a geometric nature.

1 Vectors

1.1 Vectors in $\mathbb{R}^2$

Throughout our work the symbol $\mathbb{R}$ denotes the set of real numbers. We define a *vector*[1] in $\mathbb{R}^2$ to be an ordered pair of real numbers, $\mathbf{x} = (x_1, x_2)$. This is the *algebraic* representation of the vector $\mathbf{x}$. Thanks to Descartes, we can identify the ordered pair (x_1, x_2) with a point in the Cartesian plane, $\mathbb{R}^2$. The relationship of this point to the origin $(0, 0)$ gives rise to the *geometric* interpretation of the vector $\mathbf{x}$—namely, the arrow pointing from $(0, 0)$ to (x_1, x_2), as illustrated in Figure 1.1.

The vector $\mathbf{x}$ has *length* and *direction*. The length of $\mathbf{x}$ is denoted $\|\mathbf{x}\|$ and is given by

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2},$$

whereas its *direction* can be specified, say, by the angle the arrow makes with the positive x_1-axis. We denote the zero vector $(0, 0)$ by $\mathbf{0}$ and agree that it has no direction. We say two vectors are *equal* if they have the same coordinates, or, equivalently, if they have the same length and direction.

More generally, any two points A and B in the plane determine a directed line segment from A to B, denoted $\overrightarrow{AB}$. This can be visualized as an arrow with A as its "tail" and B as its "head." If $A = (a_1, a_2)$ and $B = (b_1, b_2)$, then the arrow $\overrightarrow{AB}$ has the same length

[1]The word derives from the Latin *vector*, "carrier," from *vectus*, the past participle of *vehere*, "to carry."

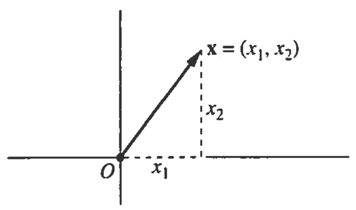

FIGURE 1.1

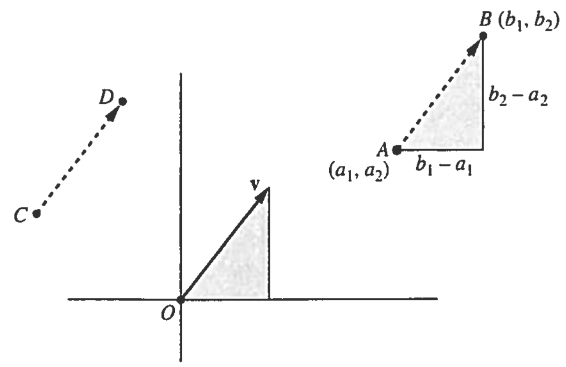

FIGURE 1.2

and direction as the vector $\mathbf{v} = (b_1 - a_1, b_2 - a_2)$. For algebraic purposes, a vector should always have its tail at the origin, but for geometric and physical applications, it is important to be able to "translate" it—to move it parallel to itself so that its tail is elsewhere. Thus, at least geometrically, we think of the arrow $\overrightarrow{AB}$ as the same thing as the vector $\mathbf{v}$. In the same vein, if $C = (c_1, c_2)$ and $D = (d_1, d_2)$, then, as indicated in Figure 1.2, the vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are equal if $(b_1 - a_1, b_2 - a_2) = (d_1 - c_1, d_2 - c_2)$.[2] This is often a bit confusing at first, so for a while we shall use dotted lines in our diagrams to denote the vectors whose tails are not at the origin.

Scalar multiplication

If c is a real number and $\mathbf{x} = (x_1, x_2)$ is a vector, then we define $c\mathbf{x}$ to be the vector with coordinates (cx_1, cx_2). Now the length of $c\mathbf{x}$ is

$$\|c\mathbf{x}\| = \sqrt{(cx_1)^2 + (cx_2)^2} = \sqrt{c^2(x_1^2 + x_2^2)} = |c|\sqrt{x_1^2 + x_2^2} = |c|\|\mathbf{x}\|.$$

When $c \neq 0$, the direction of $c\mathbf{x}$ is either exactly the same as or exactly opposite that of $\mathbf{x}$, depending on the sign of c. Thus multiplication by the real number c simply stretches (or shrinks) the vector by a factor of $|c|$ and reverses its direction when c is negative, as shown in Figure 1.3. Because this is a geometric "change of scale," we refer to the real number c as a *scalar* and to the multiplication $c\mathbf{x}$ as *scalar multiplication*.

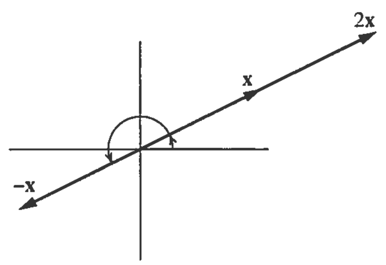

FIGURE 1.3

> *Definition.* A vector $\mathbf{x}$ is called a *unit vector* if it has length 1, i.e., if $\|\mathbf{x}\| = 1$.

[2]The sophisticated reader may recognize that we have defined an *equivalence relation* on the collection of directed line segments. A vector can then officially be interpreted as an *equivalence class* of directed line segments.

Note that whenever $\mathbf{x} \neq \mathbf{0}$, we can find a unit vector with the same direction by taking

$$\frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{1}{\|\mathbf{x}\|}\mathbf{x},$$

as shown in Figure 1.4.

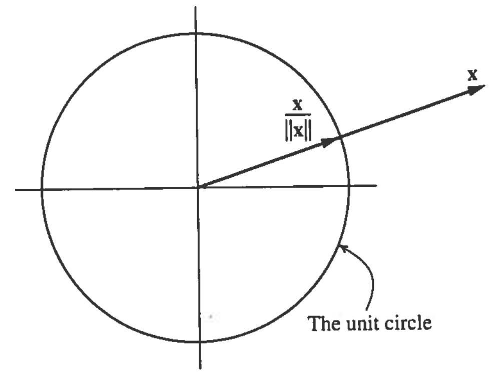

FIGURE 1.4

EXAMPLE 1

The vector $\mathbf{x} = (1, -2)$ has length $\|\mathbf{x}\| = \sqrt{1^2 + (-2)^2} = \sqrt{5}$. Thus, the vector

$$\mathbf{u} = \frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{1}{\sqrt{5}}(1, -2)$$

is a unit vector in the same direction as $\mathbf{x}$. As a check, $\|\mathbf{u}\|^2 = \left(\frac{1}{\sqrt{5}}\right)^2 + \left(\frac{-2}{\sqrt{5}}\right)^2 = \frac{1}{5} + \frac{4}{5} = 1$. ▲

Given a nonzero vector $\mathbf{x}$, any scalar multiple $c\mathbf{x}$ lies on the line that passes through the origin and the head of the vector $\mathbf{x}$. For this reason, we make the following definition.

> *Definition.* We say two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ are *parallel* if one vector is a scalar multiple of the other, i.e., if there is a scalar c such that $\mathbf{y} = c\mathbf{x}$. We say two nonzero vectors are *nonparallel* if they are not parallel. (Notice that when one of the vectors is $\mathbf{0}$, they are not considered to be either parallel or nonparallel.)

Vector addition

If $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$, then we define

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2).$$

Because addition of real numbers is commutative, it follows immediately that vector addition is commutative:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}.$$

(See Exercise 28 for an exhaustive list of the properties of vector addition and scalar multiplication.) To understand this geometrically, we move the vector $\mathbf{y}$ so that its tail is at the head of $\mathbf{x}$ and draw the arrow from the origin to the head of the shifted vector $\mathbf{y}$, as shown in Figure 1.5. This is called the *parallelogram law* for vector addition, for, as we see in Figure 1.5, $\mathbf{x}+\mathbf{y}$ is the "long" diagonal of the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$. The symmetry of the parallelogram illustrates the commutative law $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$.

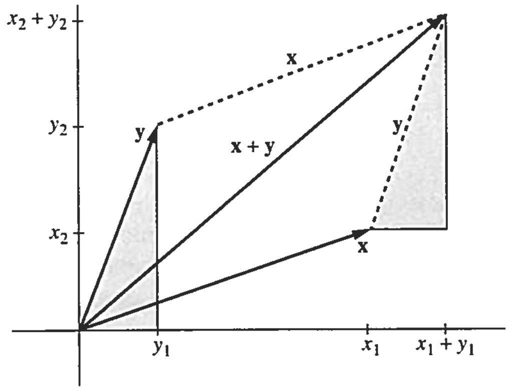

FIGURE 1.5

This would be a good place for the diligent student to grab paper and pencil and make up some numerical examples. Pick a few vectors $\mathbf{x}$ and $\mathbf{y}$, calculate their sums algebraically, and then verify your answers by making sketches to scale.

Remark. We emphasize here that the notions of vector addition and scalar multiplication make sense geometrically for vectors that do not necessarily have their tails at the origin. If we wish to add $\overrightarrow{CD}$ to $\overrightarrow{AB}$, we simply recall that $\overrightarrow{CD}$ is equal to *any* vector with the same length and direction, so we just translate $\overrightarrow{CD}$ so that C and B coincide; then the arrow from A to the point D in its new position is the sum $\overrightarrow{AB}+\overrightarrow{CD}$.

Vector subtraction

Subtraction of one vector from another is also easy to define algebraically. If $\mathbf{x}=(x_1,x_2)$ and $\mathbf{y}=(y_1,y_2)$, then we set

$$\mathbf{x}-\mathbf{y}=(x_1-y_1,x_2-y_2).$$

As is the case with real numbers, we have the following important interpretation of the difference: $\mathbf{x}-\mathbf{y}$ is the vector we must add to $\mathbf{y}$ in order to obtain $\mathbf{x}$; that is,

$$(\mathbf{x}-\mathbf{y})+\mathbf{y}=\mathbf{x}.$$

From this interpretation we can understand $\mathbf{x}-\mathbf{y}$ geometrically. The arrow representing it has its tail at (the head of) $\mathbf{y}$ and its head at (the head of) $\mathbf{x}$; when we add the resulting vector to $\mathbf{y}$, we do in fact get $\mathbf{x}$. As shown in Figure 1.6, this results in the other diagonal of the parallelogram determined by $\mathbf{x}$ and $\mathbf{y}$. Of course, we can also think of $\mathbf{x}-\mathbf{y}$ as the sum $\mathbf{x}+(-\mathbf{y})=\mathbf{x}+(-1)\mathbf{y}$, as pictured in Figure 1.7. Note that if A and B are points in the plane and O denotes the origin, then setting $\mathbf{x}=\overrightarrow{OB}$ and $\mathbf{y}=\overrightarrow{OA}$ gives $\mathbf{x}-\mathbf{y}=\overrightarrow{AB}$.

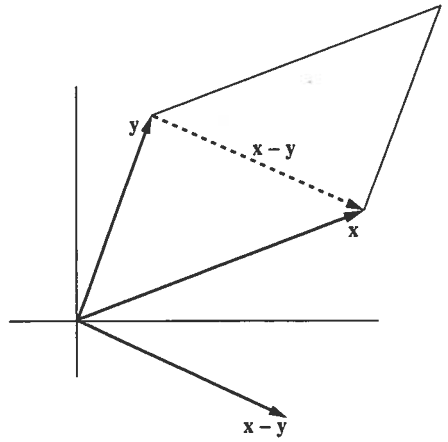

FIGURE 1.6

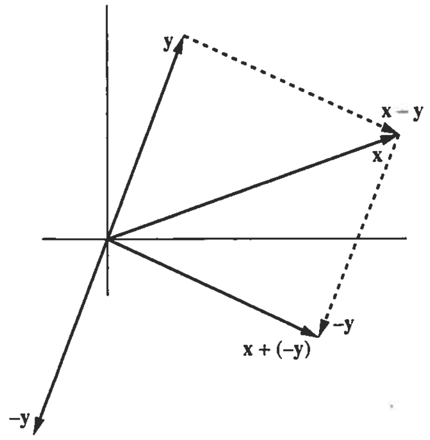

FIGURE 1.7

EXAMPLE 2

Let A and B be points in the plane. The *midpoint* M of the line segment $\overline{AB}$ is the unique point in the plane with the property that $\overrightarrow{AM}=\overrightarrow{MB}$. Since $\overrightarrow{AB}=\overrightarrow{AM}+\overrightarrow{MB}=2\overrightarrow{AM}$, we infer that $\overrightarrow{AM}=\frac{1}{2}\overrightarrow{AB}$. (See Figure 1.8.) What's more, we can find the vector $\mathbf{v}=\overrightarrow{OM}$, whose tail is at the origin and whose head is at M, as follows. As above, we set $\mathbf{x}=\overrightarrow{OB}$ and $\mathbf{y}=\overrightarrow{OA}$, so $\overrightarrow{AB}=\mathbf{x}-\mathbf{y}$ and $\overrightarrow{AM}=\frac{1}{2}\overrightarrow{AB}=\frac{1}{2}(\mathbf{x}-\mathbf{y})$. Then we have

$$\begin{aligned}\overrightarrow{OM}&=\overrightarrow{OA}+\overrightarrow{AM}\\&=\mathbf{y}+\tfrac{1}{2}(\mathbf{x}-\mathbf{y})\\&=\mathbf{y}+\tfrac{1}{2}\mathbf{x}-\tfrac{1}{2}\mathbf{y}\\&=\tfrac{1}{2}\mathbf{x}+\tfrac{1}{2}\mathbf{y}=\tfrac{1}{2}(\mathbf{x}+\mathbf{y}).\end{aligned}$$

In particular, the vector $\overrightarrow{OM}$ is the average of the vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$.

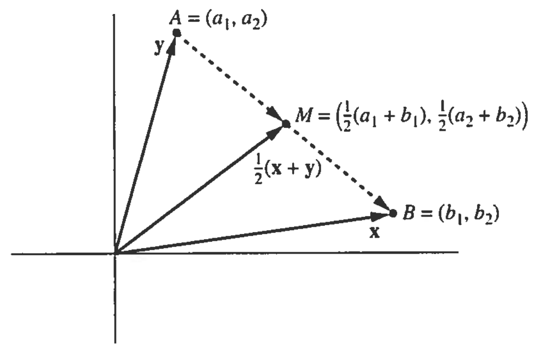

FIGURE 1.8

In coordinates, if $A=(a_1,a_2)$ and $B=(b_1,b_2)$, then the coordinates of M are the average of the respective coordinates of A and B:

$$M=\tfrac{1}{2}\big((a_1,a_2)+(b_1,b_2)\big)=\big(\tfrac{1}{2}(a_1+b_1),\tfrac{1}{2}(a_2+b_2)\big).$$

See Exercise 18 for a generalization to three vectors. ▲

We now use the result of Example 2 to derive one of the classic results from high school geometry.

Proposition 1.1. *The diagonals of a parallelogram bisect one another.*

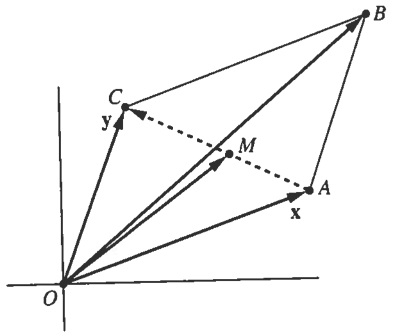

FIGURE 1.9

Proof. The strategy is this: We will find vector expressions for the midpoint of each diagonal and deduce from these expressions that these two midpoints coincide. We may assume one vertex of the parallelogram is at the origin, O, and we label the remaining vertices A, B, and C, as shown in Figure 1.9. Let $\mathbf{x} = \overrightarrow{OA}$ and $\mathbf{y} = \overrightarrow{OC}$, and let M be the midpoint of diagonal $\overline{AC}$. (In the picture, we do not place M on diagonal $\overline{OB}$, even though ultimately we will show that it is on $\overline{OB}$.) We have shown in Example 2 that

$$\overrightarrow{OM} = \tfrac{1}{2}(\mathbf{x}+\mathbf{y}).$$

Next, note that $\overrightarrow{OB} = \mathbf{x}+\mathbf{y}$ by our earlier discussion of vector addition, and so

$$\overrightarrow{ON} = \tfrac{1}{2}\overrightarrow{OB} = \tfrac{1}{2}(\mathbf{x}+\mathbf{y}) = \overrightarrow{OM}.$$

This implies that $M = N$, and so the point M is the midpoint of both diagonals. That is, the two diagonals bisect one another. ☐

> Here is some basic advice in using vectors to prove a geometric statement in $\mathbb{R}^2$. Set up an appropriate diagram and pick two convenient nonparallel vectors that arise naturally in the diagram; call these $\mathbf{x}$ and $\mathbf{y}$, and then express all other relevant quantities in terms of *only* $\mathbf{x}$ and $\mathbf{y}$.

It should now be evident that vector methods provide a great tool for translating theorems from Euclidean geometry into simple algebraic statements. Here is another example. Recall that a *median* of a triangle is a line segment from a vertex to the midpoint of the opposite side.

Proposition 1.2. *The medians of a triangle intersect at a point that is two-thirds of the way from each vertex to the opposite side.*

Proof. We may put one of the vertices of the triangle at the origin, O, so that the picture is as shown at the left in Figure 1.10: Let $\mathbf{x} = \overrightarrow{OA}$, $\mathbf{y} = \overrightarrow{OB}$, and let L, M, and N be the midpoints of $\overline{OA}$, $\overline{AB}$, and $\overline{OB}$, respectively. The battle plan is the following: We let P denote the point two-thirds of the way from B to L, Q the point two-thirds of the way from O to M, and R the point two-thirds of the way from A to N. Although we've indicated P, Q, and R as distinct points at the right in Figure 1.10, our goal is to prove that $P = Q = R$; we do this by expressing all the vectors $\overrightarrow{OP}$, $\overrightarrow{OQ}$, and $\overrightarrow{OR}$ in terms of $\mathbf{x}$ and $\mathbf{y}$. For instance, since $\overrightarrow{OB} = \mathbf{y}$ and $\overrightarrow{OL} = \frac{1}{2}\overrightarrow{OA} = \frac{1}{2}\mathbf{x}$, we get $\overrightarrow{BL} = \frac{1}{2}\mathbf{x} - \mathbf{y}$, and so

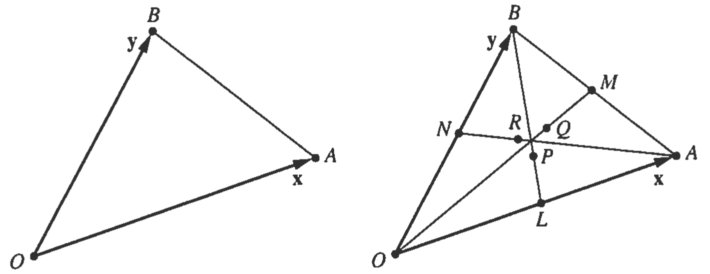

FIGURE 1.10

$$\begin{aligned}\overrightarrow{OP} &= \overrightarrow{OB} + \overrightarrow{BP} = \overrightarrow{OB} + \tfrac{2}{3}\overrightarrow{BL} = \mathbf{y} + \tfrac{2}{3}\left(\tfrac{1}{2}\mathbf{x}-\mathbf{y}\right)\\ &= \tfrac{1}{3}\mathbf{x} + \tfrac{1}{3}\mathbf{y}.\end{aligned}$$

Similarly,

$$\begin{aligned}\overrightarrow{OQ} &= \tfrac{2}{3}\overrightarrow{OM} = \tfrac{2}{3}\left(\tfrac{1}{2}(\mathbf{x}+\mathbf{y})\right) = \tfrac{1}{3}(\mathbf{x}+\mathbf{y});\ \text{and}\\ \overrightarrow{OR} &= \overrightarrow{OA} + \overrightarrow{AR} = \overrightarrow{OA} + \tfrac{2}{3}\overrightarrow{AN} = \mathbf{x} + \tfrac{2}{3}\left(\tfrac{1}{2}\mathbf{y}-\mathbf{x}\right) = \tfrac{1}{3}\mathbf{x} + \tfrac{1}{3}\mathbf{y}.\end{aligned}$$

We conclude that, as desired, $\overrightarrow{OP} = \overrightarrow{OQ} = \overrightarrow{OR}$, and so $P = Q = R$. That is, if we go two-thirds of the way down any of the medians, we end up at the same point; this is, of course, the point of intersection of the three medians. ☐

The astute reader might notice that we could have been more economical in the last proof. Suppose we merely check that the points two-thirds of the way down *two* of the medians (say, P and Q) agree. It would then follow (say, by relabeling the triangle slightly) that the same is true of a different pair of medians (say, P and R). But since any two pairs must have this point in common, we may now conclude that all three points are equal.

1.2 Lines

With these algebraic tools in hand, we now study lines[3] in $\mathbb{R}^2$. A line ℓ_0 through the origin with a given nonzero *direction vector* $\mathbf{v}$ consists of all points of the form $\mathbf{x} = t\mathbf{v}$ for some scalar t. The line ℓ parallel to ℓ_0 and passing through the point P is obtained by translating ℓ_0 by the vector $\mathbf{x}_0 = \overrightarrow{OP}$; that is, the line ℓ through P with direction $\mathbf{v}$ consists of all points of the form

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$$

as t varies over the real numbers. (It is important to remember that, geometrically, points of the line are the *heads* of the vectors $\mathbf{x}$.) It is compelling to think of t as a time *parameter*; initially (i.e., at time $t = 0$), the point starts at $\mathbf{x}_0$ and moves in the direction of $\mathbf{v}$ as time increases. For this reason, this is often called the *parametric equation* of the line.

To describe the line determined by two distinct points P and Q, we pick $\mathbf{x}_0 = \overrightarrow{OP}$ as before and set $\mathbf{y}_0 = \overrightarrow{OQ}$; we obtain a direction vector by taking

$$\mathbf{v} = \overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} = \mathbf{y}_0 - \mathbf{x}_0.$$

[3]Note: In mathematics, the word *line* is reserved for "straight" lines, and the curvy ones are usually called curves.

Thus, as indicated in Figure 1.11, any point on the line through P and Q can be expressed in the form

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} = \mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0) = (1-t)\mathbf{x}_0 + t\mathbf{y}_0.$$

As a check, when $t = 0$ and $t = 1$, we recover the points P and Q, respectively.

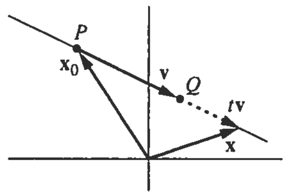

FIGURE 1.11

EXAMPLE 3

Consider the line

$$x_2 = 3x_1 + 1$$

(the usual *Cartesian equation* from high school algebra). We wish to write it in parametric form. Well, any point (x_1, x_2) lying on the line is of the form

$$\mathbf{x} = (x_1, x_2) = (x_1, 3x_1 + 1) = (0, 1) + (x_1, 3x_1) = (0, 1) + x_1(1, 3).$$

Since x_1 can have any real value, we may rename it t, and then, rewriting the equation as

$$\mathbf{x} = (0, 1) + t(1, 3),$$

we recognize this as the equation of the line through the point $P = (0, 1)$ with direction vector $\mathbf{v} = (1, 3)$.

Notice that we might have given alternative parametric equations for this line. The equations

$$\mathbf{x} = (0, 1) + s(2, 6) \qquad \text{and} \qquad \mathbf{x} = (1, 4) + u(1, 3)$$

also describe this same line. Why? ▲

The "Why?" is a sign that, once again, the reader should take pencil in hand and check that our assertion is correct.

EXAMPLE 4

Consider the line ℓ given in parametric form by

$$\mathbf{x} = (-1, 1) + t(2, 3)$$

and pictured in Figure 1.12. We wish to find a Cartesian equation of the line. Note that ℓ passes through the point $(-1, 1)$ and has direction vector $(2, 3)$. The direction vector determines the slope of the line:

$$\frac{\text{rise}}{\text{run}} = \frac{3}{2},$$

so, using the point-slope form of the equation of a line, we find

$$\frac{x_2 - 1}{x_1 + 1} = \frac{3}{2}; \quad \text{i.e.,} \quad x_2 = \frac{3}{2}x_1 + \frac{5}{2}.$$

Of course, we can rewrite this as $3x_1 - 2x_2 = -5$. ▲

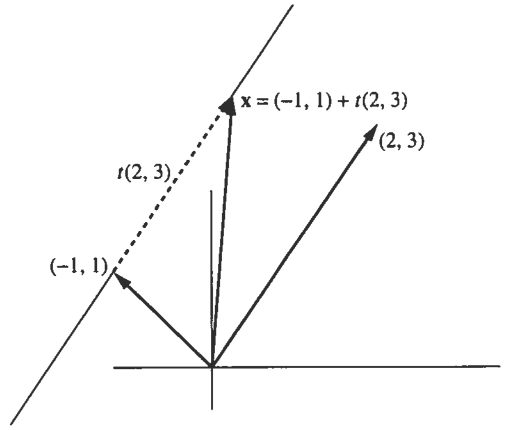

FIGURE 1.12

Mathematics is built around *sets* and relations among them. Although the precise definition of a set is surprisingly subtle, we will adopt the naïve approach that sets are just collections of objects (mathematical or not). The sets with which we shall be concerned in this text consist of vectors. In general, the objects belonging to a set are called its *elements* or *members*. If X is a set and x is an element of X, we write this as

$$x \in X.$$

We might also read the phrase "$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$" as "$\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$" or "$\mathbf{x}$ and $\mathbf{y}$ belong to $\mathbb{R}^n$."

We think of a line in $\mathbb{R}^2$ as the set of points (or vectors) with a certain property. The official notation for the parametric representation is

$$\ell = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} = (3, 0) + t(-2, 1) \text{ for some scalar } t\}.$$

Or we might describe ℓ by its Cartesian equation:

$$\ell = \{\mathbf{x} \in \mathbb{R}^2 : x_1 + 2x_2 = 3\}.$$

In words, this says that "ℓ is the set of points $\mathbf{x}$ in $\mathbb{R}^2$ such that $x_1 + 2x_2 = 3$."

Often in the text we are sloppy and speak of *the line*

$$(*) \qquad x_1 + 2x_2 = 3$$

rather than using the set notation or saying, more properly, *the line whose equation is* $(*)$.

1.3 On to $\mathbb{R}^n$

The generalizations to $\mathbb{R}^3$ and $\mathbb{R}^n$ are now quite straightforward. A vector $\mathbf{x} \in \mathbb{R}^3$ is defined to be an ordered triple of numbers (x_1, x_2, x_3), which in turn has a geometric interpretation as an arrow from the origin to the point in three-dimensional space with those Cartesian coordinates. Although our geometric intuition becomes hazy when we move to $\mathbb{R}^n$ with $n > 3$, we may still use the algebraic description of a point in n-space as an ordered n-tuple of real numbers $(x_1, x_2, \ldots, x_n)$. Thus, we write $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ for a vector in n-space. We define $\mathbb{R}^n$ to be the collection of all vectors $(x_1, x_2, \ldots, x_n)$ as $x_1, x_2, \ldots, x_n$ vary over $\mathbb{R}$. As we did in $\mathbb{R}^2$, given two points $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n) \in \mathbb{R}^n$, we associate to the directed line segment from A to B the vector $\overrightarrow{AB} = (b_1 - a_1, \ldots, b_n - a_n)$.

Remark. The beginning linear algebra student may wonder why anyone would care about $\mathbb{R}^n$ with $n > 3$. We hope that the rich structure we're going to study in this text will eventually be satisfying in and of itself. But some will be happier to know that "real-world applications" force the issue, because many applied problems require understanding the interactions of a large number of variables. For instance, to model the motion of a single particle in $\mathbb{R}^3$, we must know the three variables describing its position *and* the three variables describing its velocity, for a total of six variables. Other examples arise in economic models of a large number of industries, each of which has a supply-demand equation involving large numbers of variables, and in population models describing the interaction of large numbers of different species. In these multivariable problems, each variable accounts for one copy of $\mathbb{R}$, and so an n-variable problem naturally leads to linear (and nonlinear) problems in $\mathbb{R}^n$.

Length, scalar multiplication, and vector addition are defined algebraically in an analogous fashion: If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we define

1. $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$;
2. $c\mathbf{x} = (cx_1, cx_2, \ldots, cx_n)$;
3. $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$.

As before, scalar multiplication stretches (or shrinks or reverses) vectors, and vector addition is given by the parallelogram law. Our notion of length in $\mathbb{R}^n$ is consistent with applying the Pythagorean Theorem (or distance formula); for example, as Figure 1.13 shows, we find the length of $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ by first finding the length of the hypotenuse in the x_1x_2-plane and then using that hypotenuse as one leg of the right triangle with hypotenuse $\mathbf{x}$:

$$\|\mathbf{x}\|^2 = \left(\sqrt{x_1^2 + x_2^2}\right)^2 + x_3^2 = x_1^2 + x_2^2 + x_3^2.$$

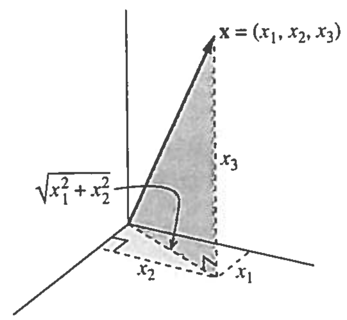

FIGURE 1.13

The *parametric* description of a line ℓ in $\mathbb{R}^n$ is exactly the same as in $\mathbb{R}^2$: If $\mathbf{x}_0 \in \mathbb{R}^n$ is a point on the line and the nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is the direction vector of the line, then points on the line are given by

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}, \quad t \in \mathbb{R}.$$

More formally, we write this as

$$\ell = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{x}_0 + t\mathbf{v} \text{ for some } t \in \mathbb{R}\}.$$

As we've already seen, two points determine a line; three or more points in $\mathbb{R}^n$ are called *collinear* if they lie on some line; they are called *noncollinear* if they do not lie on any line.

EXAMPLE 5

Consider the line determined by the points $P = (1, 2, 3)$ and $Q = (2, 1, 5)$ in $\mathbb{R}^3$. The direction vector of the line is $\mathbf{v} = \overrightarrow{PQ} = (2, 1, 5) - (1, 2, 3) = (1, -1, 2)$, and we get an initial point $\mathbf{x}_0 = \overrightarrow{OP}$, just as we did in $\mathbb{R}^2$. We now visualize Figure 1.11 as being in $\mathbb{R}^3$ and see that the general point on this line is $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} = (1, 2, 3) + t(1, -1, 2)$. ▲

The definition of *parallel* and *nonparallel* vectors in $\mathbb{R}^n$ is identical to that in $\mathbb{R}^2$. Two nonparallel vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$ determine a *plane*, $\mathcal{P}_0$, through the origin, as follows. $\mathcal{P}_0$ consists of all points of the form

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v}$$

as s and t vary over $\mathbb{R}$. Note that for fixed s, as t varies, the point moves along a line with direction vector $\mathbf{v}$; changing s gives a family of parallel lines. On the other hand, a general plane is determined by one point $\mathbf{x}_0$ and two nonparallel direction vectors $\mathbf{u}$ and $\mathbf{v}$. The plane $\mathcal{P}$ *spanned by* $\mathbf{u}$ *and* $\mathbf{v}$ *and passing through the point* $\mathbf{x}_0$ consists of all points $\mathbf{x} \in \mathbb{R}^3$ of the form

$$\mathbf{x} = \mathbf{x}_0 + s\mathbf{u} + t\mathbf{v}$$

as s and t vary over $\mathbb{R}$, as pictured in Figure 1.14. We can obtain the plane $\mathcal{P}$ by translating $\mathcal{P}_0$, the plane parallel to $\mathcal{P}$ and passing through the origin, by the vector $\mathbf{x}_0$. (Note that this *parametric* description of a plane in $\mathbb{R}^3$ makes perfect sense in n-space for any $n \geq 3$.)

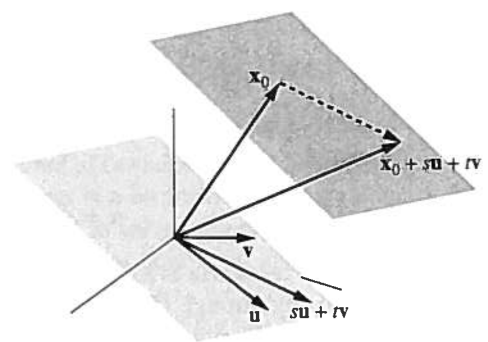

FIGURE 1.14

Before doing some examples, we define two terms that will play a crucial role throughout our study of linear algebra.

> *Definition.* Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$. If $c_1, \ldots, c_k \in \mathbb{R}$, the vector
> $$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$
> is called a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. (See Figure 1.15.)

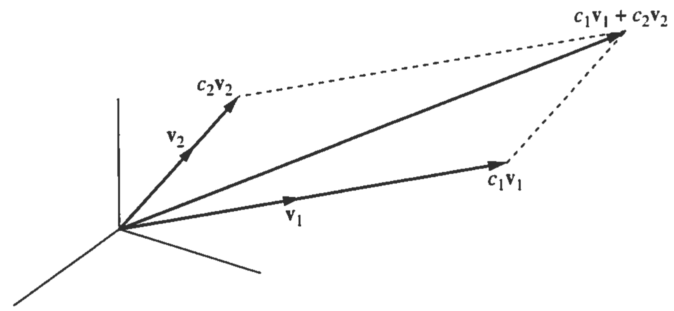

FIGURE 1.15

Definition. Let $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$. The set of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is called their *span*, denoted Span $(\mathbf{v}_1, \dots, \mathbf{v}_k)$. That is,

$$\text{Span}\,(\mathbf{v}_1, \dots, \mathbf{v}_k) = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k \text{ for some scalars } c_1, \dots, c_k\}.$$

In terms of our new language, then, the span of two nonparallel vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is a plane through the origin. (What happens if $\mathbf{u}$ and $\mathbf{v}$ are parallel? We will return to such questions in greater generality later in the text.)

EXAMPLE 6

Consider the points $\mathbf{x} \in \mathbb{R}^3$ that satisfy the *Cartesian* equation

$$(\dagger) \qquad x_1 - 2x_2 = 5.$$

The set of points $(x_1, x_2) \in \mathbb{R}^2$ satisfying this equation forms a line ℓ in $\mathbb{R}^2$; since x_3 is allowed to vary arbitrarily, we obtain a vertical plane—a fence standing upon the line ℓ. Let's write it in *parametric* form: Any $\mathbf{x}$ satisfying this equation is of the form

$$\mathbf{x} = (x_1, x_2, x_3) = (5 + 2x_2, x_2, x_3) = (5, 0, 0) + x_2(2, 1, 0) + x_3(0, 0, 1).$$

Since x_2 and x_3 can be arbitrary, we rename them s and t, respectively, obtaining the equation

$$(*) \qquad \mathbf{x} = (5, 0, 0) + s(2, 1, 0) + t(0, 0, 1),$$

which we recognize as a parametric equation of the plane spanned by $(2, 1, 0)$ and $(0, 0, 1)$ and passing through $(5, 0, 0)$. Moreover, note that any $\mathbf{x}$ of this form can be written as $\mathbf{x} = (5 + 2s, s, t)$, and so $x_1 - 2x_2 = (5 + 2s) - 2s = 5$, from which we see that $\mathbf{x}$ is indeed a solution of the equation $(\dagger)$. ▲

This may be an appropriate time to emphasize a basic technique in mathematics: How do we decide when two sets are equal? First of all, we say that X is a *subset* of Y, written

$$X \subset Y,$$

if every element of X is an element of Y. That is, $X \subset Y$ means that whenever $x \in X$, it must also be the case that $x \in Y$. (Some authors write $X \subseteq Y$ to remind us that the sets X and Y may be equal.)

To prove that two sets X and Y are equal (i.e., that every element of X is an element of Y and every element of Y is an element of X), it is often easiest to show that $X \subset Y$ and $Y \subset X$. We ask the diligent reader to check how we've done this explicitly in Example 6: Identify the two sets X and Y, and decide what justifies each of the statements $X \subset Y$ and $Y \subset X$.

EXAMPLE 7

As was the case for lines, a given plane has many different parametric representations. For example,

$$(**) \qquad \mathbf{x} = (7, 1, -5) + u(2, 1, 2) + v(2, 1, 3)$$

is another description of the plane given in Example 6, as we now proceed to check. First, we ask whether every point of $(**)$ can be expressed in the form of $(*)$ for some values of s and t; that is, fixing u and v, we must find s and t so that

$$(5, 0, 0) + s(2, 1, 0) + t(0, 0, 1) = (7, 1, -5) + u(2, 1, 2) + v(2, 1, 3).$$

This gives us the system of equations

$$\begin{aligned} 2s \quad &= 2u + 2v + 2 \\ s \quad &= u + v + 1 \\ t &= 2u + 3v - 5, \end{aligned}$$

whose solution is obviously $s = u + v + 1$ and $t = 2u + 3v - 5$. Indeed, we check the algebra:

$$\begin{aligned} (5, 0, 0) + s(2, 1, 0) + t(0, 0, 1) &= (5, 0, 0) + (u + v + 1)(2, 1, 0) \\ &\quad + (2u + 3v - 5)(0, 0, 1) \\ &= \big((5, 0, 0) + (2, 1, 0) - 5(0, 0, 1)\big) \\ &\quad + u\big((2, 1, 0) + 2(0, 0, 1)\big) + v\big((2, 1, 0) + 3(0, 0, 1)\big) \\ &= (7, 1, -5) + u(2, 1, 2) + v(2, 1, 3). \end{aligned}$$

In conclusion, every point of $(**)$ does in fact lie in the plane $(*)$.

Reversing the process is a bit trickier. Given a point of the form $(*)$ for some fixed values of s and t, we need to solve the equations for u and v. We will address this sort of problem in Section 4, but for now, we'll just notice that if we take $u = 3s - t - 8$ and $v = -2s + t + 7$ in the equation $(**)$, we get the point $(*)$. Thus, every point of the plane $(*)$ lies in the plane $(**)$. This means the two planes are, in fact, identical. ▲

EXAMPLE 8

Consider the points $\mathbf{x} \in \mathbb{R}^3$ that satisfy the equation

$$x_1 - 2x_2 + x_3 = 5.$$

Any $\mathbf{x}$ satisfying this equation is of the form

$$\mathbf{x} = (x_1, x_2, x_3) = (5 + 2x_2 - x_3, x_2, x_3) = (5, 0, 0) + x_2(2, 1, 0) + x_3(-1, 0, 1).$$

So this equation describes a plane $\mathcal{P}$ spanned by $(2, 1, 0)$ and $(-1, 0, 1)$ and passing through $(5, 0, 0)$. We leave it to the reader to check the converse—that every point in the plane $\mathcal{P}$ satisfies the original Cartesian equation. ▲

In the preceding examples, we started with a Cartesian equation of a plane in $\mathbb{R}^3$ and derived a parametric formulation. Of course, planes can be described in different ways.

EXAMPLE 9

We wish to find a parametric equation of the plane that contains the points $P = (1, 2, 1)$ and $Q = (2, 4, 0)$ and is parallel to the vector $(1, 1, 3)$. We take $\mathbf{x}_0 = (1, 2, 1)$, $\mathbf{u} = \overrightarrow{PQ} = (1, 2, -1)$, and $\mathbf{v} = (1, 1, 3)$, so the plane consists of all points of the form

$$\mathbf{x} = (1, 2, 1) + s(1, 2, -1) + t(1, 1, 3), \quad s, t \in \mathbb{R}. \qquad \blacktriangle$$

Finally, note that three noncollinear points $P, Q, R \in \mathbb{R}^3$ determine a plane. To get a parametric equation of this plane, we simply take $\mathbf{x}_0 = \overrightarrow{OP}$, $\mathbf{u} = \overrightarrow{PQ}$, and $\mathbf{v} = \overrightarrow{PR}$. We should observe that if P, Q, and R are noncollinear, then $\mathbf{u}$ and $\mathbf{v}$ are nonparallel (why?).

It is also a reasonable question to ask whether a specific point lies on a given plane.

EXAMPLE 10

Let $\mathbf{u} = (1, 1, 0, -1)$ and $\mathbf{v} = (2, 0, 1, 1)$. We ask whether the vector $\mathbf{x} = (1, 3, -1, -2)$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. That is, are there scalars s and t so that $s\mathbf{u} + t\mathbf{v} = \mathbf{x}$, i.e.,

$$s(1, 1, 0, -1) + t(2, 0, 1, 1) = (1, 3, -1, -2)?$$

Expanding, we have

$$(s + 2t, s, t, -s + t) = (1, 3, -1, -2),$$

which leads to the system of equations

$$\begin{aligned} s + 2t &= 1 \\ s &= 3 \\ t &= -1 \\ -s + t &= -2. \end{aligned}$$

From the second and third equations we infer that $s = 3$ and $t = -1$. These values also satisfy the first equation, but *not* the fourth, and so the system of equations has no solution; that is, there are no values of s and t for which *all* the equations hold. Thus, $\mathbf{x}$ is not a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Geometrically, this means that the vector $\mathbf{x}$ does not lie in the plane spanned by $\mathbf{u}$ and $\mathbf{v}$ and passing through the origin. We will learn a systematic way of solving such systems of linear equations in Section 4. ▲

EXAMPLE 11

Suppose that the nonzero vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are given in $\mathbb{R}^3$ and, moreover, that $\mathbf{v}$ and $\mathbf{w}$ are nonparallel. Consider the line ℓ given parametrically by $\mathbf{x} = \mathbf{x}_0 + r\mathbf{u}$ ($r \in \mathbb{R}$) and the plane $\mathcal{P}$ given parametrically by $\mathbf{x} = \mathbf{x}_1 + s\mathbf{v} + t\mathbf{w}$ ($s, t, \in \mathbb{R}$). Under what conditions do ℓ and $\mathcal{P}$ intersect?

> It is a good habit to begin by drawing a sketch to develop some intuition for what the problem is about (see Figure 1.16). We must start by translating the hypothesis that the line and plane have (at least) one point in common into a precise statement involving the parametric equations of the line and plane; our sentence should begin with something like "For some particular values of the real numbers r, s, and t, we have the equation ...."

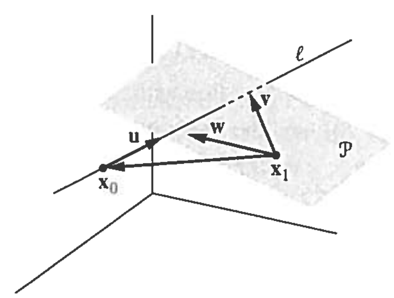

FIGURE 1.16

For ℓ and $\mathcal{P}$ to have (at least) one point $\mathbf{x}^*$ in common, that point must be represented in the form $\mathbf{x}^* = \mathbf{x}_0 + r\mathbf{u}$ for some value of r and, likewise, in the form $\mathbf{x}^* = \mathbf{x}_1 + s\mathbf{v} + t\mathbf{w}$ for some values of s and t. Setting these two expressions for $\mathbf{x}^*$ equal, we have

$$\mathbf{x}_0 + r\mathbf{u} = \mathbf{x}_1 + s\mathbf{v} + t\mathbf{w} \quad \text{for some values of } r, s, \text{ and } t,$$

which holds if and only if

$$\mathbf{x}_0 - \mathbf{x}_1 = -r\mathbf{u} + s\mathbf{v} + t\mathbf{w} \quad \text{for some values of } r, s, \text{ and } t.$$

The latter condition can be rephrased by saying that $\mathbf{x}_0 - \mathbf{x}_1$ lies in Span $(\mathbf{u}, \mathbf{v}, \mathbf{w})$.

Now, there are two ways this can happen. If Span $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ = Span $(\mathbf{v}, \mathbf{w})$, then $\mathbf{x}_0 - \mathbf{x}_1$ lies in Span $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ if and only if $\mathbf{x}_0 - \mathbf{x}_1 = s\mathbf{v} + t\mathbf{w}$ for some values of s and t, and this occurs if and only if $\mathbf{x}_0 = \mathbf{x}_1 + s\mathbf{v} + t\mathbf{w}$, i.e., $\mathbf{x}_0 \in \mathcal{P}$. (Geometrically speaking, in this case the line is parallel to the plane, and they intersect if and only if the line is a subset of the plane.) On the other hand, if Span $(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \mathbb{R}^3$, then ℓ is not parallel to $\mathcal{P}$, and they always intersect. ▲

Exercises 1.1

1. Given $\mathbf{x} = (2, 3)$ and $\mathbf{y} = (-1, 1)$, calculate the following algebraically and sketch a picture to show the geometric interpretation.
   a. $\mathbf{x} + \mathbf{y}$
   b. $\mathbf{x} - \mathbf{y}$
   c. $\mathbf{x} + 2\mathbf{y}$
   d. $\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}$
   e. $\mathbf{y} - \mathbf{x}$
   f. $2\mathbf{x} - \mathbf{y}$
   g. $\|\mathbf{x}\|$
   h. $\frac{\mathbf{x}}{\|\mathbf{x}\|}$
2. For each of the following pairs of vectors $\mathbf{x}$ and $\mathbf{y}$, compute $\mathbf{x} + \mathbf{y}$, $\mathbf{x} - \mathbf{y}$, and $\mathbf{y} - \mathbf{x}$. Also, provide sketches.
   a. $\mathbf{x} = (1, 1)$, $\mathbf{y} = (2, 3)$
   b. $\mathbf{x} = (2, -2)$, $\mathbf{y} = (0, 2)$
   c. $\mathbf{x} = (1, 2, -1)$, $\mathbf{y} = (2, 2, 2)$

*3. Three vertices of a parallelogram are $(1, 2, 1)$, $(2, 4, 3)$, and $(3, 1, 5)$. What are all the possible positions of the fourth vertex? Give your reasoning.[4]

4. Let $A = (1, -1, -1)$, $B = (-1, 1, -1)$, $C = (-1, -1, 1)$, and $D = (1, 1, 1)$. Check that the four triangles formed by these points are all equilateral.

*5. Let ℓ be the line given parametrically by $\mathbf{x} = (1, 3) + t(-2, 1)$, $t \in \mathbb{R}$. Which of the following points lie on ℓ? Give your reasoning.
   a. $\mathbf{x} = (-1, 4)$
   b. $\mathbf{x} = (7, 0)$
   c. $\mathbf{x} = (6, 2)$

[4]For exercises marked with an asterisk (*) we have provided either numerical answers or solutions at the back of the book.

6. Find a parametric equation of each of the following lines:
a. $3x_1 + 4x_2 = 6$
*b. the line with slope $1/3$ that passes through $A = (-1, 2)$
c. the line with slope $2/5$ that passes through $A = (3, 1)$
d. the line through $A = (-2, 1)$ parallel to $\mathbf{x} = (1, 4) + t(3, 5)$
e. the line through $A = (-2, 1)$ perpendicular to $\mathbf{x} = (1, 4) + t(3, 5)$
*f. the line through $A = (1, 2, 1)$ and $B = (2, 1, 0)$
g. the line through $A = (1, -2, 1)$ and $B = (2, 1, -1)$
*h. the line through $(1, 1, 0, -1)$ parallel to $\mathbf{x} = (2 + t, 1 - 2t, 3t, 4 - t)$

7. Suppose $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$ and $\mathbf{y} = \mathbf{y}_0 + s\mathbf{w}$ are two parametric representations of the same line ℓ in $\mathbb{R}^n$.
a. Show that there is a scalar t_0 so that $\mathbf{y}_0 = \mathbf{x}_0 + t_0\mathbf{v}$.
b. Show that $\mathbf{v}$ and $\mathbf{w}$ are parallel.

***8.** Decide whether each of the following vectors is a linear combination of $\mathbf{u} = (1, 0, 1)$ and $\mathbf{v} = (-2, 1, 0)$.
a. $\mathbf{x} = (1, 0, 0)$ b. $\mathbf{x} = (3, -1, 1)$ c. $\mathbf{x} = (0, 1, 2)$

***9.** Let $\mathcal{P}$ be the plane in $\mathbb{R}^3$ spanned by $\mathbf{u} = (1, 1, 0)$ and $\mathbf{v} = (1, -1, 1)$ and passing through the point $(3, 0, -2)$. Which of the following points lie on $\mathcal{P}$?
a. $\mathbf{x} = (4, -1, -1)$ c. $\mathbf{x} = (7, -2, 1)$
b. $\mathbf{x} = (1, -1, 1)$ d. $\mathbf{x} = (5, 2, 0)$

10. Find a parametric equation of each of the following planes:
a. the plane containing the point $(-1, 0, 1)$ and the line $\mathbf{x} = (1, 1, 1) + t(1, 7, -1)$
*b. the plane parallel to the vector $(1, 3, 1)$ and containing the points $(1, 1, 1)$ and $(-2, 1, 2)$
c. the plane containing the points $(1, 1, 2)$, $(2, 3, 4)$, and $(0, -1, 2)$
d. the plane in $\mathbb{R}^4$ containing the points $(1, 1, -1, 2)$, $(2, 3, 0, 1)$, and $(1, 2, 2, 3)$

11. The origin is at the center of a regular m-sided polygon.
a. What is the sum of the vectors from the origin to each of the vertices of the polygon? (The case $m = 7$ is illustrated in Figure 1.17.) Give your reasoning. (*Hint:* What happens if you rotate the vectors by $2\pi/m$?)

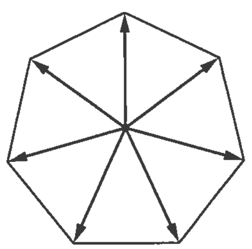

FIGURE 1.17

b. What is the sum of the vectors from one fixed vertex to each of the remaining vertices? (*Hint:* You should use an algebraic approach along with your answer to part *a*.)

***12.** Which of the following are parametric equations of the same plane?
a. $\mathcal{P}_1$: $(1, 1, 0) + s(1, 0, 1) + t(-2, 1, 0)$
b. $\mathcal{P}_2$: $(1, 1, 1) + s(0, 1, 2) + t(2, -1, 0)$
c. $\mathcal{P}_3$: $(2, 0, 0) + s(4, -1, 2) + t(0, 1, 2)$
d. $\mathcal{P}_4$: $(0, 2, 1) + s(1, -1, -1) + t(3, -1, 1)$

13. Given $\triangle ABC$, let M and N be the midpoints of $\overline{AB}$ and $\overline{AC}$, respectively. Prove that $\overrightarrow{MN} = \frac{1}{2}\overrightarrow{BC}$.

14. Let $ABCD$ be an arbitrary quadrilateral. Let P, Q, R, and S be the midpoints of $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, and $\overline{DA}$, respectively. Use Exercise 13 to prove that $PQRS$ is a parallelogram.

***15.** In $\triangle ABC$, shown in Figure 1.18, $\|\overrightarrow{AD}\| = \frac{2}{3}\|\overrightarrow{AB}\|$ and $\|\overrightarrow{CE}\| = \frac{2}{5}\|\overrightarrow{CB}\|$. Let Q denote the midpoint of $\overline{CD}$. Show that $\overrightarrow{AQ} = c\overrightarrow{AE}$ for some scalar c, and determine the ratio $c = \|\overrightarrow{AQ}\|/\|\overrightarrow{AE}\|$.

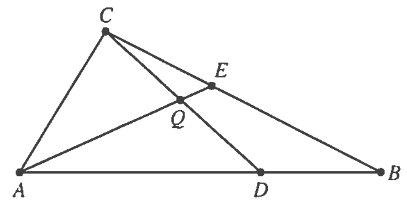

FIGURE 1.18

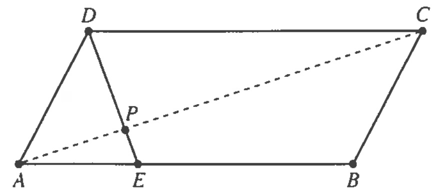

FIGURE 1.19

16. Consider parallelogram $ABCD$. Suppose $\overrightarrow{AE} = \frac{1}{3}\overrightarrow{AB}$ and $\overrightarrow{DP} = \frac{3}{4}\overrightarrow{DE}$. Show that P lies on the diagonal $\overline{AC}$. (See Figure 1.19.)

17. Given $\triangle ABC$, suppose that the point D is 3/4 of the way from A to B and that E is the midpoint of $\overline{BC}$. Use vector methods to show that the point P that is 4/7 of the way from C to D is the intersection point of $\overline{CD}$ and $\overline{AE}$.

18. Let A, B, and C be vertices of a triangle in $\mathbb{R}^3$. Let $\mathbf{x} = \overrightarrow{OA}$, $\mathbf{y} = \overrightarrow{OB}$, and $\mathbf{z} = \overrightarrow{OC}$. Show that the head of the vector $\mathbf{v} = \frac{1}{3}(\mathbf{x} + \mathbf{y} + \mathbf{z})$ lies on each median of $\triangle ABC$ (and thus is the point of intersection of the three medians). This point is called the *centroid* of the triangle ABC.

19. a. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$. Describe the vectors $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$, where $s + t = 1$. What particular subset of such $\mathbf{x}$'s is described by $s \geq 0$? By $t \geq 0$? By $s, t > 0$?
b. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Describe the vectors $\mathbf{x} = r\mathbf{u} + s\mathbf{v} + t\mathbf{w}$, where $r + s + t = 1$. What subsets of such $\mathbf{x}$'s are described by the conditions $r \geq 0$? $s \geq 0$? $t \geq 0$? $r, s, t > 0$?

20. Assume that $\mathbf{u}$ and $\mathbf{v}$ are parallel vectors in $\mathbb{R}^n$. Prove that Span $(\mathbf{u}, \mathbf{v})$ is a line.

21. Suppose $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and c is a scalar. Prove that Span $(\mathbf{v} + c\mathbf{w}, \mathbf{w}) =$ Span $(\mathbf{v}, \mathbf{w})$. (See the blue box on p. 12.)

22. Suppose the vectors $\mathbf{v}$ and $\mathbf{w}$ are both linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$.
a. Prove that for any scalar c, $c\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$.
b. Prove that $\mathbf{v} + \mathbf{w}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$.

When you are asked to "show" or "prove" something, you should make it a point to write down clearly the information you are *given* and what it is you are *to show*. One word of warning regarding part *b*: To say that $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is to say that $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k$ for *some* scalars $c_1, \dots, c_k$. These scalars will surely be different when you express a different vector $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$, so be sure you give the scalars for $\mathbf{w}$ different names.

***23.** Consider the line ℓ: $\mathbf{x} = \mathbf{x}_0 + r\mathbf{v}$ ($r \in \mathbb{R}$) and the plane $\mathcal{P}$: $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$ ($s, t \in \mathbb{R}$). Show that if ℓ and $\mathcal{P}$ intersect, then $\mathbf{x}_0 \in \mathcal{P}$.

24. Consider the lines ℓ: $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$ and m: $\mathbf{x} = \mathbf{x}_1 + s\mathbf{u}$. Show that ℓ and m intersect if and only if $\mathbf{x}_0 - \mathbf{x}_1$ lies in Span $(\mathbf{u}, \mathbf{v})$.

25. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are nonparallel vectors. (Recall the definition on p. 3.)
    a. Prove that if $s\mathbf{x} + t\mathbf{y} = \mathbf{0}$, then $s = t = 0$. (*Hint:* Show that neither $s \neq 0$ nor $t \neq 0$ is possible.)
    b. Prove that if $a\mathbf{x} + b\mathbf{y} = c\mathbf{x} + d\mathbf{y}$, then $a = c$ and $b = d$.

Two important points emerge in this exercise. First is the appearance of *proof by contradiction*. Although it seems impossible to prove the result of part a directly, it is equivalent to prove that if we assume the hypotheses and the *failure* of the conclusion, then we arrive at a contradiction. In this case, if you assume $s\mathbf{x} + t\mathbf{y} = \mathbf{0}$ *and* $s \neq 0$ (or $t \neq 0$), you should be able to see rather easily that $\mathbf{x}$ and $\mathbf{y}$ *are* parallel. In sum, the desired result must be true because it cannot be false.

Next, it is a common (and powerful) technique to prove a result (for example, part b of Exercise 25) by first proving a special case (part a) and then using it to derive the general case. (Another instance you may have seen in a calculus course is the proof of the Mean Value Theorem by reducing to Rolle's Theorem.)

---

26. "Discover" the fraction 2/3 that appears in Proposition 1.2 by finding the intersection of two medians. (Parametrize the line through O and M and the line through A and N, and solve for their point of intersection. You will need to use the result of Exercise 25.)

27. Given $\triangle ABC$, which triangles with vertices on the edges of the original triangle have the same centroid? (See Exercises 18 and 19. At some point, the result of Exercise 25 may be needed, as well.)

28. Verify algebraically that the following properties of vector arithmetic hold. (Do so for $n = 2$ if the general case is too intimidating.) Give the geometric interpretation of each property.
    a. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
    b. For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
    c. $\mathbf{0} + \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.
    d. For each $\mathbf{x} \in \mathbb{R}^n$, there is a vector $-\mathbf{x}$ so that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
    e. For all $c, d \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $c(d\mathbf{x}) = (cd)\mathbf{x}$.
    f. For all $c \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$.
    g. For all $c, d \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$.
    h. For all $\mathbf{x} \in \mathbb{R}^n$, $1\mathbf{x} = \mathbf{x}$.

29. a. Using only the properties listed in Exercise 28, prove that for any $\mathbf{x} \in \mathbb{R}^n$, we have $0\mathbf{x} = \mathbf{0}$. (It often surprises students that this is a consequence of the properties in Exercise 28.)
    b. Using the result of part a, prove that $(-1)\mathbf{x} = -\mathbf{x}$. (Be sure that you didn't use this fact in your proof of part a!)

2 Dot Product

We discuss next one of the crucial constructions in linear algebra, the dot product $\mathbf{x} \cdot \mathbf{y}$ of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By way of motivation, let's recall some basic results from plane geometry. Let $P = (x_1, x_2)$ and $Q = (y_1, y_2)$ be points in the plane, as shown in Figure 2.1. We observe that when $\angle POQ$ is a right angle, $\triangle OAP$ is similar to $\triangle OBQ$, and so $x_2/x_1 = -y_1/y_2$, whence $x_1y_1 + x_2y_2 = 0$.

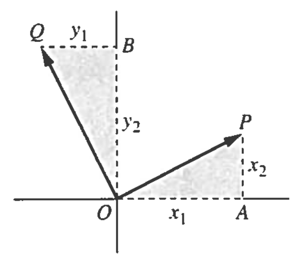

FIGURE 2.1

This leads us to make the following definition.

> ***Definition.*** Given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, define their *dot product*
> $$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2.$$
> More generally, given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, define their dot product
> $$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Remark. The dot product of two vectors is a scalar. For this reason, the dot product is also called the *scalar product*, but it should not be confused with the multiplication of a vector by a scalar, the result of which is a vector. The dot product is also an example of an inner product, which we will study in Section 6 of Chapter 3.

We know that when the vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^2$ are perpendicular, their dot product is 0. By starting with the algebraic properties of the dot product, we are able to get a great deal of geometry out of it.

Proposition 2.1. *The dot product has the following properties:*

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *(the* commutative *property);*
2. $\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2 \geq 0$ *and* $\mathbf{x} \cdot \mathbf{x} = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$*;*
3. $(c\mathbf{x}) \cdot \mathbf{y} = c(\mathbf{x} \cdot \mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$*;*
4. $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$ *for all* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ *(the* distributive *property).*

Proof. In order to simplify the notation, we give the proof with $n = 2$; the general argument would include all n terms with the obligatory .... Because multiplication of real numbers is commutative, we have

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 = y_1x_1 + y_2x_2 = \mathbf{y} \cdot \mathbf{x}.$$

The square of a real number is nonnegative and the sum of nonnegative numbers is nonnegative, so $\mathbf{x} \cdot \mathbf{x} = x_1^2 + x_2^2 \geq 0$ and is equal to 0 only when $x_1 = x_2 = 0$.

The next property follows from the associative and distributive properties of real numbers:

$$\begin{aligned}(c\mathbf{x}) \cdot \mathbf{y} &= (cx_1)y_1 + (cx_2)y_2 = c(x_1y_1) + c(x_2y_2)\\ &= c(x_1y_1 + x_2y_2) = c(\mathbf{x} \cdot \mathbf{y}).\end{aligned}$$

The last result follows from the commutative, associative, and distributive properties of real numbers:

$$\begin{aligned}\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) &= x_1(y_1 + z_1) + x_2(y_2 + z_2) = x_1y_1 + x_1z_1 + x_2y_2 + x_2z_2\\ &= (x_1y_1 + x_2y_2) + (x_1z_1 + x_2z_2) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}.\end{aligned}$$ ☐

Corollary 2.2. $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2$.

Proof. Using the properties of Proposition 2.1 repeatedly, we have

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})\\ &= \mathbf{x}\cdot\mathbf{x}+\mathbf{x}\cdot\mathbf{y}+\mathbf{y}\cdot\mathbf{x}+\mathbf{y}\cdot\mathbf{y}\\ &= \|\mathbf{x}\|^2 + 2\mathbf{x}\cdot\mathbf{y}+\|\mathbf{y}\|^2,\end{aligned}$$

as desired. □

> Although we use coordinates to define the dot product and to derive its algebraic properties in Proposition 2.1, from this point on we should try to use the properties *themselves* to prove results (e.g., Corollary 2.2). This will tend to avoid an algebraic mess and emphasize the geometry.

The geometric meaning of this result comes from the Pythagorean Theorem: When $\mathbf{x}$ and $\mathbf{y}$ are perpendicular vectors in $\mathbb{R}^2$, as shown in Figure 2.2, we have $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, and so, by Corollary 2.2, it must be the case that $\mathbf{x}\cdot\mathbf{y} = 0$. (And the converse follows, too, from the converse of the Pythagorean Theorem, which follows from the Law of Cosines. See Exercise 14.) That is, two vectors in $\mathbb{R}^2$ are perpendicular if and only if their dot product is 0.

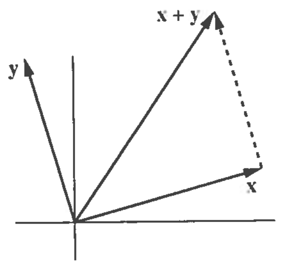

FIGURE 2.2

Motivated by this, we use the algebraic definition of the dot product of vectors in $\mathbb{R}^n$ to bring in the geometry.

> ***Definition.*** We say vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ are *orthogonal*[5] if $\mathbf{x}\cdot\mathbf{y} = 0$.

Orthogonal and *perpendicular* are synonyms, but we shall stick to the former, because that is the common terminology in linear algebra texts.

EXAMPLE 1

To illustrate the power of the algebraic properties of the dot product, we prove that the diagonals of a parallelogram are orthogonal if and only if the parallelogram is a rhombus (that is, all sides have equal length). As usual, we place one vertex at the origin (see Figure 2.3),

[5]This word derives from the Greek *orthos*, meaning "straight," "right," or "true," and *gōnia*, meaning "angle."

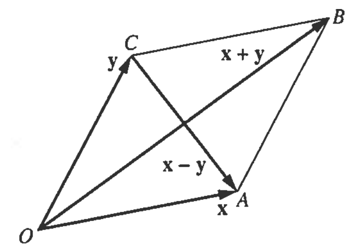

FIGURE 2.3

and we let $\mathbf{x} = \overrightarrow{OA}$ and $\mathbf{y} = \overrightarrow{OC}$ be vectors representing adjacent sides emanating from the origin. We have the diagonals $\overrightarrow{OB} = \mathbf{x}+\mathbf{y}$ and $\overrightarrow{CA} = \mathbf{x}-\mathbf{y}$, so the diagonals are orthogonal if and only if

$$(\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) = 0.$$

Using the properties of dot product to expand this expression, we obtain

$$(\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) = \mathbf{x}\cdot\mathbf{x}+\mathbf{y}\cdot\mathbf{x}-\mathbf{x}\cdot\mathbf{y}-\mathbf{y}\cdot\mathbf{y} = \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2,$$

so the diagonals are orthogonal if and only if $\|\mathbf{x}\|^2 = \|\mathbf{y}\|^2$. Since the length of a vector is nonnegative, this occurs if and only if $\|\mathbf{x}\| = \|\mathbf{y}\|$, which means that all the sides of the parallelogram have equal length. ▲

> In general, when you are asked to prove a statement of the form *P if and only if Q*, this means that you must prove two statements: If P is true, then Q is also true ("only if"); and if Q is true, then P is also true ("if"). In this example, we gave the two arguments simultaneously, because they relied essentially only on algebraic identities.
>
> A useful shorthand for writing proofs is the *implication* symbol, $\Longrightarrow$. The sentence
>
> $$P \Longrightarrow Q$$
>
> can be read in numerous ways:
>
> - "if P, then Q"
> - "P implies Q"
> - "P only if Q"
> - "Q whenever P"
> - "P is sufficient for Q" (because when P is true, then Q is true as well)
> - "Q is necessary for P" (because P can't be true unless Q is true)
>
> The "reverse implication" symbol, $\Longleftarrow$, occurs less frequently, because we ordinarily write "$P \Longleftarrow Q$" as "$Q \Longrightarrow P$." This is called the *converse* of the original implication. To convince yourself that a proposition and its converse are logically distinct, consider the sentence "If students major in mathematics, then they take a linear algebra course." The converse is "If students take a linear algebra course, then they major in mathematics." How many of the students in this class are mathematics majors??
>
> We often use the symbol $\Longleftrightarrow$ to denote "if and only if": $P \Longleftrightarrow Q$ means "$P \Longrightarrow Q$ and $Q \Longrightarrow P$." This is often read "P is necessary and sufficient for Q"; here necessity corresponds to "$Q \Longrightarrow P$" and sufficiency corresponds to "$P \Longrightarrow Q$."

Armed with the definition of orthogonal vectors, we proceed to a construction that will be important in much of our future work. Starting with two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\mathbf{y} \neq \mathbf{0}$, Figure 2.4 suggests that we should be able to write $\mathbf{x}$ as the sum of a vector, $\mathbf{x}^{\parallel}$ (read "x-parallel"), that is a scalar multiple of $\mathbf{y}$ and a vector, $\mathbf{x}^{\perp}$ (read "x-perp"), that is orthogonal to $\mathbf{y}$. Let's suppose we have such an equation:

$$\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}, \quad \text{where}$$
$$\mathbf{x}^{\parallel} \text{ is a scalar multiple of } \mathbf{y} \quad \text{and} \quad \mathbf{x}^{\perp} \text{ is orthogonal to } \mathbf{y}.$$

To say that $\mathbf{x}^{\parallel}$ is a scalar multiple of $\mathbf{y}$ means that we can write $\mathbf{x}^{\parallel} = c\mathbf{y}$ for some scalar c. Now, assuming such an expression exists, we can determine c by taking the dot product of both sides of the equation with $\mathbf{y}$:

$$\mathbf{x}\cdot\mathbf{y} = (\mathbf{x}^{\parallel} + \mathbf{x}^{\perp})\cdot\mathbf{y} = (\mathbf{x}^{\parallel}\cdot\mathbf{y}) + (\mathbf{x}^{\perp}\cdot\mathbf{y}) = \mathbf{x}^{\parallel}\cdot\mathbf{y} = (c\mathbf{y})\cdot\mathbf{y} = c\|\mathbf{y}\|^2.$$

This means that

$$c = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}, \quad \text{and so} \quad \mathbf{x}^{\parallel} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}.$$

The vector $\mathbf{x}^{\parallel}$ is called the *projection of* $\mathbf{x}$ *onto* $\mathbf{y}$, written $\text{proj}_{\mathbf{y}}\mathbf{x}$.

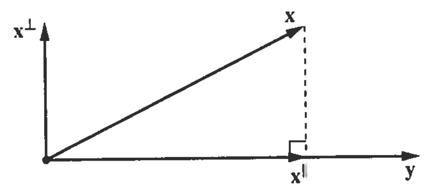

FIGURE 2.4

The fastidious reader may be puzzled by the logic here. We have apparently assumed that we can write $\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}$ in order to prove that we can do so. Of course, as it stands, this is no fair. Here's how we fix it. We now *define*

$$\mathbf{x}^{\parallel} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}$$
$$\mathbf{x}^{\perp} = \mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}.$$

Obviously, $\mathbf{x}^{\parallel} + \mathbf{x}^{\perp} = \mathbf{x}$ and $\mathbf{x}^{\parallel}$ is a scalar multiple of $\mathbf{y}$. All we need to check is that $\mathbf{x}^{\perp}$ is in fact orthogonal to $\mathbf{y}$. Well,

$$\begin{aligned}\mathbf{x}^{\perp}\cdot\mathbf{y} &= \left(\mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}\right)\cdot\mathbf{y} \\ &= \mathbf{x}\cdot\mathbf{y} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}\cdot\mathbf{y} \\ &= \mathbf{x}\cdot\mathbf{y} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\|\mathbf{y}\|^2 \\ &= \mathbf{x}\cdot\mathbf{y} - \mathbf{x}\cdot\mathbf{y} = 0,\end{aligned}$$

as required. Note that by finding a formula for c above, we have shown that $\mathbf{x}^{\parallel}$ is the *unique* multiple of $\mathbf{y}$ that satisfies the equation $(\mathbf{x} - \mathbf{x}^{\parallel})\cdot\mathbf{y} = 0$.

The pattern of reasoning we've just been through is really not that foreign. When we "solve" the equation

$$\sqrt{x+2} = 2,$$

we assume x satisfies this equation and proceed to find candidates for x. At the end of the process, we must check to see which of our answers work. In this case, of course, we assume x satisfies the equation, square both sides, and conclude that $x = 2$. (That is, *if* $\sqrt{x+2} = 2$, *then* x must equal 2.) But we check the converse: If $x = 2$, then $\sqrt{x+2} = \sqrt{4} = 2$.

It is a bit more interesting if we try solving

$$\sqrt{x+2} = x.$$

Now, squaring both sides leads to the equation

$$x^2 - x - 2 = (x-2)(x+1) = 0,$$

and so we conclude that *if* x satisfies the given equation, *then* $x = 2$ *or* $x = -1$. As before, $x = 2$ is a fine solution, but $x = -1$ is not.

EXAMPLE 2

Let $\mathbf{x} = (2, 3, 1)$ and $\mathbf{y} = (-1, 1, 1)$. Then

$$\mathbf{x}^{\parallel} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} = \frac{(2,3,1)\cdot(-1,1,1)}{\|(-1,1,1)\|^2}(-1,1,1) = \tfrac{2}{3}(-1,1,1) \quad \text{and}$$
$$\mathbf{x}^{\perp} = (2,3,1) - \tfrac{2}{3}(-1,1,1) = \left(\tfrac{8}{3}, \tfrac{7}{3}, \tfrac{1}{3}\right).$$

To double-check, we compute $\mathbf{x}^{\perp}\cdot\mathbf{y} = \left(\tfrac{8}{3}, \tfrac{7}{3}, \tfrac{1}{3}\right)\cdot(-1,1,1) = 0$, as it should be. ▲

Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. We shall see next that the formula for the projection of $\mathbf{x}$ onto $\mathbf{y}$ enables us to calculate the *angle* between the vectors $\mathbf{x}$ and $\mathbf{y}$. Consider the right triangle in Figure 2.5; let θ denote the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. Remembering that the

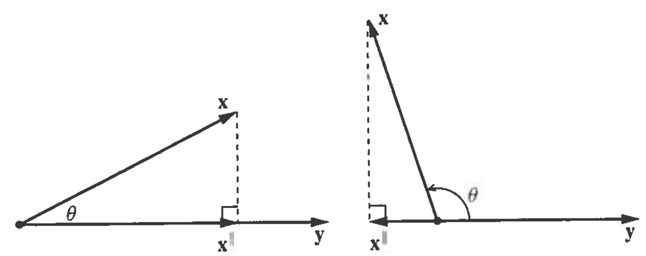

FIGURE 2.5

cosine of an angle is the ratio of the *signed* length of the adjacent side to the length of the hypotenuse, we see that

$$\cos\theta = \frac{\text{signed length of } \mathbf{x}^{\parallel}}{\text{length of } \mathbf{x}} = \frac{c\|\mathbf{y}\|}{\|\mathbf{x}\|} = \frac{\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\|\mathbf{y}\|}{\|\mathbf{x}\|} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

This, then, is the geometric interpretation of the dot product:

$$\boxed{\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta.}$$

Note that if the angle θ is obtuse, i.e., $\pi/2 < |\theta| < \pi$, then $c < 0$ (the signed length of $\mathbf{x}^{\parallel}$ is negative) and $\mathbf{x}\cdot\mathbf{y}$ is negative.

Will this formula still make sense even when $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$? Geometrically, we simply restrict our attention to the plane spanned by $\mathbf{x}$ and $\mathbf{y}$ and measure the angle θ in that plane, and so we blithely make the following definition.

> ***Definition.*** Let $\mathbf{x}$ and $\mathbf{y}$ be nonzero vectors in $\mathbb{R}^n$. We define the *angle* between them to be the unique θ satisfying $0 \le \theta \le \pi$ so that
>
> $$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

EXAMPLE 3

Set $A = (1,-1,-1)$, $B = (-1,1,-1)$, and $C = (-1,-1,1)$. Then $\overrightarrow{AB} = (-2,2,0)$ and $\overrightarrow{AC} = (-2,0,2)$, so

$$\cos\angle BAC = \frac{\overrightarrow{AB}\cdot\overrightarrow{AC}}{\|\overrightarrow{AB}\|\|\overrightarrow{AC}\|} = \frac{4}{(2\sqrt{2})^2} = \frac{1}{2}.$$

We conclude that $\angle BAC = \pi/3$. ▲

Since our geometric intuition may be misleading in $\mathbb{R}^n$, we should check *algebraically* that this definition makes sense. Since $|\cos\theta| \le 1$, the following result gives us what is needed.

Proposition 2.3 (Cauchy-Schwarz Inequality). *If* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *then*

$$|\mathbf{x}\cdot\mathbf{y}| \le \|\mathbf{x}\|\|\mathbf{y}\|.$$

Moreover, equality holds if and only if one of the vectors is a scalar multiple of the other.

Proof. If one of the vectors is the zero vector, the result is immediate, so we assume both vectors are nonzero. Suppose first that both $\mathbf{x}$ and $\mathbf{y}$ are unit vectors. Each of the vectors $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ (which we can picture as the diagonals of the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ when the vectors are nonparallel, as shown in Figure 2.6) has nonnegative length.

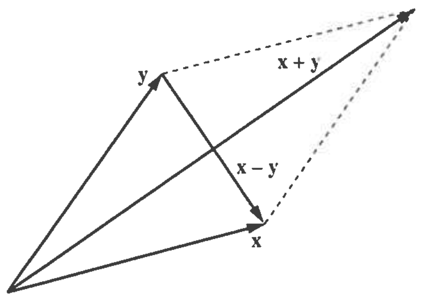

FIGURE 2.6

Using Corollary 2.2, we have

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2 = 2(\mathbf{x}\cdot\mathbf{y}+1)$$
$$\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 - 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2 = 2(-\mathbf{x}\cdot\mathbf{y}+1).$$

Since $\|\mathbf{x}+\mathbf{y}\|^2 \ge 0$ and $\|\mathbf{x}-\mathbf{y}\|^2 \ge 0$, we see that $\mathbf{x}\cdot\mathbf{y}+1 \ge 0$ and $-\mathbf{x}\cdot\mathbf{y}+1 \ge 0$. Thus,

$$-1 \le \mathbf{x}\cdot\mathbf{y} \le 1, \quad \text{and so} \quad |\mathbf{x}\cdot\mathbf{y}| \le 1.$$

Note that equality holds if and only if either $\mathbf{x}+\mathbf{y} = \mathbf{0}$ or $\mathbf{x}-\mathbf{y} = \mathbf{0}$, i.e., if and only if $\mathbf{x} = \pm\mathbf{y}$.

In general, since $\mathbf{x}/\|\mathbf{x}\|$ and $\mathbf{y}/\|\mathbf{y}\|$ are unit vectors, we have

$$\left|\frac{\mathbf{x}}{\|\mathbf{x}\|}\cdot\frac{\mathbf{y}}{\|\mathbf{y}\|}\right| \le 1, \quad \text{and so} \quad |\mathbf{x}\cdot\mathbf{y}| \le \|\mathbf{x}\|\|\mathbf{y}\|,$$

as required. Equality holds if and only if $\dfrac{\mathbf{x}}{\|\mathbf{x}\|} = \pm\dfrac{\mathbf{y}}{\|\mathbf{y}\|}$; that is, equality holds if and only if $\mathbf{x}$ and $\mathbf{y}$ are parallel. □

Remark. The dot product also arises in situations removed from geometry. The economist introduces the *commodity vector*, whose entries are the quantities of various commodities that happen to be of interest. For example, we might consider $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5$, where x_1 represents the number of pounds of flour, x_2 the number of dozens of eggs, x_3 the number of pounds of chocolate chips, x_4 the number of pounds of walnuts, and x_5 the number of pounds of butter needed to produce a certain massive quantity of chocolate chip cookies. The economist next introduces the *price vector* $\mathbf{p} = (p_1, p_2, p_3, p_4, p_5) \in \mathbb{R}^5$, where p_i is the price (in dollars) of a unit of the i^{th} commodity (for example, p_2 is the price of a dozen eggs). Then it follows that

$$\mathbf{p}\cdot\mathbf{x} = p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 + p_5x_5$$

is the total cost of producing the massive quantity of cookies. (To be realistic, we might also want to include x_6 as the number of hours of labor, with corresponding hourly wage p_6.) We will return to this interpretation in Section 5 of Chapter 2.

The gambler uses the dot product to compute the *expected value* of a lottery that has multiple payoffs with various probabilities. If the possible payoffs for a given lottery are given by $\mathbf{w} = (w_1, \dots, w_n)$ and the probabilities of winning the respective payoffs are given by $\mathbf{p} = (p_1, \dots, p_n)$, with $p_1 + \dots + p_n = 1$, then the expected value of the lottery is $\mathbf{p}\cdot\mathbf{w} = p_1w_1 + \dots + p_nw_n$. For example, if the possible prizes, in dollars, for a particular lottery are given by the payoff vector $\mathbf{w} = (0, 1, 5, 100)$ and the probability vector is $\mathbf{p} = (0.5, 0.4, 0.09, 0.01)$, then the expected value is $\mathbf{p}\cdot\mathbf{w} = 0.4 + 0.45 + 1 = 1.85$. Thus, if the lottery ticket costs more than \$1.85, the gambler should expect to lose money in the long run.

Exercises 1.2

1. For each of the following pairs of vectors $\mathbf{x}$ and $\mathbf{y}$, calculate $\mathbf{x}\cdot\mathbf{y}$ and the angle θ between the vectors.
   a. $\mathbf{x} = (2,5)$, $\mathbf{y} = (-5,2)$
   b. $\mathbf{x} = (2,1)$, $\mathbf{y} = (-1,1)$
   *c. $\mathbf{x} = (1,8)$, $\mathbf{y} = (7,-4)$
   d. $\mathbf{x} = (1,4,-3)$, $\mathbf{y} = (5,1,3)$
   e. $\mathbf{x} = (1,-1,6)$, $\mathbf{y} = (5,3,2)$
   *f. $\mathbf{x} = (3,-4,5)$, $\mathbf{y} = (-1,0,1)$
   g. $\mathbf{x} = (1,1,1,1)$, $\mathbf{y} = (1,-3,-1,5)$

***2.** For each pair of vectors in Exercise 1, calculate $\text{proj}_{\mathbf{y}}\mathbf{x}$ and $\text{proj}_{\mathbf{x}}\mathbf{y}$.

3. A methane molecule has four hydrogen (H) atoms at the points indicated in Figure 2.7 and a carbon (C) atom at the origin. Find the H − C − H bond angle. (Because of the result of Exercise 1.1.4, this configuration is called a regular tetrahedron.)

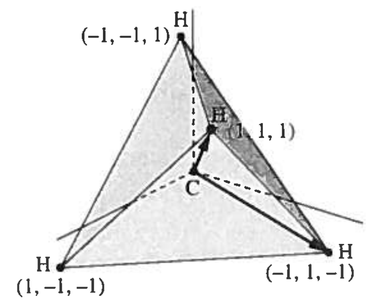

FIGURE 2.7

***4.** Find the angle between the long diagonal of a cube and a face diagonal.

5. Find the angle that the long diagonal of a $3 \times 4 \times 5$ rectangular box makes with the longest edge.

***6.** Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x}\| = 3$, $\|\mathbf{y}\| = 2$, and the angle θ between $\mathbf{x}$ and $\mathbf{y}$ is $\theta = \arccos(-1/6)$. Show that the vectors $\mathbf{x} + 2\mathbf{y}$ and $\mathbf{x} - \mathbf{y}$ are orthogonal.

7. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x}\| = \sqrt{2}$, $\|\mathbf{y}\| = 1$, and the angle between $\mathbf{x}$ and $\mathbf{y}$ is $3\pi/4$. Show that the vectors $2\mathbf{x} + 3\mathbf{y}$ and $\mathbf{x} - \mathbf{y}$ are orthogonal.

8. Suppose $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ are unit vectors satisfying $\mathbf{x} + \mathbf{y} + \mathbf{z} = \mathbf{0}$. Determine the angles between each pair of vectors.

9. Let $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, and $\mathbf{e}_3 = (0, 0, 1)$ be the so-called *standard basis* for $\mathbb{R}^3$. Let $\mathbf{x} \in \mathbb{R}^3$ be a nonzero vector. For $i = 1, 2, 3$, let θ_i denote the angle between $\mathbf{x}$ and $\mathbf{e}_i$. Compute $\cos^2\theta_1 + \cos^2\theta_2 + \cos^2\theta_3$.

***10.** Let $\mathbf{x} = (1, 1, 1, \dots, 1) \in \mathbb{R}^n$ and $\mathbf{y} = (1, 2, 3, \dots, n) \in \mathbb{R}^n$. Let θ_n be the angle between $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. Find $\lim_{n\to\infty} \theta_n$. (The formulas $1 + 2 + \dots + n = n(n+1)/2$ and $1^2 + 2^2 + \dots + n^2 = n(n+1)(2n+1)/6$ may be useful.)

♯11. Suppose $\mathbf{x}, \mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ and $\mathbf{x}$ is orthogonal to each of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. Show that $\mathbf{x}$ is orthogonal to any linear combination $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k$.[6]

12. Use vector methods to prove that a parallelogram is a rectangle if and only if its diagonals have the same length.

13. Use the algebraic properties of the dot product to show that

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right).$$

Interpret the result geometrically.

***14.** Use the dot product to prove the law of cosines: As shown in Figure 2.8,

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

15. Use vector methods to prove that a triangle that is inscribed in a circle and has a diameter as one of its sides must be a right triangle. (*Hint:* See Figure 2.9. Express the vectors $\mathbf{u}$ and $\mathbf{v}$ in terms of $\mathbf{x}$ and $\mathbf{y}$.)

[6]The symbol ♯ indicates that the result of this problem will be used later.

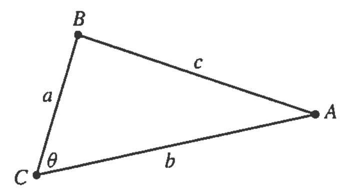

FIGURE 2.8

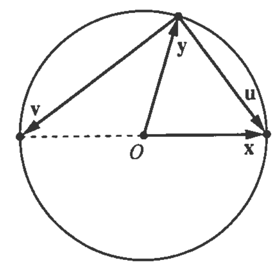

FIGURE 2.9

16. a. Let $\mathbf{y} \in \mathbb{R}^n$. If $\mathbf{x} \cdot \mathbf{y} = 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$, then prove that $\mathbf{y} = \mathbf{0}$.

> When you know some equation holds *for all values of* $\mathbf{x}$, you should often choose some strategic, *particular* value(s) for $\mathbf{x}$.

b. Suppose $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{z}$ for all $\mathbf{x} \in \mathbb{R}^n$. What can you conclude? (*Hint:* Apply the result of part *a*.)

17. If $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, set $\rho(\mathbf{x}) = (-x_2, x_1)$.

a. Check that $\rho(\mathbf{x})$ is orthogonal to $\mathbf{x}$. (Indeed, $\rho(\mathbf{x})$ is obtained by rotating $\mathbf{x}$ an angle $\pi/2$ counterclockwise.)

b. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, show that $\mathbf{x} \cdot \rho(\mathbf{y}) = -\rho(\mathbf{x}) \cdot \mathbf{y}$. Interpret this statement geometrically.

18. Prove the *triangle inequality*: For any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$. (*Hint:* Use the dot product to calculate $\|\mathbf{x} + \mathbf{y}\|^2$.)

19. a. Give an alternative proof of the Cauchy-Schwarz Inequality by minimizing the quadratic function $Q(t) = \|\mathbf{x} - t\mathbf{y}\|^2$. Note that $Q(t) \ge 0$ for all t.

b. If $Q(t_0) \le Q(t)$ for all t, how is $t_0\mathbf{y}$ related to $\mathbf{x}^{\parallel}$? What does this say about $\text{proj}_{\mathbf{y}}\mathbf{x}$?

20. Use the Cauchy-Schwarz inequality to solve the following max/min problem: If the (long) diagonal of a rectangular box has length c, what is the greatest that the sum of the length, width, and height of the box can be? For what shape box does the maximum occur?

21. a. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors with $\|\mathbf{x}\| = \|\mathbf{y}\|$. Prove that the vector $\mathbf{x} + \mathbf{y}$ bisects the angle between $\mathbf{x}$ and $\mathbf{y}$. (*Hint:* Because $\mathbf{x} + \mathbf{y}$ lies in the plane spanned by $\mathbf{x}$ and $\mathbf{y}$, one has only to check that the angle between $\mathbf{x}$ and $\mathbf{x} + \mathbf{y}$ equals the angle between $\mathbf{y}$ and $\mathbf{x} + \mathbf{y}$.)

b. More generally, if $\mathbf{x}$ and $\mathbf{y}$ are arbitrary nonzero vectors, let $a = \|\mathbf{x}\|$ and $b = \|\mathbf{y}\|$. Prove that the vector $b\mathbf{x} + a\mathbf{y}$ bisects the angle between $\mathbf{x}$ and $\mathbf{y}$.

22. Use vector methods to prove that the diagonals of a parallelogram bisect the vertex angles if and only if the parallelogram is a rhombus. (*Hint:* Use Exercise 21.)

23. Given $\triangle ABC$ with D on $\overline{BC}$, as shown in Figure 2.10, prove that if $\overline{AD}$ bisects $\angle BAC$, then $\|\overrightarrow{BD}\|/\|\overrightarrow{CD}\| = \|\overrightarrow{AB}\|/\|\overrightarrow{AC}\|$. (*Hint:* Use part *b* of Exercise 21. Let $\mathbf{x} = \overrightarrow{AB}$

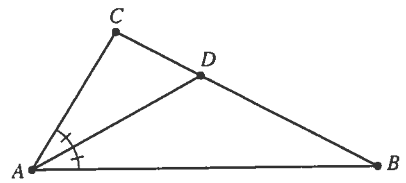

FIGURE 2.10

and $\mathbf{y} = \overrightarrow{AC}$; express $\overrightarrow{AD}$ in two ways as a linear combination of $\mathbf{x}$ and $\mathbf{y}$ and use Exercise 1.1.25.)

24. Use vector methods to show that the angle bisectors of a triangle have a common point. (*Hint:* Given $\triangle OAB$, let $\mathbf{x} = \overrightarrow{OA}$, $\mathbf{y} = \overrightarrow{OB}$, $a = \|\overrightarrow{OA}\|$, $b = \|\overrightarrow{OB}\|$, and $c = \|\overrightarrow{AB}\|$. If we define the point P by $\overrightarrow{OP} = \frac{1}{a+b+c}(b\mathbf{x} + a\mathbf{y})$, use part b of Exercise 21 to show that P lies on all three angle bisectors.)

25. Use vector methods to show that the altitudes of a triangle have a common point. Recall that altitudes of a triangle are the lines passing through a vertex and perpendicular to the line through the remaining vertices. (*Hint:* See Figure 2.11. Let C be the point of intersection of the altitude from B and the altitude from A. Show that $\overrightarrow{OC}$ is orthogonal to $\overrightarrow{AB}$.)

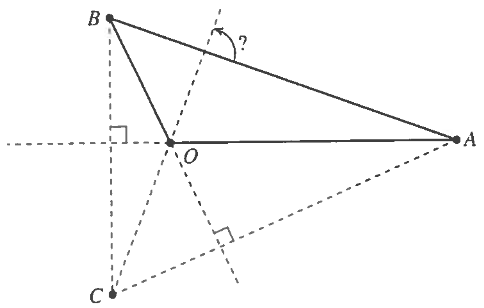

FIGURE 2.11

26. Use vector methods to show that the perpendicular bisectors of the sides of a triangle intersect in a point, as follows. Assume the triangle OAB has one vertex at the origin, and let $\mathbf{x} = \overrightarrow{OA}$ and $\mathbf{y} = \overrightarrow{OB}$. Let $\mathbf{z}$ be the point of intersection of the perpendicular bisectors of $\overline{OA}$ and $\overline{OB}$. Show that $\mathbf{z}$ lies on the perpendicular bisector of $\overline{AB}$. (*Hint:* What is the dot product of $\mathbf{z} - \frac{1}{2}(\mathbf{x} + \mathbf{y})$ with $\mathbf{x} - \mathbf{y}$?)

3 Hyperplanes in $\mathbb{R}^n$

We emphasized earlier a parametric description of lines in $\mathbb{R}^2$ and planes in $\mathbb{R}^3$. Let's begin by revisiting the Cartesian equation of a line passing through the origin in $\mathbb{R}^2$, e.g.,

$$2x_1 + x_2 = 0.$$

We recognize that the left-hand side of this equation is the dot product of the vector $\mathbf{a} = (2, 1)$ with $\mathbf{x} = (x_1, x_2)$. That is, the vector $\mathbf{x}$ satisfies this equation precisely when it is orthogonal to the vector $\mathbf{a}$, as indicated in Figure 3.1, and we have described the line as the set of vectors in the plane orthogonal to the given vector $\mathbf{a} = (2, 1)$:

$$(*) \qquad \mathbf{a} \cdot \mathbf{x} = 0.$$

It is customary to say that $\mathbf{a}$ is a *normal*[7] *vector* to the line. (Note that any nonzero scalar multiple of $\mathbf{a}$ will do just as well, but we often abuse language by referring to "the" normal vector.)

[7] This is the first of several occurrences of the word *normal*—evidence of mathematicians' propensity to use a word repeatedly with different meanings. Here the meaning derives from the Latin *norma*, "carpenter's square."

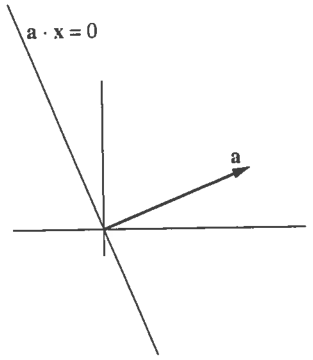

FIGURE 3.1

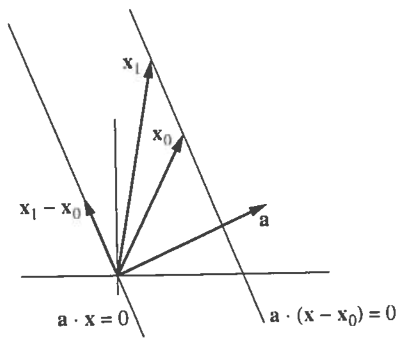

FIGURE 3.2

It is easy to see that specifying a normal vector to a line through the origin is equivalent to specifying its slope. Specifically, if the normal vector is (a, b), then the line has slope $-a/b$. What is the effect of varying the constant on the right-hand side of the equation $(*)$? We get different lines parallel to the one with which we started. In particular, consider a parallel line passing through the point $\mathbf{x}_0$, as shown in Figure 3.2. If $\mathbf{x}$ is on the line, then $\mathbf{x} - \mathbf{x}_0$ will be orthogonal to $\mathbf{a}$, and hence the Cartesian equation of the line is

$$\mathbf{a} \cdot (\mathbf{x} - \mathbf{x}_0) = 0,$$

which we can rewrite in the form

$$\mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot \mathbf{x}_0$$

or

$$\mathbf{a} \cdot \mathbf{x} = c,$$

where c is the fixed real number $\mathbf{a} \cdot \mathbf{x}_0$.[8] (Why is this quantity the same for every point $\mathbf{x}_0$ on the line?)

EXAMPLE 1

Consider the line ℓ_0 through the origin in $\mathbb{R}^2$ with direction vector $\mathbf{v} = (1, -3)$. The points on this line are all of the form

$$\mathbf{x} = t(1, -3), \quad t \in \mathbb{R}.$$

Because $(3, 1) \cdot (1, -3) = 0$, we may take $\mathbf{a} = (3, 1)$ to be the normal vector to the line, and the Cartesian equation of ℓ_0 is

$$\mathbf{a} \cdot \mathbf{x} = 3x_1 + x_2 = 0.$$

(As a check, suppose we start with $3x_1 + x_2 = 0$. Then we can write $x_1 = -\frac{1}{3}x_2$, and so the solutions consist of vectors of the form

$$\mathbf{x} = (x_1, x_2) = \left(-\tfrac{1}{3}x_2, x_2\right) = -\tfrac{1}{3}x_2(1, -3), \quad x_2 \in \mathbb{R}.$$

Letting $t = -\frac{1}{3}x_2$, we recover the original parametric equation.)

[8] The sophisticated reader should compare this to the study of level curves of functions in multivariable calculus. Here our function is $f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$.

Now consider the line ℓ passing through $\mathbf{x}_0 = (2, 1)$ with direction vector $\mathbf{v} = (1, -3)$. Then the points on ℓ are all of the form

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} = (2, 1) + t(1, -3), \quad t \in \mathbb{R}.$$

As promised, we take the same vector $\mathbf{a} = (3, 1)$ and compute that

$$3x_1 + x_2 = \mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot (\mathbf{x}_0 + t\mathbf{v}) = \mathbf{a} \cdot \mathbf{x}_0 + t(\mathbf{a} \cdot \mathbf{v}) = \mathbf{a} \cdot \mathbf{x}_0 = (3, 1) \cdot (2, 1) = 7.$$

This is the Cartesian equation of ℓ. ▲

We can give a geometric interpretation of the constant c on the right-hand side of the equation $\mathbf{a} \cdot \mathbf{x} = c$. Recall that

$$\mathrm{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a},$$

and so, as indicated in Figure 3.3, the line consists of all vectors whose projection onto the normal vector $\mathbf{a}$ is the constant vector

$$\frac{c}{\|\mathbf{a}\|^2}\mathbf{a}.$$

In particular, since the hypotenuse of a right triangle is longer than either leg,

$$\frac{c}{\|\mathbf{a}\|^2}\mathbf{a}$$

is the point on the line closest to the origin, and we say that the *distance from the origin to the line* is

$$\left\| \frac{c}{\|\mathbf{a}\|^2}\mathbf{a} \right\| = \frac{|c|}{\|\mathbf{a}\|} = \|\mathrm{proj}_{\mathbf{a}}\mathbf{x}_0\|$$

for any point $\mathbf{x}_0$ on the line.

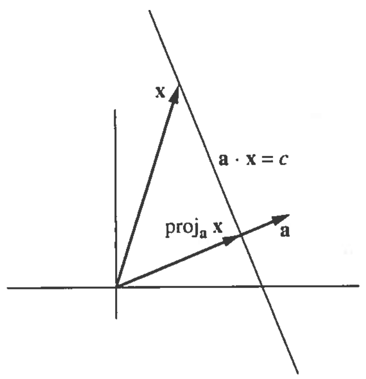

FIGURE 3.3

We now move on to see that planes in $\mathbb{R}^3$ can also be described by using normal vectors.

EXAMPLE 2

Consider the plane $\mathcal{P}_0$ passing through the origin spanned by $\mathbf{u} = (1, 0, 1)$ and $\mathbf{v} = (2, 1, 1)$, as indicated schematically in Figure 3.4. Our intuition suggests that there is a line orthogonal to $\mathcal{P}_0$, so we look for a vector $\mathbf{a} = (a_1, a_2, a_3)$ that is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. It must satisfy the equations

$$\begin{aligned} a_1 \qquad + a_3 &= 0 \\ 2a_1 + a_2 + a_3 &= 0. \end{aligned}$$

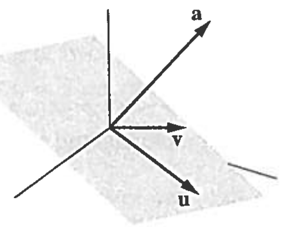

FIGURE 3.4

Substituting $a_3 = -a_1$ into the second equation, we obtain $a_1 + a_2 = 0$, so $a_2 = -a_1$ as well. Thus, any candidate for $\mathbf{a}$ must be a scalar multiple of the vector $(1, -1, -1)$, and so we take $\mathbf{a} = (1, -1, -1)$ and try the equation

$$\mathbf{a} \cdot \mathbf{x} = (1, -1, -1) \cdot \mathbf{x} = x_1 - x_2 - x_3 = 0$$

for $\mathcal{P}_0$. Now, we know that $\mathbf{a} \cdot \mathbf{u} = \mathbf{a} \cdot \mathbf{v} = 0$. Does it follow that $\mathbf{a}$ is orthogonal to every linear combination of $\mathbf{u}$ and $\mathbf{v}$? We just compute: If $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$, then

$$\begin{aligned} \mathbf{a} \cdot \mathbf{x} &= \mathbf{a} \cdot (s\mathbf{u} + t\mathbf{v}) \\ &= s(\mathbf{a} \cdot \mathbf{u}) + t(\mathbf{a} \cdot \mathbf{v}) = 0, \end{aligned}$$

as desired.

As before, if we want the equation of the plane $\mathcal{P}$ parallel to $\mathcal{P}_0$ and passing through $\mathbf{x}_0 = (2, 3, -2)$, we take

$$\begin{aligned} x_1 - x_2 - x_3 = \mathbf{a} \cdot \mathbf{x} &= \mathbf{a} \cdot (\mathbf{x}_0 + s\mathbf{u} + t\mathbf{v}) \\ &= \mathbf{a} \cdot \mathbf{x}_0 + s(\mathbf{a} \cdot \mathbf{u}) + t(\mathbf{a} \cdot \mathbf{v}) \\ &= \mathbf{a} \cdot \mathbf{x}_0 = (1, -1, -1) \cdot (2, 3, -2) = 1. \end{aligned}$$ ▲

As this example suggests, a point $\mathbf{x}_0$ and a normal vector $\mathbf{a}$ give rise to the *Cartesian equation* of a plane in $\mathbb{R}^3$:

$$\mathbf{a} \cdot (\mathbf{x} - \mathbf{x}_0) = 0, \quad \text{or, equivalently,} \quad \mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot \mathbf{x}_0.$$

Thus, every plane in $\mathbb{R}^3$ has an equation of the form

$$a_1x_1 + a_2x_2 + a_3x_3 = c,$$

where $\mathbf{a} = (a_1, a_2, a_3)$ is the normal vector and $c \in \mathbb{R}$.

EXAMPLE 3

Consider the set of points $\mathbf{x} = (x_1, x_2, x_3)$ defined by the equation

$$x_1 - 2x_2 + 5x_3 = 3.$$

Let's verify that this is, in fact, a plane in $\mathbb{R}^3$ according to our original parametric definition. If $\mathbf{x}$ satisfies this equation, then $x_1 = 3 + 2x_2 - 5x_3$ and so we may write

$$\begin{aligned} \mathbf{x} = (x_1, x_2, x_3) &= (3 + 2x_2 - 5x_3, x_2, x_3) \\ &= (3, 0, 0) + x_2(2, 1, 0) + x_3(-5, 0, 1). \end{aligned}$$

So, if we let $\mathbf{x}_0 = (3, 0, 0)$, $\mathbf{u} = (2, 1, 0)$, and $\mathbf{v} = (-5, 0, 1)$, we see that $\mathbf{x} = \mathbf{x}_0 + x_2\mathbf{u} + x_3\mathbf{v}$, where x_2 and x_3 are arbitrary scalars. This is in accordance with our original definition of a plane in $\mathbb{R}^3$. ▲

As in the case of lines in $\mathbb{R}^2$, the *distance from the origin to the* (closest point on the) *plane* $\mathbf{a}\cdot\mathbf{x}=c$ is

$$\frac{|c|}{\|\mathbf{a}\|}.$$

Again, note that the point on the plane closest to the origin is

$$\frac{c}{\|\mathbf{a}\|^2}\mathbf{a},$$

which is the point where the line through the origin with direction vector $\mathbf{a}$ intersects the plane, as shown in Figure 3.5. (Indeed, the origin, this point, and any other point $\mathbf{b}$ on the plane form a right triangle, and the hypotenuse of that right triangle has length $\|\mathbf{b}\|$.)

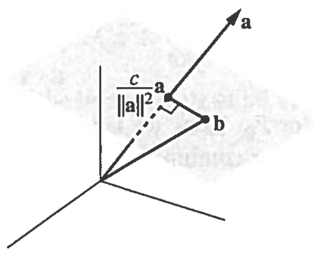

FIGURE 3.5

Finally, generalizing to n dimensions, if $\mathbf{a}\in\mathbb{R}^n$ is a nonzero vector and $c\in\mathbb{R}$, then the equation

$$\mathbf{a}\cdot\mathbf{x}=c$$

defines a *hyperplane* in $\mathbb{R}^n$. As we shall see in Chapter 3, this means that the solution set has "dimension" $n-1$, i.e., 1 less than the dimension of the ambient space $\mathbb{R}^n$. Let's write an explicit formula for the general vector $\mathbf{x}$ satisfying this equation: If $\mathbf{a}=(a_1,a_2,\dots,a_n)$ and $a_1\neq 0$, then we rewrite the equation

$$a_1x_1+a_2x_2+\dots+a_nx_n=c$$

to solve for x_1:

$$x_1=\frac{1}{a_1}(c-a_2x_2-\dots-a_nx_n),$$

and so the *general solution* is of the form

$$\begin{aligned}\mathbf{x}=(x_1,\dots,x_n)&=\left(\frac{1}{a_1}(c-a_2x_2-\dots-a_nx_n),x_2,\dots,x_n\right)\\&=\left(\frac{c}{a_1},0,\dots,0\right)+x_2\left(-\frac{a_2}{a_1},1,0,\dots,0\right)+x_3\left(-\frac{a_3}{a_1},0,1,\dots,0\right)\\&\qquad+\dots+x_n\left(-\frac{a_n}{a_1},0,\dots,0,1\right).\end{aligned}$$

(We leave it to the reader to write down the formula in the event that $a_1=0$.)

EXAMPLE 4

Consider the hyperplane

$$x_1+x_2-x_3+2x_4+x_5=2$$

in $\mathbb{R}^5$. Then a parametric description of the general solution of this equation can be written as follows:

$$\begin{aligned}\mathbf{x}&=(-x_2+x_3-2x_4-x_5+2,x_2,x_3,x_4,x_5)\\&=(2,0,0,0,0)+x_2(-1,1,0,0,0)+x_3(1,0,1,0,0)\\&\qquad+x_4(-2,0,0,1,0)+x_5(-1,0,0,0,1).\end{aligned}$$

▲

To close this section, let's consider the set of simultaneous solutions of two linear equations in $\mathbb{R}^3$, i.e., the intersection of two planes:

$$\begin{aligned}\mathbf{a}\cdot\mathbf{x}&=a_1x_1+a_2x_2+a_3x_3=c\\\mathbf{b}\cdot\mathbf{x}&=b_1x_1+b_2x_2+b_3x_3=d.\end{aligned}$$

If a vector $\mathbf{x}$ satisfies both equations, then the point (x_1,x_2,x_3) must lie on both the planes; i.e., it lies in the intersection of the planes. Geometrically, we see that there are three possibilities, as illustrated in Figure 3.6:

1. *A plane*: In this case, both equations describe the same plane.
2. *The empty set*: In this case, the equations describe parallel planes.
3. *A line*: This is the expected situation.

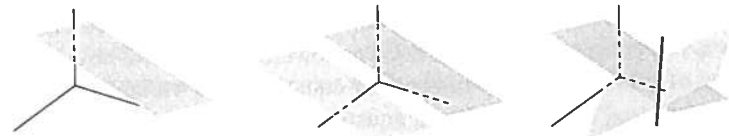

FIGURE 3.6

Notice that if the two planes are identical or parallel, then the normal vectors will be the same (up to a scalar multiple). That is, there will be a nonzero real number r so that $r\mathbf{a}=\mathbf{b}$; if we multiply the equation

$$\mathbf{a}\cdot\mathbf{x}=a_1x_1+a_2x_2+a_3x_3=c$$

by r, we get

$$\mathbf{b}\cdot\mathbf{x}=r\mathbf{a}\cdot\mathbf{x}=b_1x_1+b_2x_2+b_3x_3=rc.$$

If a point (x_1,x_2,x_3) satisfying this equation is also to satisfy the equation

$$\mathbf{b}\cdot\mathbf{x}=b_1x_1+b_2x_2+b_3x_3=d,$$

then we must have $d=rc$; i.e., the two planes coincide. On the other hand, if $d\neq rc$, then there is no solution of the pair of equations, and the two planes are parallel.

More interestingly, if the normal vectors $\mathbf{a}$ and $\mathbf{b}$ are nonparallel, then the planes intersect in a line, and that line is described as the set of solutions of the simultaneous equations. Geometrically, the direction vector of the line must be orthogonal to both $\mathbf{a}$ and $\mathbf{b}$.

EXAMPLE 5

We give a parametric description of the line of intersection of the planes

$$\begin{aligned}x_1+2x_2-x_3&=2\\x_1-x_2+2x_3&=5.\end{aligned}$$

Subtracting the first equation from the second yields

$$\begin{aligned}-3x_2+3x_3&=3,\quad\text{or}\\-x_2+x_3&=1.\end{aligned}$$

Adding twice the latter equation to the first equation in the original system yields

$$x_1+x_3=4.$$

Thus, we can determine both x_1 and x_2 in terms of x_3:

$$\begin{aligned} x_1 &= 4 - x_3 \\ x_2 &= -1 + x_3. \end{aligned}$$

Then the general solution is of the form

$$\mathbf{x} = (x_1, x_2, x_3) = (4 - x_3, -1 + x_3, x_3) = (4, -1, 0) + x_3(-1, 1, 1).$$

Indeed, as we mentioned earlier, the direction vector $(-1, 1, 1)$ is orthogonal to $\mathbf{a} = (1, 2, -1)$ and $\mathbf{b} = (1, -1, 2)$. ▲

Much of the remainder of this course will be devoted to understanding higher-dimensional analogues of lines and planes in $\mathbb{R}^3$. In particular, we will be concerned with the relation between their parametric description and their description as the set of solutions of a system of linear equations (geometrically, the intersection of a collection of hyperplanes). The first step toward this goal will be to develop techniques and notation for solving systems of m linear equations in n variables (as in Example 5, where we solved a system of two linear equations in three variables). This is the subject of the next section.

Exercises 1.3

1. Give Cartesian equations of the given hyperplanes:
a. $\mathbf{x} = (-1, 2) + t(3, 2)$
*b. The plane passing through $(1, 2, 2)$ and orthogonal to the line $\mathbf{x} = (5, 1, -1) + t(-1, 1, -1)$
c. The plane passing through $(2, 0, 1)$ and orthogonal to the line $\mathbf{x} = (2, -1, 3) + t(1, 2, -2)$
*d. The plane spanned by $(1, 1, 1)$ and $(2, 1, 0)$ and passing through $(1, 1, 2)$
e. The plane spanned by $(1, 0, 1)$ and $(1, 2, 2)$ and passing through $(-1, 1, 1)$
*f. The hyperplane in $\mathbb{R}^4$ through the origin spanned by $(1, -1, 1, -1)$, $(1, 1, -1, -1)$, and $(1, -1, -1, 1)$.

***2.** Redo Exercise 1.1.12 by finding Cartesian equations of the respective planes.

3. Find the general solution of each of the following equations (presented, as in the text, as a combination of an appropriate number of vectors).
a. $x_1 - 2x_2 + 3x_3 = 4$ (in $\mathbb{R}^3$)
b. $x_1 + x_2 - x_3 + 2x_4 = 0$ (in $\mathbb{R}^4$)
*c. $x_1 + x_2 - x_3 + 2x_4 = 5$ (in $\mathbb{R}^4$)
*d. $x_1 - 2x_2 + 3x_3 = 4$ (in $\mathbb{R}^4$)
e. $x_2 + x_3 - 3x_4 = 2$ (in $\mathbb{R}^4$)

4. Find a normal vector to the given hyperplane and use it to find the distance from the origin to the hyperplane.
a. $\mathbf{x} = (-1, 2) + t(3, 2)$
b. The plane in $\mathbb{R}^3$ given by the equation $2x_1 + x_2 - x_3 = 5$
*c. The plane passing through $(1, 2, 2)$ and orthogonal to the line $\mathbf{x} = (3, 1, -1) + t(-1, 1, -1)$
d. The plane passing through $(2, -1, 1)$ and orthogonal to the line $\mathbf{x} = (3, 1, 1) + t(-1, 2, 1)$
*e. The plane spanned by $(1, 1, 4)$ and $(2, 1, 0)$ and passing through $(1, 1, 2)$
f. The plane spanned by $(1, 1, 1)$ and $(2, 1, 0)$ and passing through $(3, 0, 2)$

g. The hyperplane in $\mathbb{R}^4$ spanned by $(1, -1, 1, -1)$, $(1, 1, -1, -1)$, and $(1, -1, -1, 1)$ and passing through $(2, 1, 0, 1)$

5. Find parametric equations of the line of intersection of the given planes in $\mathbb{R}^3$.
a. $x_1 + x_2 + x_3 = 1$, $2x_1 + x_2 + 2x_3 = 1$
b. $x_1 - x_2 = 1$, $x_1 + x_2 + 2x_3 = 5$

***6.** a. Give the general solution of the equation $x_1 + 5x_2 - 2x_3 = 0$ in $\mathbb{R}^3$ (as a linear combination of two vectors, as in the text).
b. Find a specific solution of the equation $x_1 + 5x_2 - 2x_3 = 3$ in $\mathbb{R}^3$; give the general solution.
c. Give the general solution of the equation $x_1 + 5x_2 - 2x_3 + x_4 = 0$ in $\mathbb{R}^4$. Now give the general solution of the equation $x_1 + 5x_2 - 2x_3 + x_4 = 3$.

***7.** The equation $2x_1 - 3x_2 = 5$ defines a line in $\mathbb{R}^2$.
a. Give a normal vector $\mathbf{a}$ to the line.
b. Find the distance from the origin to the line by using projection.
c. Find the point on the line closest to the origin by using the parametric equation of the line through $\mathbf{0}$ with direction vector $\mathbf{a}$. Double-check your answer to part *b*.
d. Find the distance from the point $\mathbf{w} = (3, 1)$ to the line by using projection.
e. Find the point on the line closest to $\mathbf{w}$ by using the parametric equation of the line through $\mathbf{w}$ with direction vector $\mathbf{a}$. Double-check your answer to part *d*.

8. The equation $2x_1 - 3x_2 - 6x_3 = -4$ defines a plane in $\mathbb{R}^3$.
a. Give its normal vector $\mathbf{a}$.
b. Find the distance from the origin to the plane by using projection.
c. Find the point on the plane closest to the origin by using the parametric equation of the line through $\mathbf{0}$ with direction vector $\mathbf{a}$. Double-check your answer to part *b*.
d. Find the distance from the point $\mathbf{w} = (3, -3, -5)$ to the plane by using projection.
e. Find the point on the plane closest to $\mathbf{w}$ by using the parametric equation of the line through $\mathbf{w}$ with direction vector $\mathbf{a}$. Double-check your answer to part *d*.

9. The equation $2x_1 + 2x_2 - 3x_3 + 8x_4 = 6$ defines a hyperplane in $\mathbb{R}^4$.
a. Give a normal vector $\mathbf{a}$ to the hyperplane.
b. Find the distance from the origin to the hyperplane using projection.
c. Find the point on the hyperplane closest to the origin by using the parametric equation of the line through $\mathbf{0}$ with direction vector $\mathbf{a}$. Double-check your answer to part *b*.
d. Find the distance from the point $\mathbf{w} = (1, 1, 1, 1)$ to the hyperplane using dot products.
e. Find the point on the hyperplane closest to $\mathbf{w}$ by using the parametric equation of the line through $\mathbf{w}$ with direction vector $\mathbf{a}$. Double-check your answer to part *d*.

10. *a. The equations $x_1 = 0$ and $x_2 = 0$ describe planes in $\mathbb{R}^3$ that contain the x_3-axis. Write down the Cartesian equation of a general such plane.
b. The equations $x_1 - x_2 = 0$ and $x_1 - x_3 = 0$ describe planes in $\mathbb{R}^3$ that contain the line through the origin with direction vector $(1, 1, 1)$. Write down the Cartesian equation of a general such plane.

11. a. Assume $\mathbf{b}$ and $\mathbf{c}$ are nonparallel vectors in $\mathbb{R}^3$. Generalizing the result of Exercise 10, show that the plane $\mathbf{a} \cdot \mathbf{x} = 0$ contains the intersection of the planes $\mathbf{b} \cdot \mathbf{x} = 0$ and $\mathbf{c} \cdot \mathbf{x} = 0$ if and only if $\mathbf{a} = s\mathbf{b} + t\mathbf{c}$ for some $s, t \in \mathbb{R}$, not both 0. Describe this result geometrically.
b. Assume $\mathbf{b}$ and $\mathbf{c}$ are nonparallel vectors in $\mathbb{R}^n$. Formulate a conjecture about which hyperplanes $\mathbf{a} \cdot \mathbf{x} = 0$ in $\mathbb{R}^n$ contain the intersection of the hyperplanes $\mathbf{b} \cdot \mathbf{x} = 0$ and $\mathbf{c} \cdot \mathbf{x} = 0$. Prove as much of your conjecture as you can.

12. Suppose $\mathbf{a} \neq \mathbf{0}$ and $\mathcal{P} \subset \mathbb{R}^3$ is the plane through the origin with normal vector $\mathbf{a}$. Suppose $\mathcal{P}$ is spanned by $\mathbf{u}$ and $\mathbf{v}$.

a. Suppose $\mathbf{u} \cdot \mathbf{v} = 0$. Show that for every $\mathbf{x} \in \mathcal{P}$, we have

$$\mathbf{x} = \mathrm{proj}_{\mathbf{u}}\mathbf{x} + \mathrm{proj}_{\mathbf{v}}\mathbf{x}.$$

b. Suppose $\mathbf{u} \cdot \mathbf{v} = 0$. Show that for every $\mathbf{x} \in \mathbb{R}^3$, we have

$$\mathbf{x} = \mathrm{proj}_{\mathbf{a}}\mathbf{x} + \mathrm{proj}_{\mathbf{u}}\mathbf{x} + \mathrm{proj}_{\mathbf{v}}\mathbf{x}.$$

(*Hint:* Apply part *a* to the vector $\mathbf{x} - \mathrm{proj}_{\mathbf{a}}\mathbf{x}$.)

c. Give an example to show the result of part *a* is false when $\mathbf{u}$ and $\mathbf{v}$ are not orthogonal.

13. Consider the line ℓ in $\mathbb{R}^3$ given parametrically by $\mathbf{x} = \mathbf{x}_0 + t\mathbf{a}$. Let $\mathcal{P}_0$ denote the plane through the origin with normal vector $\mathbf{a}$ (so it is orthogonal to ℓ).

a. Show that ℓ and $\mathcal{P}_0$ intersect in the point $\mathbf{x}_0 - \mathrm{proj}_{\mathbf{a}}\mathbf{x}_0$.

b. Conclude that the distance from the origin to ℓ is $\|\mathbf{x}_0 - \mathrm{proj}_{\mathbf{a}}\mathbf{x}_0\|$.

4 Systems of Linear Equations and Gaussian Elimination

In this section we give an explicit algorithm for solving a system of m linear equations in n variables. Unfortunately, this is a little bit like giving the technical description of tying a shoe—it is much easier to do it than to read how to do it. For that reason, before embarking on the technicalities of the process, we will present here a few examples and introduce the notation of matrices. On the other hand, once the technique is mastered, it will be important for us to understand why it yields *all* solutions of the system of equations. For this reason, it is essential to understand Theorem 4.1.

To begin with, a linear equation in the n variables $x_1, x_2, \dots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where the *coefficients* a_i, $i = 1, \dots, n$, are fixed real numbers and b is a fixed real number. Notice that if we let $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{x} = (x_1, \dots, x_n)$, then we can write this equation in vector notation as

$$\mathbf{a} \cdot \mathbf{x} = b.$$

We recognize this as the equation of a hyperplane in $\mathbb{R}^n$, and a vector $\mathbf{x}$ solves the equation precisely when the point $\mathbf{x}$ lies on that hyperplane.

A system of m linear equations in n variables consists of m such equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

The notation appears cumbersome, but we have to live with it. A pair of subscripts is needed on the coefficient a_{ij} to indicate in which equation it appears (the first index, i) and to which variable it is associated (the second index, j). A solution $\mathbf{x} = (x_1, \dots, x_n)$ is an n-tuple of real numbers that satisfies all m of the equations. Thus, a solution gives a point in the intersection of the m hyperplanes.

To *solve* a system of linear equations, we want to give a complete *parametric* description of the solutions, as we did for hyperplanes and for the intersection of two planes in Example 5 in the preceding section. We will call this the *general solution* of the system. Some systems are relatively simple to solve. For example, the system

$$\begin{aligned} x_1 \quad\quad\quad &= 1 \\ x_2 \quad\quad &= 2 \\ x_3 &= -1 \end{aligned}$$

has exactly one solution, namely $\mathbf{x} = (1, 2, -1)$. This is the only point common to the three planes described by the three equations. A slightly more complicated example is

$$\begin{aligned} x_1 \quad - x_3 &= 1 \\ x_2 + 2x_3 &= 2. \end{aligned}$$

These equations enable us to determine x_1 and x_2 in terms of x_3; in particular, we can write $x_1 = 1 + x_3$ and $x_2 = 2 - 2x_3$, where x_3 is *free* to take on any real value. Thus, any solution of this system is of the form

$$\mathbf{x} = (1 + t, 2 - 2t, t) = (1, 2, 0) + t(1, -2, 1) \quad \text{for some } t \in \mathbb{R}.$$

It is easily checked that every vector of this form is in fact a solution, as $(1 + t) - t = 1$ and $(2 - 2t) + 2t = 2$ for every $t \in \mathbb{R}$. Thus, we see that the intersection of the two given planes is the line in $\mathbb{R}^3$ passing through $(1, 2, 0)$ with direction vector $(1, -2, 1)$.

One should note that in the preceding example, we chose to solve for x_1 and x_2 in terms of x_3. We could just as well have solved, say, for x_2 and x_3 in terms of x_1 by first writing $x_3 = x_1 - 1$ and then substituting to obtain $x_2 = 4 - 2x_1$. Then we would end up writing

$$\mathbf{x} = (s, 4 - 2s, -1 + s) = (0, 4, -1) + s(1, -2, 1) \quad \text{for some } s \in \mathbb{R}.$$

We will soon give an algorithm for solving systems of linear equations that will eliminate the ambiguity in deciding which variables should be taken as parameters. The variables that are allowed to vary freely (as parameters) are called *free variables*, and the remaining variables, which can be expressed in terms of the free variables, are called *pivot variables*. Broadly speaking, if there are m equations, whenever possible we will try to solve for the first m variables (assuming there are that many) in terms of the remaining variables. This is not always possible (for example, the first variable may not even appear in any of the equations), so we will need to specify a general procedure to select which will be pivot variables and which will be free.

When we are solving a system of equations, there are three basic algebraic operations we can perform that will not affect the solution set. They are the following *elementary operations*:

(i) Interchange any pair of equations.

(ii) Multiply any equation by a nonzero real number.

(iii) Replace any equation by its sum with a multiple of any other equation.

The first two are probably so obvious that it seems silly to write them down; however, soon you will see their importance. It is not obvious that the third operation does not change the solution set; we will address this in Theorem 4.1. First, let's consider an example of solving a system of linear equations using these operations.

EXAMPLE 1

Consider the system of linear equations

$$\begin{aligned} 3x_1 - 2x_2 + 2x_3 + 9x_4 &= 4 \\ 2x_1 + 2x_2 - 2x_3 - 4x_4 &= 6. \end{aligned}$$

We can use operation (i) to replace this system with

$$\begin{aligned} 2x_1 + 2x_2 - 2x_3 - 4x_4 &= 6 \\ 3x_1 - 2x_2 + 2x_3 + 9x_4 &= 4; \end{aligned}$$

then we use operation (ii), multiplying the first equation by $1/2$, to get

$$\begin{aligned} x_1 + x_2 - x_3 - 2x_4 &= 3 \\ 3x_1 - 2x_2 + 2x_3 + 9x_4 &= 4; \end{aligned}$$

now we use operation (iii), adding -3 times the first equation to the second:

$$\begin{aligned} x_1 + x_2 - x_3 - 2x_4 &= 3 \\ - 5x_2 + 5x_3 + 15x_4 &= -5. \end{aligned}$$

Next we use operation (ii) again, multiplying the second equation by $-1/5$, to obtain

$$\begin{aligned} x_1 + x_2 - x_3 - 2x_4 &= 3 \\ x_2 - x_3 - 3x_4 &= 1; \end{aligned}$$

finally, we use operation (iii), adding -1 times the second equation to the first:

$$\begin{aligned} x_1 \qquad\quad + x_4 &= 2 \\ x_2 - x_3 - 3x_4 &= 1. \end{aligned}$$

From this we see that x_1 and x_2 are determined by x_3 and x_4, both of which are free to take on any values. Thus, we read off the general solution of the system of equations:

$$\begin{aligned} x_1 &= 2 \qquad\quad - x_4 \\ x_2 &= 1 + x_3 + 3x_4 \\ x_3 &= \qquad x_3 \\ x_4 &= \qquad\qquad x_4 \end{aligned}$$

In vector form, the general solution is

$$\mathbf{x} = (x_1, x_2, x_3, x_4) = (2, 1, 0, 0) + x_3(0, 1, 1, 0) + x_4(-1, 3, 0, 1),$$

which is the parametric representation of a plane in $\mathbb{R}^4$. ▲

Before describing the algorithm for solving a general system of linear equations, we want to introduce some notation to make the calculations less cumbersome to write out. We begin with a system of m equations in n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

We can simplify our notation somewhat by writing the equations in vector notation:

$$\begin{aligned} \mathbf{A}_1 \cdot \mathbf{x} &= b_1 \\ \mathbf{A}_2 \cdot \mathbf{x} &= b_2 \\ &\vdots \\ \mathbf{A}_m \cdot \mathbf{x} &= b_m, \end{aligned}$$

where $\mathbf{A}_i = (a_{i1}, a_{i2}, \dots, a_{in}) \in \mathbb{R}^n$, $i = 1, 2, \dots, m$. To simplify the notation further, we introduce the $m \times n$ (read "m by n") *matrix*[9]

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

and the *column vectors*[10]

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \quad\text{and}\quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m,$$

and write our equations as

$$A\mathbf{x} = \mathbf{b},$$

where the multiplication on the left-hand side is defined to be

$$A\mathbf{x} = \begin{bmatrix} \mathbf{A}_1 \cdot \mathbf{x} \\ \mathbf{A}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}.$$

We will discuss the algebraic and geometric properties of matrices a bit later, but for now we simply use them as convenient shorthand notation for systems of equations. We emphasize that an $m \times n$ matrix has m rows and n columns. The coefficient a_{ij} appearing in the i^{th} row and the j^{th} column is called the ij-entry of A. We say that two matrices are *equal* if they have the same *shape* (that is, if they have equal numbers of rows and equal numbers of columns) and their corresponding entries are equal. As we did above, we will customarily denote the *row vectors* of the matrix A by $\mathbf{A}_1, \dots, \mathbf{A}_m \in \mathbb{R}^n$.

We reiterate that a solution $\mathbf{x}$ of the system of equations $A\mathbf{x} = \mathbf{b}$ is a vector having the requisite dot products with the row vectors $\mathbf{A}_i$:

$$\mathbf{A}_i \cdot \mathbf{x} = b_i \quad \text{for all } i = 1, 2, \dots, m.$$

That is, the system of equations describes the intersection of the m hyperplanes with normal vectors $\mathbf{A}_i$ and at (signed) distance $b_i/\|\mathbf{A}_i\|$ from the origin. To give the general solution, we must find a parametric representation of this intersection.

[9]The word *matrix* derives from the Latin *mātrix*, "womb" (originally, "pregnant animal"), from *māter*, "mother."

[10]We shall henceforth try to write vectors as columns, unless doing so might cause undue typographical hardship.

Notice that the first two types of elementary operations do not change this collection of hyperplanes, so it is no surprise that these operations do not affect the solution set of the system of equations. On the other hand, the third type actually changes one of the hyperplanes without changing the intersection. To see why, suppose $\mathbf{a}$ and $\mathbf{b}$ are nonparallel and consider the pairs of equations

$$\begin{aligned} \mathbf{a}\cdot\mathbf{x} &= 0 \\ \mathbf{b}\cdot\mathbf{x} &= 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} (\mathbf{a}+c\mathbf{b})\cdot\mathbf{x} &= 0 \\ \mathbf{b}\cdot\mathbf{x} &= 0. \end{aligned}$$

Suppose $\mathbf{x}$ satisfies the first set of equations, so $\mathbf{a}\cdot\mathbf{x}=0$ and $\mathbf{b}\cdot\mathbf{x}=0$; then $\mathbf{x}$ satisfies the second set as well, since $(\mathbf{a}+c\mathbf{b})\cdot\mathbf{x} = (\mathbf{a}\cdot\mathbf{x}) + c(\mathbf{b}\cdot\mathbf{x}) = 0 + c0 = 0$ and $\mathbf{b}\cdot\mathbf{x}=0$ remains true. Conversely, if $\mathbf{x}$ satisfies the second set of equations, we have $\mathbf{b}\cdot\mathbf{x}=0$ and $\mathbf{a}\cdot\mathbf{x} = (\mathbf{a}+c\mathbf{b})\cdot\mathbf{x} - c(\mathbf{b}\cdot\mathbf{x}) = 0 - c0 = 0$, so $\mathbf{x}$ also satisfies the first set. Thus the solution sets are identical. Geometrically, as shown in Figure 4.1, taking a bit of poetic license, we can think of the hyperplanes $\mathbf{a}\cdot\mathbf{x}=0$ and $\mathbf{b}\cdot\mathbf{x}=0$ as the covers of a book, and the solutions $\mathbf{x}$ will form the "spine" of the book. The typical equation $(\mathbf{a}+c\mathbf{b})\cdot\mathbf{x}=0$ describes one of the pages of the book, and that page intersects either of the covers precisely in the same spine. This follows from the fact that the spine consists of all vectors orthogonal to the plane spanned by $\mathbf{a}$ and $\mathbf{b}$; this is identical to the plane spanned by $\mathbf{a}+c\mathbf{b}$ and $\mathbf{b}$ (or $\mathbf{a}$).

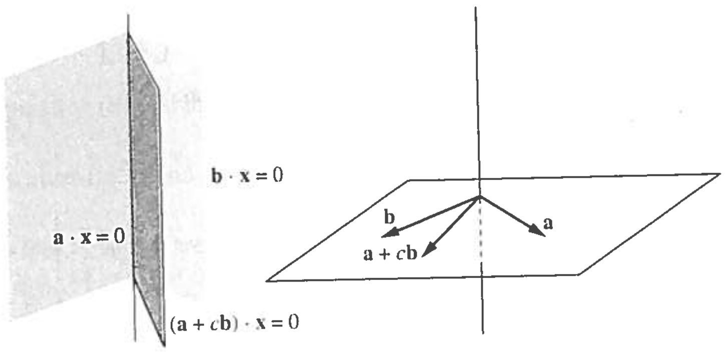

FIGURE 4.1

The general result is the following:

Theorem 4.1. *If a system of equations $A\mathbf{x}=\mathbf{b}$ is changed into the new system $C\mathbf{x}=\mathbf{d}$ by elementary operations, then the systems have the same set of solutions.*

Proof. We need to show that every solution of $A\mathbf{x}=\mathbf{b}$ is also a solution of $C\mathbf{x}=\mathbf{d}$, and vice versa. Start with a solution $\mathbf{u}$ of $A\mathbf{x}=\mathbf{b}$. Denoting the rows of A by $\mathbf{A}_1,\dots,\mathbf{A}_m$, we have

$$\begin{aligned} \mathbf{A}_1\cdot\mathbf{u} &= b_1 \\ \mathbf{A}_2\cdot\mathbf{u} &= b_2 \\ &\vdots \\ \mathbf{A}_m\cdot\mathbf{u} &= b_m. \end{aligned}$$

If we apply an elementary operation of type (i), $\mathbf{u}$ still satisfies precisely the same list of equations. If we apply an elementary operation of type (ii), say multiplying the k^{th} equation by $r \neq 0$, we note that if $\mathbf{u}$ satisfies $\mathbf{A}_k\cdot\mathbf{u}=b_k$, then it must satisfy $(r\mathbf{A}_k)\cdot\mathbf{u}=rb_k$. As for an elementary operation of type (iii), suppose we add r times the k^{th} equation to the ℓ^{th}; since $\mathbf{A}_k\cdot\mathbf{u}=b_k$ and $\mathbf{A}_\ell\cdot\mathbf{u}=b_\ell$, it follows that

$$(r\mathbf{A}_k+\mathbf{A}_\ell)\cdot\mathbf{u} = (r\mathbf{A}_k\cdot\mathbf{u}) + (\mathbf{A}_\ell\cdot\mathbf{u}) = rb_k + b_\ell,$$

and so $\mathbf{u}$ satisfies the "new" ℓ^{th} equation.

To prove conversely that if $\mathbf{u}$ satisfies $C\mathbf{x}=\mathbf{d}$, then it satisfies $A\mathbf{x}=\mathbf{b}$, we merely note that each argument we've given can be reversed; in particular, *the inverse of an elementary operation is again an elementary operation*. Note that it is important here that $r\neq 0$ for an operation of type (ii). □

We introduce one further piece of shorthand notation, the *augmented matrix*

$$[\,A \mid \mathbf{b}\,] = \left[\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{21} & \dots & a_{2n} & b_2 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right].$$

Notice that the augmented matrix contains all of the information of the original system of equations, because we can recover the latter by filling in the x_i's, $+$'s, and $=$'s as needed.

The elementary operations on a system of equations become operations on the rows of the augmented matrix; in this setting, we refer to them as *elementary row operations* of the corresponding three types:

(i) Interchange any pair of rows.

(ii) Multiply all the entries of any row by a nonzero real number.

(iii) Replace any row by its sum with a multiple of any other row.

Since we have established that elementary operations do not affect the solution set of a system of equations, we can freely perform elementary row operations on the augmented matrix of a system of equations with the goal of finding an "equivalent" augmented matrix from which we can easily read off the general solution.

EXAMPLE 2

We revisit Example 1 in the notation of augmented matrices. To solve

$$\begin{aligned} 3x_1 - 2x_2 + 2x_3 + 9x_4 &= 4 \\ 2x_1 + 2x_2 - 2x_3 - 4x_4 &= 6, \end{aligned}$$

we begin by forming the appropriate augmented matrix

$$\left[\begin{array}{cccc|c} 3 & -2 & 2 & 9 & 4 \\ 2 & 2 & -2 & -4 & 6 \end{array}\right].$$

We denote the process of performing row operations by the symbol $\rightsquigarrow$ and (in this example) we indicate above it the type of operation we are performing:

$$\left[\begin{array}{cccc|c} 3 & -2 & 2 & 9 & 4 \\ 2 & 2 & -2 & -4 & 6 \end{array}\right] \overset{\text{(i)}}{\rightsquigarrow} \left[\begin{array}{cccc|c} 2 & 2 & -2 & -4 & 6 \\ 3 & -2 & 2 & 9 & 4 \end{array}\right] \overset{\text{(ii)}}{\rightsquigarrow} \left[\begin{array}{cccc|c} 1 & 1 & -1 & -2 & 3 \\ 3 & -2 & 2 & 9 & 4 \end{array}\right]$$

$$\overset{\text{(iii)}}{\rightsquigarrow} \left[\begin{array}{cccc|c} 1 & 1 & -1 & -2 & 3 \\ 0 & -5 & 5 & 15 & -5 \end{array}\right] \overset{\text{(ii)}}{\rightsquigarrow} \left[\begin{array}{cccc|c} 1 & 1 & -1 & -2 & 3 \\ 0 & 1 & -1 & -3 & 1 \end{array}\right] \overset{\text{(iii)}}{\rightsquigarrow} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & -1 & -3 & 1 \end{array}\right].$$

From the final augmented matrix we are able to recover the simpler form of the equations,

$$\begin{aligned} x_1 \qquad\qquad + x_4 &= 2 \\ x_2 - x_3 - 3x_4 &= 1, \end{aligned}$$

and read off the general solution just as before. ▲

Remark. It is important to distinguish between the symbols $=$ and $\rightsquigarrow$; when we convert one matrix to another by performing one or more row operations, we do *not* have equal matrices.

To recap, we have discussed the elementary operations that can be performed on a system of linear equations without changing the solution set, and we have introduced the shorthand notation of augmented matrices. To proceed, we need to discuss the final form our system should have in order for us to be able to read off the solutions easily. To understand this goal, let's consider a few more examples.

EXAMPLE 3

(a) Consider the system

$$\begin{aligned} x_1 + 2x_2 \qquad - x_4 &= 1 \\ x_3 + 2x_4 &= 2. \end{aligned}$$

We see that using the second equation, we can determine x_3 in terms of x_4 and that using the first, we can determine x_1 in terms of x_2 and x_4. In particular, the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1-2x_2+x_4 \\ x_2 \\ 2 \quad -2x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}.$$

(b) The system

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

requires some algebraic manipulation before we can read off the solution. Although the second equation determines x_3 in terms of x_4, the first describes x_1 in terms of x_2, x_3, and x_4; but x_2, x_3, and x_4 are not *all* allowed to vary arbitrarily: We would like to modify the first equation by removing x_3. Indeed, if we subtract the second equation from the first, we will recover the system in (a).

(c) The system

$$\begin{aligned} x_1 + 2x_2 \qquad &= 3 \\ x_1 \qquad - x_3 &= 2 \end{aligned}$$

involves similar difficulties. The value of x_1 seems to be determined, on the one hand, by x_2 and, on the other, by x_3; this is problematic (try $x_2 = 1$ and $x_3 = 3$). Indeed, we

recognize that this system of equations describes the intersection of two planes in $\mathbb{R}^3$ (that are distinct and not parallel); this should be a line, whose parametric expression should depend on only one variable. The point is that we cannot choose both x_2 and x_3 to be free variables. We first need to manipulate the system of equations so that we can determine one of them in terms of the other (for example, we might subtract the first equation from the second). ▲

The point of this discussion is to use elementary row operations to manipulate systems of linear equations like those in Examples 3(b) and (c) above into equivalent systems from which the solutions can be easily recognized, as in Example 3(a). But what distinguishes Example 3(a)?

Definition. We call the first *nonzero* entry of a row (reading left to right) its *leading entry*. A matrix is in *echelon*[11] *form* if

1. The leading entries move to the right in successive rows.
2. The entries of the column *below* each leading entry are all 0.[12]
3. All rows of 0's are at the bottom of the matrix.

A matrix is in *reduced echelon form* if it is in echelon form and, in addition,

4. Every leading entry is 1.
5. All the entries of the column *above* each leading entry are 0 as well.

If a matrix is in echelon form, we call the leading entry of any (nonzero) row a *pivot*. We refer to the columns in which a pivot appears as *pivot columns* and to the corresponding variables (in the original system of equations) as *pivot variables*. The remaining variables are called *free variables*.

What do we learn from the respective augmented matrices for our earlier examples?

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 1 & 2 & 2 \end{array}\right], \quad \left[\begin{array}{cccc|c} 1 & 2 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 & 1 \end{array}\right], \quad \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 1 & 0 & -1 & 2 \end{array}\right]$$

Of the augmented matrices from Example 3, (a) is in reduced echelon form, (b) is in echelon form, and (c) is in neither. The key point is this: When the matrix is in reduced echelon form, we are able to determine the general solution by expressing each of the *pivot* variables in terms of the *free* variables.

[11]The word *echelon* derives from the French *échelle*, "ladder." Although we don't usually draw the rungs of the ladder, they are there: $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}$. OK, perhaps it looks more like a staircase.

[12]Condition **2** is actually a consequence of **1**, but we state it anyway for clarity.

Here are a few further examples.

EXAMPLE 4

The matrix

$$\begin{bmatrix} 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 3 & 0 & 2 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

is in echelon form. The pivot variables are x_2, x_3, and x_4; the free variables are x_1 and x_5. However, the matrix

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

is not in echelon form, because the row of 0's is not at the bottom; the matrix

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 4 \\ 0 & 0 & 3 & 0 & 2 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$$

is not in echelon form, since the entry below the leading entry of the second row is nonzero. And the matrix

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

in also not in echelon form, because the leading entries do not move to the right. ▲

EXAMPLE 5

The augmented matrix

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{array}\right]$$

is in reduced echelon form. The corresponding system of equations is

$$\begin{aligned} x_1 + 2x_2 \qquad\qquad + 4x_5 &= 1 \\ x_3 \qquad - 2x_5 &= 2 \\ x_4 + \ x_5 &= 1. \end{aligned}$$

Notice that the pivot variables, x_1, x_3, and x_4, are completely determined by the free variables x_2 and x_5. As usual, we can write the general solution in terms of the free variables only:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1-2x_2-4x_5 \\ x_2 \\ 2 \quad +2x_5 \\ 1 \quad - x_5 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -4 \\ 0 \\ 2 \\ -1 \\ 1 \end{bmatrix}.$$

▲

We stop for a moment to formalize the manner in which we have expressed the parametric form of the general solution of a system of linear equations once it's been put in *reduced echelon form*.

> ***Definition.*** We say that we've written the general solution in *standard form* when it is expressed as the sum of a *particular solution*—obtained by setting all the free variables equal to 0—and a linear combination of vectors, one for each free variable—obtained by setting that free variable equal to 1 and the remaining free variables equal to 0 and ignoring the particular solution.[13]

Our strategy now is to transform the augmented matrix of any system of linear equations into echelon form by performing a sequence of elementary row operations. The algorithm goes by the name of *Gaussian elimination*. The first step is to identify the first column (starting at the left) that does not consist only of 0's; usually this is the first column, but it may not be. Pick a row whose entry in this column is nonzero—usually the uppermost such row, but you may choose another if it helps with the arithmetic—and interchange this with the first row; now the first entry of the first nonzero column is nonzero. This will be our first *pivot*. Next, we add the appropriate multiple of the top row to all the remaining rows to make all the entries below the pivot equal to 0. For example, if we begin with the matrix

$$A = \begin{bmatrix} 3 & -1 & 2 & 7 \\ 2 & 1 & 3 & 3 \\ 2 & 2 & 4 & 2 \end{bmatrix},$$

then we can switch the first and third rows of A (to avoid fractions) and clear out the first pivot column to obtain

$$A' = \begin{bmatrix} 2 & 2 & 4 & 2 \\ 0 & -1 & -1 & 1 \\ 0 & -4 & -4 & 4 \end{bmatrix}.$$

We have circled the pivot for emphasis. (If we are headed for the reduced echelon form, we might replace the first row of A' by $(1, 1, 2, 1)$, but this can wait.)

The next step is to find the first column (again, starting at the left) in the *new* matrix having a nonzero entry *below the first row*. Pick a row below the first that has a nonzero entry in this column, and, if necessary, interchange it with the second row. Now the second entry of this column is nonzero; this is our second pivot. (Once again, if we're calculating the reduced echelon form, we multiply by the reciprocal of this entry to make the pivot 1.) We then add appropriate multiples of the second row to the rows beneath it to make all the

[13] In other words, if x_j is a free variable, the corresponding vector in the general solution has j^{th} coordinate equal to 1 and k^{th} coordinate equal to 0 for all the other free variables x_k. Concentrate on the circled entries in the vectors from Example 5:

$$x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -4 \\ 0 \\ 2 \\ -1 \\ 1 \end{bmatrix}.$$

entries beneath the pivot equal to 0. Continuing with our example, we obtain

$$A'' = \begin{bmatrix} 2 & 2 & 4 & 2 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

At this point, A'' is in echelon form; note that the zero row is at the bottom and that the pivots move toward the right and down.

In general, the process continues until we can find no more pivots—either because we have a pivot in each row or because we're left with nothing but rows of zeroes. At this stage, if we are interested in finding the reduced echelon form, we clear out the entries in the pivot columns *above* the pivots and then make all the pivots equal to 1. (A few words of advice here: If we start at the *right* and work our way up and to the left, we in general minimize the amount of arithmetic that must be done. Also, we always do our best to avoid fractions.) Continuing with our example, we find that the reduced echelon form of A is

$$A'' = \begin{bmatrix} 2 & 2 & 4 & 2 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R_A.$$

It should be evident that there are many choices involved in the process of Gaussian elimination. For example, at the outset, we chose to interchange the first and third rows of A. We might just as well have used either the first or the second row to obtain our first pivot, but we chose the third because we noticed that it would simplify the arithmetic to do so. This lack of specificity in our algorithm is perhaps disconcerting at first, because we are afraid that we might make the "wrong" choice. But so long as we choose a row with a nonzero entry in the appropriate column, we can proceed. It's just a matter of making the arithmetic more or less convenient, and—as in our experience with techniques of integration—practice brings the ability to make more advantageous choices.

Given all the choices we make along the way, we might wonder whether we always arrive at the same answer. Evidently, the echelon form may well depend on the choices. But despite the fact that a matrix may have lots of different echelon forms, they all must have the same number of *nonzero rows*; that number is called the *rank* of the matrix.

Proposition 4.2. *All echelon forms of an $m \times n$ matrix A have the same number of nonzero rows.*

Proof. Suppose B and C are two echelon forms of A, and suppose C has (at least) one more row of zeroes than B. Because there is a pivot in each nonzero row, there is (at least) one pivot variable for B that is a free variable for C, say x_j. Since x_j is a free variable for C, there is a vector $\mathbf{v} = (a_1, a_2, a_3, \dots, a_{j-1}, 1, 0, \dots, 0)$ that satisfies $C\mathbf{v} = \mathbf{0}$. We obtain this vector by setting $x_j = 1$ and the other free variables (for C) equal to 0, and then solving for the remaining (pivot) variables.[14]

On the other hand, x_j is a pivot variable for B; assume that it is the pivot in the ℓ^{th} row. That is, the first nonzero entry of the ℓ^{th} row of B occurs in the j^{th} column. Then the ℓ^{th}

[14]To see why $\mathbf{v}$ has this form, we must understand why the k^{th} entry of $\mathbf{v}$ is 0 whenever $k > j$. So suppose $k > j$. If x_k is a free variable, then by construction the k^{th} entry of $\mathbf{v}$ is 0. On the other hand, if x_k is a pivot variable, then the value of x_k is determined *only* by the values of the pivot variables x_ℓ with $\ell > k$; since, by construction, these are all 0, once again, the k^{th} entry of $\mathbf{v}$ is 0.

entry of $B\mathbf{v}$ is 1. This contradicts Theorem 4.1, for if $C\mathbf{v} = \mathbf{0}$, then $A\mathbf{v} = \mathbf{0}$, and so $B\mathbf{v} = \mathbf{0}$ as well. □

In fact, it is not difficult to see that more is true, as we ask the ambitious reader to check in Exercise 16:

Theorem 4.3. *Each matrix has a unique* reduced *echelon form.*

We conclude with a few examples illustrating Gaussian elimination and its applications.

EXAMPLE 6

Give the general solution of the following system of linear equations:

$$\begin{aligned} x_1 + x_2 + 3x_3 - x_4 \phantom{{}+2x_5} &= 0 \\ -x_1 + x_2 + x_3 + x_4 + 2x_5 &= -4 \\ x_2 + 2x_3 + 2x_4 - x_5 &= 0 \\ 2x_1 - x_2 \phantom{{}+2x_3} + x_4 - 6x_5 &= 9. \end{aligned}$$

We begin with the augmented matrix of coefficients and put it in reduced echelon form:

$$\left[\begin{array}{ccccc|c} 1 & 1 & 3 & -1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 2 & -4 \\ 0 & 1 & 2 & 2 & -1 & 0 \\ 2 & -1 & 0 & 1 & -6 & 9 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 1 & 3 & -1 & 0 & 0 \\ 0 & 2 & 4 & 0 & 2 & -4 \\ 0 & 1 & 2 & 2 & -1 & 0 \\ 0 & -3 & -6 & 3 & -6 & 9 \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 1 & 3 & -1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & -2 \\ 0 & 0 & 0 & 2 & -2 & 2 \\ 0 & 0 & 0 & 3 & -3 & 3 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 1 & 3 & -1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -2 & 3 \\ 0 & 1 & 2 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Thus, the system of equations is given in reduced echelon form by

$$\begin{aligned} x_1 \phantom{{}+x_2} + x_3 \phantom{{}+x_4} - 2x_5 &= 3 \\ x_2 + 2x_3 \phantom{{}+x_4} + x_5 &= -2 \\ x_4 - x_5 &= 1, \end{aligned}$$

from which we read off

$$\begin{aligned} x_1 &= 3 - x_3 + 2x_5 \\ x_2 &= -2 - 2x_3 - x_5 \\ x_3 &= \phantom{-2 - 2{}} x_3 \\ x_4 &= 1 \phantom{{}- 2x_3} + x_5 \\ x_5 &= \phantom{-2 - 2x_3 - {}} x_5, \end{aligned}$$

and so the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

▲

EXAMPLE 7

We wish to find a normal vector to the hyperplane in $\mathbb{R}^4$ spanned by the vectors $\mathbf{v}_1 = (1, 0, 1, 0)$, $\mathbf{v}_2 = (0, 1, 0, 1)$, and $\mathbf{v}_3 = (1, 2, 3, 4)$. That is, we want a vector $\mathbf{x} \in \mathbb{R}^4$ satisfying the system of equations $\mathbf{v}_1 \cdot \mathbf{x} = \mathbf{v}_2 \cdot \mathbf{x} = \mathbf{v}_3 \cdot \mathbf{x} = 0$. Such a vector $\mathbf{x}$ must satisfy the system of equations

$$\begin{aligned} x_1 \quad\quad + x_3 \quad\quad &= 0 \\ x_2 \quad\quad + x_4 &= 0 \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 0. \end{aligned}$$

Putting the augmented matrix in reduced echelon form, we find

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 2 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right].$$

From this we read off

$$\begin{aligned} x_1 &= x_4 \\ x_2 &= -x_4 \\ x_3 &= -x_4 \\ x_4 &= x_4, \end{aligned}$$

and so the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix};$$

that is, a normal vector to the plane is (any nonzero scalar multiple of) $(1, -1, -1, 1)$. The reader should check that this vector actually *is* orthogonal to the three given vectors. ▲

Recalling that solving the system of linear equations

$$\mathbf{A}_1 \cdot \mathbf{x} = b_1, \quad \mathbf{A}_2 \cdot \mathbf{x} = b_2, \quad \ldots, \quad \mathbf{A}_m \cdot \mathbf{x} = b_m$$

amounts to finding a parametric representation of the intersection of these m hyperplanes, we consider one last example.

EXAMPLE 8

We seek a parametric description of the intersection of the three hyperplanes in $\mathbb{R}^4$ given by

$$\begin{aligned} x_1 - x_2 + 2x_3 + 3x_4 &= 2 \\ 2x_1 + x_2 + x_3 \quad\quad &= 1 \\ x_1 + 2x_2 - x_3 - 3x_4 &= 7. \end{aligned}$$

Again, we start with the augmented matrix and put it in echelon form:

$$\left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 2 & 1 & 1 & 0 & 1 \\ 1 & 2 & -1 & -3 & 7 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 3 & -3 & -6 & -3 \\ 0 & 3 & -3 & -6 & 5 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 3 & -3 & -6 & -3 \\ 0 & 0 & 0 & 0 & 8 \end{array}\right].$$

Without even continuing to reduced echelon form, we see that the new augmented matrix gives the system of equations

$$\begin{aligned} x_1 - x_2 + 2x_3 + 3x_4 &= 2 \\ 3x_2 - 3x_3 - 6x_4 &= -3 \\ 0 &= 8. \end{aligned}$$

The last equation, $0 = 8$, is, of course, absurd. What happened? There can be no values of x_1, x_2, x_3, and x_4 that make this system of equations hold: The three hyperplanes described by our equations have no point in common. A system of linear equations may not have any solutions; in this case it is called *inconsistent*. We study this notion carefully in the next section. ▲

Exercises 1.4

1. Use elementary operations to find the general solution of each of the following systems of equations. Use the method of Example 1 as a prototype.

a. $$\begin{aligned} x_1 + x_2 \quad\quad &= 1 \\ x_1 + 2x_2 + x_3 &= 1 \\ x_2 + 2x_3 &= 1 \end{aligned}$$

*b. $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1 \\ 2x_1 + 4x_2 + 5x_3 &= 1 \\ x_1 + 2x_2 + 2x_3 &= 0 \end{aligned}$$

c. $$\begin{aligned} 3x_1 - 6x_2 - x_3 + x_4 &= 6 \\ -x_1 + 2x_2 + 2x_3 + 3x_4 &= 3 \\ 4x_1 - 8x_2 - 3x_3 - 2x_4 &= 3 \end{aligned}$$

*2. Decide which of the following matrices are in echelon form, which are in reduced echelon form, and which are neither. Justify your answers.

a. $\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$

b. $\begin{bmatrix} 2 & 1 & 3 \\ 0 & 1 & -1 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \end{bmatrix}$

d. $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

e. $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

f. $\begin{bmatrix} 1 & 1 & 0 & -1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ g. $\begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$

3. For each of the following matrices A, determine its reduced echelon form and give the general solution of $A\mathbf{x} = \mathbf{0}$ in standard form.

*a. $A = \begin{bmatrix} 1 & 0 & -1 \\ -2 & 3 & -1 \\ 3 & -3 & 0 \end{bmatrix}$

*b. $A = \begin{bmatrix} 2 & -2 & 4 \\ -1 & 1 & -2 \\ 3 & -3 & 6 \end{bmatrix}$

c. $A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 3 & 1 \\ 2 & 4 & 3 \\ -1 & 1 & 6 \end{bmatrix}$

d. $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 \\ 1 & 3 & 2 & 4 \\ 1 & 2 & 2 & 3 \end{bmatrix}$

*e. $A = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 2 & -4 & 3 & -1 \end{bmatrix}$

f. $A = \begin{bmatrix} 1 & 2 & 0 & -1 & -1 \\ -1 & -3 & 1 & 2 & 3 \\ 1 & -1 & 3 & 1 & 1 \\ 2 & -3 & 7 & 3 & 4 \end{bmatrix}$

*g. $A = \begin{bmatrix} 1 & -1 & 1 & 1 & 0 \\ 1 & 0 & 2 & 1 & 1 \\ 0 & 2 & 2 & 2 & 0 \\ -1 & 1 & -1 & 0 & -1 \end{bmatrix}$

h. $A = \begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 \\ 0 & 1 & 1 & 3 & -2 & 0 \\ -1 & 2 & 3 & 4 & 1 & -6 \\ 0 & 4 & 4 & 12 & -1 & -7 \end{bmatrix}$

4. Give the general solution of the equation $A\mathbf{x} = \mathbf{b}$ in standard form.

*a. $A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 0 \\ -3 \end{bmatrix}$

b. $A = \begin{bmatrix} 2 & -1 \\ 2 & 1 \\ -1 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix}$

*c. $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix}$

d. $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$

e. $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 3 & 3 & 2 & 0 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ 17 \end{bmatrix}$

f. $A = \begin{bmatrix} 1 & 1 & 1 & -1 & 0 \\ 2 & 0 & 4 & 1 & -1 \\ 1 & 2 & 0 & -2 & 2 \\ 0 & 1 & -1 & 2 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ 10 \\ -3 \\ 7 \end{bmatrix}$

5. For the following matrices A, give the general solution of the equation $A\mathbf{x} = \mathbf{x}$ in standard form. (*Hint:* Rewrite this as $B\mathbf{x} = \mathbf{0}$ for an appropriate matrix B.)

a. $A = \begin{bmatrix} 10 & -6 \\ 18 & -11 \end{bmatrix}$ *b. $A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 2 \\ -2 & 0 & 3 \end{bmatrix}$ c. $A = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}$

6. For the following matrices A, give the general solution of the equation $A\mathbf{x} = 2\mathbf{x}$ in standard form.

a. $A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$ b. $A = \begin{bmatrix} 3 & 16 & -15 \\ 1 & 12 & -9 \\ 1 & 16 & -13 \end{bmatrix}$

7. One might need to find solutions of $A\mathbf{x} = \mathbf{b}$ for several different $\mathbf{b}$'s, say $\mathbf{b}_1, \dots, \mathbf{b}_k$. In this event, one can augment the matrix A with all the $\mathbf{b}$'s simultaneously, forming the "multi-augmented" matrix $[\, A \mid \mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_k \,]$. One can then read off the various solutions from the reduced echelon form of the multi-augmented matrix. Use this method to solve $A\mathbf{x} = \mathbf{b}_j$ for the given matrices A and vectors $\mathbf{b}_j$.

a. $A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -1 \\ -1 & 2 & 2 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 3 & 2 & 1 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

c. $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

*8. Find all the unit vectors $\mathbf{x} \in \mathbb{R}^3$ that make an angle of $\pi/3$ with each of the vectors $(1, 0, -1)$ and $(0, 1, 1)$.

9. Find all the unit vectors $\mathbf{x} \in \mathbb{R}^3$ that make an angle of $\pi/4$ with $(1, 0, 1)$ and an angle of $\pi/3$ with $(0, 1, 0)$.

10. Find a normal vector to the hyperplane in $\mathbb{R}^4$ spanned by
*a. $(1, 1, 1, 1)$, $(1, 2, 1, 2)$, and $(1, 3, 2, 4)$;
b. $(1, 1, 1, 1)$, $(2, 2, 1, 2)$, and $(1, 3, 2, 3)$.

11. Find all vectors $\mathbf{x} \in \mathbb{R}^4$ that are orthogonal to both
*a. $(1, 0, 1, 1)$ and $(0, 1, -1, 2)$;
b. $(1, 1, 1, -1)$ and $(1, 2, -1, 1)$.

12. Find all the unit vectors in $\mathbb{R}^4$ that make an angle of $\pi/3$ with $(1, 1, 1, 1)$ and an angle of $\pi/4$ with both $(1, 1, 0, 0)$ and $(1, 0, 0, 1)$.

♯*13. Let A be an $m \times n$ matrix, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and let c be a scalar. Show that
a. $A(c\mathbf{x}) = c(A\mathbf{x})$
b. $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$

14. Let A be an $m \times n$ matrix, and let $\mathbf{b} \in \mathbb{R}^m$.

a. Show that if $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^n$ are both solutions of $A\mathbf{x} = \mathbf{b}$, then $\mathbf{u} - \mathbf{v}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

b. Suppose $\mathbf{u}$ is a solution of $A\mathbf{x} = \mathbf{0}$ and $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$. Show that $\mathbf{u} + \mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

Hint: Use Exercise 13.

15. a. Prove or give a counterexample: If A is an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$, then either every entry of A is 0 or $\mathbf{x} = \mathbf{0}$.

b. Prove or give a counterexample: If A is an $m \times n$ matrix and $A\mathbf{x} = \mathbf{0}$ for every vector $\mathbf{x} \in \mathbb{R}^n$, then every entry of A is 0.

Although an example does not constitute a proof, a *counterexample* is a fine disproof: A counterexample is merely an explicit example illustrating that the statement is false.

Here, the evil authors are asking you first to decide whether the statement is true or false. It is important to try examples to develop your intuition. In a problem like this that contains arbitrary positive integers m and n, it is often good to start with small values. Of course, if we take $m = n = 1$, we get the statement

If a is a real number and $ax = 0$ for every real number x, then $a = 0$.

Here you might say, "Well, if $a \neq 0$, then I can divide both sides of the equation by a and obtain $x = 0$. Since the equation must hold for *all* real numbers x, we must have $a = 0$." But this doesn't give us any insight into the general case, as we can't divide by vectors or matrices.

What are some alternative approaches? You might try picking a particular value of x that will shed light on the situation. For example, if we take $x = 1$, then we immediately get $a = 0$. How might you use this idea to handle the general case? If you wanted to show that a particular entry, say a_{25}, of the matrix A was 0, could you pick the vector $\mathbf{x}$ appropriately?

There's another way to pick a particular value of x that leads to information. Since the only given object in the problem is the real number a, we might try letting $x = a$ and see what happens. Here we get $ax = a^2 = 0$, from which we conclude immediately that $a = 0$. How does this idea help us with the general case? Remember that the entries of the vector $A\mathbf{x}$ are the dot products $\mathbf{A}_i \cdot \mathbf{x}$. Looking back at part *a* of Exercise 1.2.16, we learned there that if $\mathbf{a} \cdot \mathbf{x} = 0$ for all $\mathbf{x}$, then $\mathbf{a} = \mathbf{0}$. How does our current path of reasoning lead us to this?

16. Prove that the reduced echelon form of a matrix is unique, as follows. Suppose B and C are reduced echelon forms of a given nonzero $m \times n$ matrix A.

a. Deduce from the proof of Proposition 4.2 that B and C have the same pivot variables.

b. Explain why the pivots of B and C are in the identical positions. (This is true even without the assumption that the matrices are in *reduced* echelon form.)

c. By considering the solutions in standard form of $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$, deduce that $B = C$.

17. In rating the efficiency of different computer algorithms for solving a system of equations, it is usually considered sufficient to compare the number of multiplications required to carry out the algorithm.

a. Show that

$$n(n-1) + (n-1)(n-2) + \cdots + (2)(1) = \sum_{k=1}^{n}(k^2 - k)$$

multiplications are required to bring a general $n \times n$ matrix to echelon form by (forward) Gaussian elimination.

b. Show that $\sum_{k=1}^{n}(k^2 - k) = \frac{1}{3}(n^3 - n)$. (*Hint:* For some appropriate formulas, see Exercise 1.2.10.)

c. Now show that it takes $n + (n-1) + (n-2) + \cdots + 1 = n(n+1)/2$ multiplications to bring the matrix to reduced echelon form by clearing out the columns above the pivots, working right to left. Show that it therefore takes a total of $\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$ multiplications to put A in reduced echelon form.

d. Gauss-Jordan elimination is a slightly different algorithm used to bring a matrix to reduced echelon form: Here each column is cleared out, both below and above the pivot, before moving on to the next column. Show that in general this procedure requires $n^2(n-1)/2$ multiplications. For large n, which method is preferred?

5 The Theory of Linear Systems

We developed Gaussian elimination as a technique for finding a parametric description of the solutions of a system of linear Cartesian equations. Now we shall see that this same technique allows us to proceed in the opposite direction. That is, given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, we would like to find a set of Cartesian equations whose solution is precisely Span $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. In addition, we will rephrase in somewhat more general terms the observations we have already made about solutions of systems of linear equations.

5.1 Existence, Constraint Equations, and Rank

Suppose A is an $m \times n$ matrix. There are two equally important ways to interpret the system of equations $A\mathbf{x} = \mathbf{b}$. In the preceding section, we concentrated on the row vectors of A: If $\mathbf{A}_1, \ldots, \mathbf{A}_m$ denote the *row vectors* of A, then the vector $\mathbf{c}$ is a solution of $A\mathbf{x} = \mathbf{b}$ if and only if

$$\mathbf{A}_1 \cdot \mathbf{c} = b_1, \quad \mathbf{A}_2 \cdot \mathbf{c} = b_2, \quad \ldots, \quad \mathbf{A}_m \cdot \mathbf{c} = b_m.$$

Geometrically, $\mathbf{c}$ is a solution precisely when it lies in the intersection of all the hyperplanes defined by the system of equations.

On the other hand, we can define the *column vectors* of the $m \times n$ matrix A as follows:

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{R}^m, \quad j = 1, 2, \ldots, n.$$

We now make an observation that will be crucial in our future work: The matrix product $A\mathbf{x}$ can also be written as

$$(*) \quad A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

$$= x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Thus, a solution $\mathbf{c} = (c_1, \dots, c_n)$ of the linear system $A\mathbf{x} = \mathbf{b}$ provides scalars $c_1, \dots, c_n$ so that

$$\mathbf{b} = c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n.$$

This is our second geometric interpretation of the system of linear equations: A solution $\mathbf{c}$ gives a representation of the vector $\mathbf{b}$ as a linear combination, $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$, of the column vectors of A.

EXAMPLE 1

Consider the four vectors

$$\mathbf{b} = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}.$$

Suppose we want to express the vector $\mathbf{b}$ as a linear combination of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Writing out the expression

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = x_1 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 2 \end{bmatrix},$$

we obtain the system of equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= 4 \\ x_2 + x_3 &= 3 \\ x_1 + x_2 + x_3 &= 1 \\ 2x_1 + x_2 + 2x_3 &= 2. \end{aligned}$$

In matrix notation, we must solve $A\mathbf{x} = \mathbf{b}$, where the columns of A are $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 2 \end{bmatrix}.$$

So we take the augmented matrix to reduced echelon form:

$$[A \mid \mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 3 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 2 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & -1 & -3 \\ 0 & -1 & -2 & -6 \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

This tells us that the solution is

$$\mathbf{x} = \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}, \quad \text{so} \quad \mathbf{b} = -2\mathbf{v}_1 + 0\mathbf{v}_2 + 3\mathbf{v}_3,$$

which, as the reader can check, works. ▲

Now we modify the preceding example slightly.

EXAMPLE 2

We would like to express the vector

$$\mathbf{b}' = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

as a linear combination of the same vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. This then leads analogously to the system of equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= 1 \\ x_2 + x_3 &= 1 \\ x_1 + x_2 + x_3 &= 0 \\ 2x_1 + x_2 + 2x_3 &= 1 \end{aligned}$$

and to the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 2 & 1 & 2 & 1 \end{array}\right],$$

whose echelon form is

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

The last row of the augmented matrix corresponds to the equation

$$0x_1 + 0x_2 + 0x_3 = 1,$$

which obviously has no solution. Thus, the original system of equations has no solution: The vector $\mathbf{b}'$ in this example cannot be written as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. ▲

These examples lead us to make the following definition.

Definition. If the system of equations $A\mathbf{x} = \mathbf{b}$ has no solutions, the system is said to be *inconsistent*; if it has at least one solution, then it is said to be *consistent*.

A system of equations is consistent precisely when a solution *exists*. We see that the system of equations in Example 2 is inconsistent and the system of equations in Example 1 is consistent. It is easy to recognize an inconsistent system of equations from the echelon form of its augmented matrix: The system is inconsistent precisely when there is an equation that reads

$$0x_1 + 0x_2 + \cdots + 0x_n = c$$

for some nonzero scalar c, i.e., when there is a row in the echelon form of the augmented matrix all of whose entries are 0 except for the rightmost.

Turning this around a bit, let $[\,U \mid \mathbf{c}\,]$ denote an echelon form of the augmented matrix $[\,A \mid \mathbf{b}\,]$. The system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if any zero row in U corresponds to a zero entry in the vector $\mathbf{c}$.

There are two geometric interpretations of consistency. From the standpoint of row vectors, the system $A\mathbf{x} = \mathbf{b}$ is consistent precisely when the intersection of the hyperplanes

$$\mathbf{A}_1 \cdot \mathbf{x} = b_1, \quad \ldots, \quad \mathbf{A}_m \cdot \mathbf{x} = b_m$$

is nonempty. From the point of view of column vectors, the system $A\mathbf{x} = \mathbf{b}$ is consistent precisely when the vector $\mathbf{b}$ can be written as a linear combination of the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of A; in other words, it is consistent when $\mathbf{b} \in \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

In the next example, we characterize those vectors $\mathbf{b} \in \mathbb{R}^4$ that can be expressed as a linear combination of the three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ from Examples 1 and 2.

EXAMPLE 3

For what vectors

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

will the system of equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= b_1 \\ x_2 + x_3 &= b_2 \\ x_1 + x_2 + x_3 &= b_3 \\ 2x_1 + x_2 + 2x_3 &= b_4 \end{aligned}$$

have a solution? We form the augmented matrix $[\,A \mid \mathbf{b}\,]$ and put it in echelon form:

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & b_1 \\ 0 & 1 & 1 & b_2 \\ 1 & 1 & 1 & b_3 \\ 2 & 1 & 2 & b_4 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & b_1 \\ 0 & 1 & 1 & b_2 \\ 0 & 0 & -1 & b_3 - b_1 \\ 0 & -1 & -2 & b_4 - 2b_1 \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & b_1 \\ 0 & 1 & 1 & b_2 \\ 0 & 0 & 1 & b_1 - b_3 \\ 0 & 0 & 0 & -b_1 + b_2 - b_3 + b_4 \end{array}\right].$$

We deduce that the original system of equations will have a solution if and only if

$$(**) \qquad -b_1 + b_2 - b_3 + b_4 = 0.$$

That is, the vector $\mathbf{b}$ belongs to $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ precisely when $\mathbf{b}$ satisfies the *constraint equation* (**). Changing letters slightly, we infer that a Cartesian equation of the hyperplane spanned by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in $\mathbb{R}^4$ is $-x_1 + x_2 - x_3 + x_4 = 0$. ▲

EXAMPLE 4

As a further example, we take

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 3 & 2 & -1 \\ 1 & 4 & -3 \\ 3 & -3 & 3 \end{bmatrix},$$

and we look for constraint equations that describe the vectors $\mathbf{b} \in \mathbb{R}^4$ for which $A\mathbf{x} = \mathbf{b}$ is consistent, i.e., all vectors $\mathbf{b}$ that can be expressed as a linear combination of the columns of A.

As before, we consider the augmented matrix $[\,A \mid \mathbf{b}\,]$ and determine an echelon form $[\,U \mid \mathbf{c}\,]$. In order for the system to be consistent, every entry of $\mathbf{c}$ corresponding to a row of 0's in U must be 0 as well:

$$[\,A \mid \mathbf{b}\,] = \left[\begin{array}{ccc|c} 1 & -1 & 1 & b_1 \\ 3 & 2 & -1 & b_2 \\ 1 & 4 & -3 & b_3 \\ 3 & -3 & 3 & b_4 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & b_1 \\ 0 & 5 & -4 & b_2 - 3b_1 \\ 0 & 5 & -4 & b_3 - b_1 \\ 0 & 0 & 0 & b_4 - 3b_1 \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & b_1 \\ 0 & 5 & -4 & b_2 - 3b_1 \\ 0 & 0 & 0 & b_3 - b_2 + 2b_1 \\ 0 & 0 & 0 & b_4 - 3b_1 \end{array}\right].$$

Here we have two rows of 0's in U, so we conclude that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ satisfies the two *constraint equations*

$$2b_1 - b_2 + b_3 = 0 \quad \text{and} \quad -3b_1 + b_4 = 0.$$

These equations describe the intersection of two hyperplanes through the origin in $\mathbb{R}^4$ with respective normal vectors $(2, -1, 1, 0)$ and $(-3, 0, 0, 1)$. ▲

Notice that in the last two examples, we have reversed the process of Sections 3 and 4. There we expressed the general solution of a system of linear equations as a linear combination of certain vectors, just as we described lines, planes, and hyperplanes parametrically earlier. Here, starting with the column vectors of the matrix A, we have found the *constraint equations* that a vector $\mathbf{b}$ must satisfy in order to be a linear combination of them (that is, to be in their span). This is the process of determining Cartesian equations of a space that is defined parametrically.

Remark. It is worth noting that since A has different echelon forms, one can arrive at different constraint equations. We will investigate this more deeply in Chapter 3.

EXAMPLE 5

Find a Cartesian equation of the plane in $\mathbb{R}^3$ given parametrically by

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

We ask which vectors $\mathbf{b} = (b_1, b_2, b_3)$ can be written in the form

$$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

This system of equations can be rewritten as

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} b_1 - 1 \\ b_2 - 2 \\ b_3 - 1 \end{bmatrix},$$

and so we want to know when this system of equations is consistent. Reducing the augmented matrix to echelon form, we have

$$\left[\begin{array}{cc|c} 1 & 2 & b_1 - 1 \\ 0 & 1 & b_2 - 2 \\ 1 & 1 & b_3 - 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 2 & b_1 - 1 \\ 0 & 1 & b_2 - 2 \\ 0 & 0 & b_3 - b_1 + b_2 - 2 \end{array}\right].$$

Thus, the constraint equation is $-b_1 + b_2 + b_3 - 2 = 0$. A Cartesian equation of the given plane is $x_1 - x_2 - x_3 = -2$. ▲

In general, given an $m \times n$ matrix, we might wonder how many conditions a vector $\mathbf{b} \in \mathbb{R}^m$ must satisfy in order to be a linear combination of the columns of A. From the procedure we've just followed, the answer is quite clear: Each row of 0's in the echelon form of A contributes one constraint. This leads us to our next definition.

> ***Definition.*** The *rank* of a matrix A is the number of nonzero rows (the number of pivots) in any echelon form of A. It is usually denoted by r.

Then the number of rows of 0's in the echelon form is $m - r$, and $\mathbf{b}$ must satisfy $m - r$ constraint equations. Note that it is a consequence of Proposition 4.2 that the rank of a matrix is well-defined, i.e., independent of the choice of echelon form.

Now, given a system of m linear equations in n variables, let A denote its coefficient matrix and r the rank of A. We summarize the above remarks as follows.

Proposition 5.1. *The linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if the rank of the augmented matrix $[\,A \mid \mathbf{b}\,]$ equals the rank of A. In particular, when the rank of A equals m, the system $A\mathbf{x} = \mathbf{b}$ will be consistent for all vectors $\mathbf{b} \in \mathbb{R}^m$.*

Proof. Let $[\,U \mid \mathbf{c}\,]$ denote the echelon form of the augmented matrix $[\,A \mid \mathbf{b}\,]$. We know that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if any zero row in U corresponds to a zero entry in the vector $\mathbf{c}$, which occurs if and only if the number of nonzero rows in the augmented matrix

$[\,U \mid \mathbf{c}\,]$ equals the number of nonzero rows in U, i.e., the rank of A. When $r = m$, there is no row of 0's in U and hence no possibility of inconsistency. □

5.2 Uniqueness and Nonuniqueness of Solutions

We now turn our attention to the question of how many solutions a given *consistent* system of equations has. Our experience with solving systems of equations in Sections 3 and 4 suggests that the solutions of a consistent linear system $A\mathbf{x} = \mathbf{b}$ are intimately related to the solutions of the system $A\mathbf{x} = \mathbf{0}$.

> ***Definition.*** A system $A\mathbf{x} = \mathbf{b}$ of linear equations is called *inhomogeneous* when $\mathbf{b} \neq \mathbf{0}$; the corresponding equation $A\mathbf{x} = \mathbf{0}$ is called the associated *homogeneous system.*

To relate the solutions of the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ and those of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$, we need the following fundamental algebraic observation.

Proposition 5.2. *Let A be an $m \times n$ matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then*

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}.$$

(This is called the distributive property *of matrix multiplication.)*

Proof. Recall that, by definition, the i^{th} entry of the product $A\mathbf{x}$ is equal to the dot product $\mathbf{A}_i \cdot \mathbf{x}$. The distributive property of dot product (the last property listed in Proposition 2.1) dictates that

$$\mathbf{A}_i \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{A}_i \cdot \mathbf{x} + \mathbf{A}_i \cdot \mathbf{y},$$

and so the i^{th} entry of $A(\mathbf{x} + \mathbf{y})$ equals the i^{th} entry of $A\mathbf{x} + A\mathbf{y}$. Since this holds for all $i = 1, \dots, m$, the vectors are equal. □

This argument establishes the first part of the following theorem.

Theorem 5.3. *Assume the system $A\mathbf{x} = \mathbf{b}$ is consistent, and let $\mathbf{u}_1$ be a particular solution.*[15] *Then all the solutions are of the form*

$$\mathbf{u} = \mathbf{u}_1 + \mathbf{v}$$

for some solution $\mathbf{v}$ of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.

Proof. First we observe that any such vector $\mathbf{u}$ is a solution of $A\mathbf{x} = \mathbf{b}$. Using Proposition 5.2, we have

$$A\mathbf{u} = A(\mathbf{u}_1 + \mathbf{v}) = A\mathbf{u}_1 + A\mathbf{v} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Conversely, *every* solution of $A\mathbf{x} = \mathbf{b}$ can be written in this form, for if $\mathbf{u}$ is an arbitrary solution of $A\mathbf{x} = \mathbf{b}$, then, by distributivity again,

$$A(\mathbf{u} - \mathbf{u}_1) = A\mathbf{u} - A\mathbf{u}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

so $\mathbf{v} = \mathbf{u} - \mathbf{u}_1$ is a solution of the associated homogeneous system; now we just solve for $\mathbf{u}$, obtaining $\mathbf{u} = \mathbf{u}_1 + \mathbf{v}$, as required. □

Remark. As Figure 5.1 suggests, when the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ is consistent, its solutions are obtained by *translating* the set of solutions of the associated homogeneous

[15] This is classical terminology for *any* single solution of the inhomogeneous system. There need not be anything special about it. In Example 5 on p. 44, we saw a way to pick a *particular* particular solution.

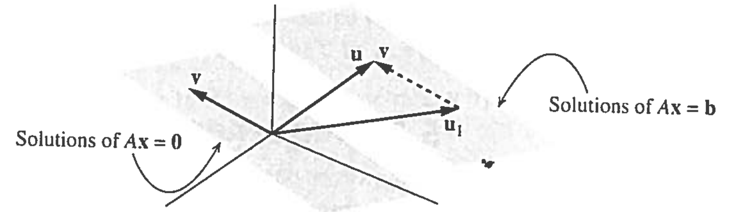

FIGURE 5.1

system by a particular solution $\mathbf{u}_1$. Since $\mathbf{u}_1$ lies on each of the hyperplanes $\mathbf{A}_i \cdot \mathbf{x} = b$ $i = 1, \dots, m$, we can translate each of the hyperplanes $\mathbf{A}_i \cdot \mathbf{x} = 0$, which pass through th origin, by the vector $\mathbf{u}_1$. Thus, translating the intersection of the hyperplanes $\mathbf{A}_i \cdot \mathbf{x} = ($ $i = 1, \dots, m$, by the vector $\mathbf{u}_1$ gives us the intersection of the hyperplanes $\mathbf{A}_i \cdot \mathbf{x} = b$ $i = 1, \dots, m$, as indicated in Figure 5.2.

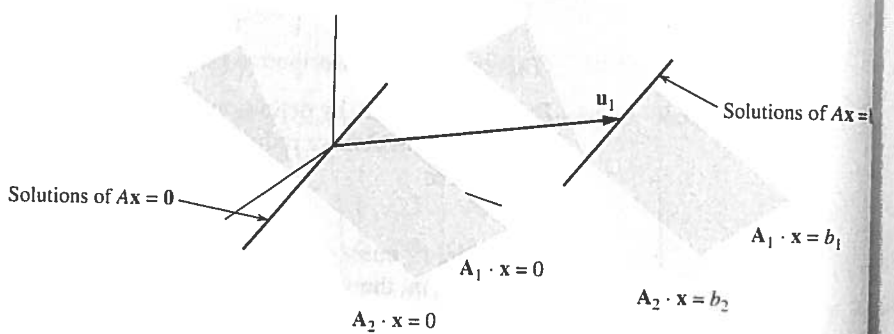

FIGURE 5.2

Of course, a homogeneous system is always consistent, because the *trivial solution*, $\mathbf{x} = \mathbf{0}$, is always a solution of $A\mathbf{x} = \mathbf{0}$. Now, if the rank of A is r, then there will be r pivot variables and $n - r$ free variables in the general solution of $A\mathbf{x} = \mathbf{0}$. In particular, if $r = n$, then $\mathbf{x} = \mathbf{0}$ is the only solution of $A\mathbf{x} = \mathbf{0}$.

> ***Definition.*** If the system of equations $A\mathbf{x} = \mathbf{b}$ has precisely one solution, then we say that the system has a *unique* solution.

Thus, a homogeneous system $A\mathbf{x} = \mathbf{0}$ has a *unique* solution when $r = n$ and *infinitely many* solutions when $r < n$. Note that it is impossible to have $r > n$, since there cannot be more pivots than columns. Similarly, there cannot be more pivots than rows in the matrix, so it follows that whenever $n > m$ (i.e., there are more variables than equations), the homogeneous system $A\mathbf{x} = \mathbf{0}$ must have infinitely many solutions.

From Theorem 5.3 we know that if the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ is consistent, then its solutions are obtained by translating the solutions of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$ by a particular solution. So we have the following proposition.

Proposition 5.4. *Suppose the system $A\mathbf{x} = \mathbf{b}$ is consistent. Then it has a unique solution if and only if the associated homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This happens exactly when $r = n$.*

We conclude this discussion with an important special case. It is natural to ask when the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ has a *unique* solution for *every* $\mathbf{b} \in \mathbb{R}^m$. From Proposition 5.1 we infer that for the system always to be consistent, we must have $r = m$; from Proposition 5.4 we infer that for solutions to be unique, we must have $r = n$. And so we see that we can have both conditions only when $r = m = n$.

> ***Definition.*** An $n \times n$ matrix of rank $r = n$ is called *nonsingular*. An $n \times n$ matrix of rank $r < n$ is called *singular*.

We observe that an $n \times n$ matrix is nonsingular if and only if there is a pivot in each row, hence in each column, of its echelon form. Thus, its reduced echelon form must be the $n \times n$ matrix

$$\begin{bmatrix} 1 & & & & \\ & 1 & & \text{\Large 0} & \\ & & \ddots & & \\ & \text{\Large 0} & & 1 & \\ & & & & 1 \end{bmatrix}.$$

It seems silly to remark that when $m = n$, if $r = n$, then $r = m$, and conversely. But the following result, which will be extremely important in the next few chapters, is an immediate consequence of this observation.

Proposition 5.5. *Let A be an $n \times n$ matrix. The following are equivalent:*

1. *A is nonsingular.*
2. *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
3. *For every $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution (indeed, a unique solution).*

Exercises 1.5

*1. By solving a system of equations, find the linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ that gives $\mathbf{b} = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix}$.

*2. For each of the following vectors $\mathbf{b} \in \mathbb{R}^4$, decide whether $\mathbf{b}$ is a linear combination of

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \quad \text{and } \mathbf{v}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}.$$

a. $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ b. $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$ c. $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -2 \end{bmatrix}$

3. Find constraint equations (if any) that $\mathbf{b}$ must satisfy in order for $A\mathbf{x} = \mathbf{b}$ to be consistent.

a. $A = \begin{bmatrix} 3 & -1 \\ 6 & -2 \\ -9 & 3 \end{bmatrix}$

*b. $A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \\ 1 & 3 & 4 \end{bmatrix}$

c. $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 1 & -1 \end{bmatrix}$

*d. $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 2 \\ 1 & -2 & 1 \end{bmatrix}$

e. $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 4 \\ -2 & -1 & 1 \end{bmatrix}$

f. $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & 3 \end{bmatrix}$

4. Find constraint equations that $\mathbf{b}$ must satisfy in order to be an element of

a. $V = \text{Span}\left((-1, 2, 1), (2, -4, -2)\right)$

b. $V = \text{Span}\left((1, 0, 1, 1), (0, 1, 1, 2), (1, 1, 1, 0)\right)$

c. $V = \text{Span}\left((1, 0, 1, 1), (0, 1, 1, 2), (2, -1, 1, 0)\right)$

d. $V = \text{Span}\left((1, 2, 3), (-1, 0, -2), (1, -2, 1)\right)$

5. By finding appropriate constraint equations, give a Cartesian equation of each of the following planes in $\mathbb{R}^3$.

a. $\mathbf{x} = s(1, -2, -2) + t(2, 0, -1),\ s, t \in \mathbb{R}$

b. $\mathbf{x} = (1, 2, 3) + s(1, -2, -2) + t(2, 0, -1),\ s, t \in \mathbb{R}$

c. $\mathbf{x} = (4, 2, 1) + s(1, 0, 1) + t(1, 2, -1),\ s, t \in \mathbb{R}$

6. Suppose A is a 3×4 matrix satisfying the equations

$$A\begin{bmatrix} 1 \\ 2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{and} \quad A\begin{bmatrix} 0 \\ 3 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Find a vector $\mathbf{x} \in \mathbb{R}^4$ such that $A\mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$. Give your reasoning. (*Hint:* Look carefully at the vectors on the right-hand side of the equations.)

7. Find a matrix A with the given property or explain why none can exist.

a. One of the rows of A is $(1, 0, 1)$, and for some $\mathbf{b} \in \mathbb{R}^2$ both the vectors $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ are solutions of the equation $A\mathbf{x} = \mathbf{b}$.

*b. The rows of A are linear combinations of $(0, 1, 0, 1)$ and $(0, 0, 1, 1)$, and for some $\mathbf{b} \in \mathbb{R}^2$ both the vectors $\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 1 \\ 0 \\ 3 \end{bmatrix}$ are solution of the equation $A\mathbf{x} = \mathbf{b}$.

c. The rows of A are orthogonal to $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, and for some nonzero vector $\mathbf{b} \in \mathbb{R}^2$ both the vectors $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ are solutions of the equation $A\mathbf{x} = \mathbf{b}$.

*d. For some vectors $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^2$ both the vectors $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ are solutions of the equation $A\mathbf{x} = \mathbf{b}_1$, and both the vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ are solutions of the equation $A\mathbf{x} = \mathbf{b}_2$.

***8.** Let $A = \begin{bmatrix} 1 & \alpha \\ \alpha & 3\alpha \end{bmatrix}$.

a. For which numbers α will A be singular?

b. For all numbers α *not* on your list in part *a*, we can solve $A\mathbf{x} = \mathbf{b}$ for every vector $\mathbf{b} \in \mathbb{R}^2$. For each of the numbers α on your list, give the vectors $\mathbf{b}$ for which we can solve $A\mathbf{x} = \mathbf{b}$.

9. Let $A = \begin{bmatrix} 1 & \alpha & \alpha \\ \alpha & 2 & 1 \\ \alpha & \alpha & 1 \end{bmatrix}$.

a. For which numbers α will A be singular?

b. For all numbers α *not* on your list in part *a*, we can solve $A\mathbf{x} = \mathbf{b}$ for every vector $\mathbf{b} \in \mathbb{R}^3$. For each of the numbers α on your list, give the vectors $\mathbf{b}$ for which we can solve $A\mathbf{x} = \mathbf{b}$.

10. Let A be an $m \times n$ matrix. Prove or give a counterexample: If $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$, then $A\mathbf{x} = \mathbf{b}$ always has a unique solution.

11. Let A and B be $m \times n$ matrices. Prove or give a counterexample: If $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solutions, then the set of vectors $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ is consistent is the same as the set of the vectors $\mathbf{b}$ such that $B\mathbf{x} = \mathbf{b}$ is consistent.

12. In each case, give positive integers m and n and an example of an $m \times n$ matrix A with the stated property, or explain why none can exist.

*a. $A\mathbf{x} = \mathbf{b}$ is inconsistent for every $\mathbf{b} \in \mathbb{R}^m$.

*b. $A\mathbf{x} = \mathbf{b}$ has one solution for every $\mathbf{b} \in \mathbb{R}^m$.

c. $A\mathbf{x} = \mathbf{b}$ has no solutions for some $\mathbf{b} \in \mathbb{R}^m$ and one solution for every other $\mathbf{b} \in \mathbb{R}^m$.

d. $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions for every $\mathbf{b} \in \mathbb{R}^m$.

*e. $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b} \in \mathbb{R}^m$ and has infinitely many solutions whenever it is consistent.

f. There are vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ so that $A\mathbf{x} = \mathbf{b}_1$ has no solution, $A\mathbf{x} = \mathbf{b}_2$ has exactly one solution, and $A\mathbf{x} = \mathbf{b}_3$ has infinitely many solutions.

♯**13.** Suppose A is an $m \times n$ matrix with rank m and $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are vectors with $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathbb{R}^n$. Prove that $\text{Span}(A\mathbf{v}_1, \dots, A\mathbf{v}_k) = \mathbb{R}^m$.

14. Let A be an $m \times n$ matrix with row vectors $\mathbf{A}_1, \dots, \mathbf{A}_m \in \mathbb{R}^n$.

*a. Suppose $\mathbf{A}_1 + \dots + \mathbf{A}_m = \mathbf{0}$. Deduce that $\text{rank}(A) < m$. (*Hint:* Why must there be a row of 0's in the echelon form of A?)

b. More generally, suppose there is some linear combination $c_1\mathbf{A}_1 + \dots + c_m\mathbf{A}_m = \mathbf{0}$, where some $c_i \neq 0$. Show that $\text{rank}(A) < m$.

15. Let A be an $m \times n$ matrix with column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$.
   a. Suppose $\mathbf{a}_1 + \cdots + \mathbf{a}_n = \mathbf{0}$. Prove that $\text{rank}(A) < n$. (*Hint:* Consider solutions of $A\mathbf{x} = \mathbf{0}$.)
   b. More generally, suppose there is some linear combination $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{0}$, where some $c_i \neq 0$. Prove that $\text{rank}(A) < n$.

6 Some Applications

We whet the reader's appetite with a few simple applications of systems of linear equations. In later chapters, when we begin to think of matrices as representing functions, we will find further applications of linear algebra.

6.1 Curve Fitting

The first application is to fitting data points to a certain class of curves.

EXAMPLE 1

We want to find the equation of the line passing through the points $(1, 1)$, $(2, 5)$, and $(-2, -11)$. Of course, none of us needs any linear algebra to solve this problem—the point-slope formula will do; but let's proceed anyhow.

We hope to find an equation of the form

$$y = mx + b$$

that is satisfied by each of the three points. (See Figure 6.1.) That gives us a system of

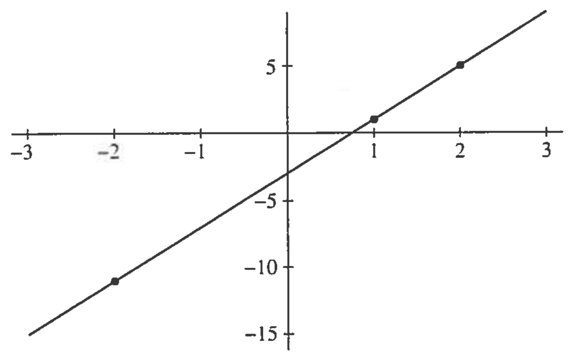

FIGURE 6.1

three equations in the two variables m and b when we substitute the respective points into the equation:

$$\begin{aligned} 1m + b &= 1 \\ 2m + b &= 5 \\ -2m + b &= -11. \end{aligned}$$

It is easy enough to solve this system of equations using Gaussian elimination:

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 2 & 1 & 5 \\ -2 & 1 & -11 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -1 & 3 \\ 0 & 3 & -9 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{array}\right],$$

and so the line we sought is $y = 4x - 3$. The reader should check that all three points indeed lie on this line. ▲

Of course, with three data points, we would expect this system of equations to be inconsistent. In Chapter 4 we will see a beautiful application of dot products and projection to find the line of regression ("least squares line") giving the best fit to the data points in that situation.

Given three points, it is plausible that if they are not collinear, then we should be able to fit a parabola

$$y = ax^2 + bx + c$$

to them (provided no two lie on a vertical line). You are asked to prove this in Exercise 7, but let's do a numerical example here.

EXAMPLE 2

Given the points $(0, 3)$, $(2, -5)$, and $(7, 10)$, we wish to find the parabola $y = ax^2 + bx + c$ passing through them. (See Figure 6.2.) Now we write down the system of equations in

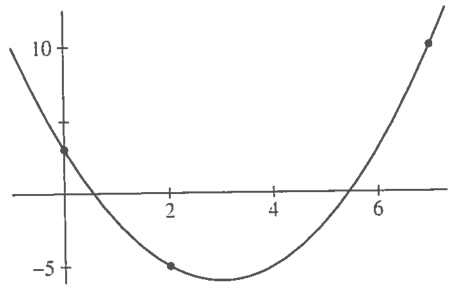

FIGURE 6.2

the variables a, b, and c:

$$\begin{aligned} 0a + 0b + c &= 3 \\ 4a + 2b + c &= -5 \\ 49a + 7b + c &= 10. \end{aligned}$$

We're supposed to solve this system by Gaussian elimination, but we can't resist the temptation to use the fact that $c = 3$ and then rewrite the remaining equations as

$$\begin{aligned} 2a + b &= -4 \\ 7a + b &= 1, \end{aligned}$$

which we can solve easily to obtain $a = 1$ and $b = -6$. Thus, our desired parabola is $y = x^2 - 6x + 3$; once again, the reader should check that each of the three data points lies on this curve. ▲

The curious reader might wonder whether, given $n + 1$ points in the plane (no two with the same x-coordinate), there is a polynomial $P(x)$ of degree at most n so that all $n + 1$ points lie on the graph $y = P(x)$. The answer is yes, as we will prove with the *Lagrange interpolation formula* in Chapter 3. It is widely used in numerical applications.