7-23-2015

Linear Algebra

Daniel Scully
djscully@stcloudstate.edu

Linear Algebra

Daniel Scully

July 23, 2015

Contents

1 Systems of Linear Equations and Matrices

1.1 Systems of Linear Equations

Definition 1.1. A **linear equation** in the variables (or unknowns) $x_1, x_2, \cdots, x_n$ is an equation that can be written in the form

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

where b and the coefficients $a_1, a_2, \cdots, a_n$ are (usually) real or complex numbers that are known.

Equations like $3(x_1 - x_3) = 4(x_2 + x_3 + 2)$ are linear because they can be algebraically rearranged into the above form (called **standard form**). For example,

$$3x_1 - 3x_3 = 4x_2 + 4x_3 + 8$$

can be rewritten in the standard form as

$$3x_1 - 4x_2 - 7x_3 = 8.$$

Equations involving powers (other than one) of variables, products of variables, variables in exponents or under radicals are usually not linear equations and are not studied here unless there is some related linear equation involved.

Example 1.1

The following equations are not linear.

$$x_1x_2 + x_3 = 5, \quad x_1^2 + x_2^2 = 5, \quad e^{x_1} + e^{x_2} = 5,$$

$$\sqrt{x_1} + \sqrt{x_2} = 1, \quad x_1 + \sin x_2 = 3$$

When only two variables are involved, we usually use x and y rather than x_1 and x_2 as variables. When three variables are used, we use x, y, and z. With four variables, we use x_1, x_2, x_3, and x_4. Of course these are not hard and fast rules.

Definition 1.2. A **solution** to a linear equation $a_1x_1 + \cdots + a_nx_n = b$ is an ordered n-tuple $(s_1, \cdots, s_n)$ of numbers with the property that if the number s_1 replaces the variable x_1 and s_2 replaces x_2 and so on in the equation, then the equation becomes a true statement. When this happens, we say the n-tuple **satisfies** the equation.

Example 1.2

The ordered pairs $(1,1), (-3,4), (0,\frac{7}{4}), (\frac{7}{3},0)$ are solutions to the equation $3x+4y=7$ (see Figure 1.1). That is, they all satisfy this equation. The ordered pairs $(0,0), (-1,2), (1,-1)$ are not solutions to this equation.

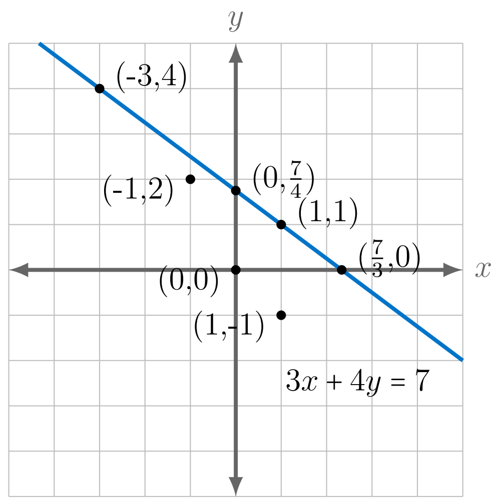


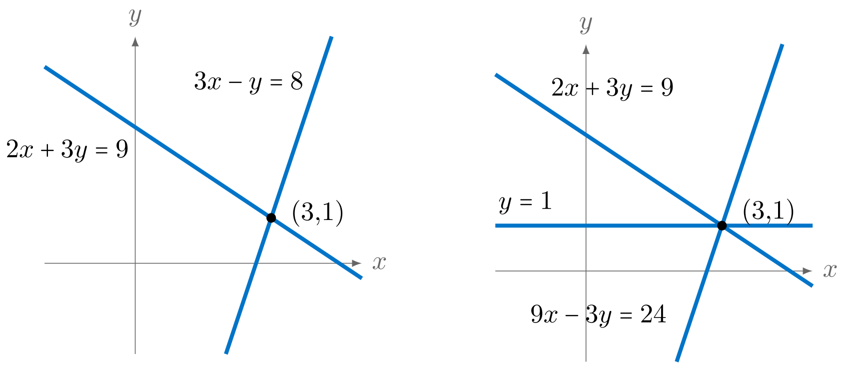
Figure 1.1 *The use of the term linear comes from the two-dimensional case.*

The name *linear* comes from the two-variable case because, as you know, the set of all solutions of a linear equation in two variables line up when graphed on the xy-plane (see Figure 1.1). We call this the **graph** of the equation. As we shall see, the graphs in 3-space of linear equations in three variables are planes. We simply don't have enough geometric dimensions to view graphs of linear equations in four or more variables. We use the phrase *linear equations* to apply to these equations regardless of the number of variables.

A **system of linear equations** is a collection of one or more linear equations. A **solution to a system** is an ordered n-tuple that satisfies *all* the equations in the system. A system is in **standard form** if each equation in the system is in standard form and like variables line up in columns. We call a system of m equation in n variables (as shown below) an $m \times n$ **system**. The numbers m and n are called the **dimensions** of the system.

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

We use double subscripts to describe the coefficients. In the above system, a_{ij} is the coefficient in the i^{th} equation of the j^{th} variable x_j.



(a) The ordered pair $(3, 1)$ satisfies the 2×2 linear system: $3x - y = 8, 2x + 3y = 9$.

(b) A 3×2 linear system with a unique solution.

Figure 1.2

Example 1.3

In the 3×3 system

$$\begin{array}{rcrcrcr} 3x & - & 2y & + & 4z & = & 3 \\ x & + & y & - & 2z & = & 1 \\ 2x & + & y & + & z & = & 5 \end{array}$$

we see that $a_{1,3} = 4$ and $b_3 = 5$.

Example 1.4

Because of the geometry of lines in a plane (see Figure 1.2a), we know that the 2×2 system

$$\begin{array}{rcrcr} 2x & + & 3y & = & 9 \\ 3x & - & y & = & 8 \end{array}$$

has exactly one solution. You have likely solved many systems like this using the method of elimination. By multiplying the second equation by 3, and adding it to the first, we get $11x = 33$, or $x = 3$. By substituting $x = 3$ into the first equation and solving we get $y = 1$. That is, $(3, 1)$ is the only solution.

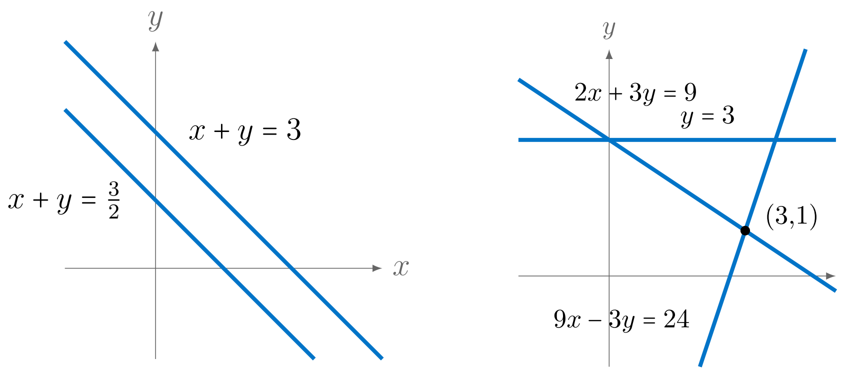
The system found in Example 1.3 is a little more difficult to solve using elimination.

The linear systems found in Example 1.3 and Example 1.4 are called **square systems** because they have the same number of variables as equations. Square systems play an important role in linear algebra and deserve special study, but square systems are not the only linear systems.

Example 1.5

There are tall skinny systems (more equations than variables) like the 3×2 system

$$\begin{array}{rcrcr} x & + & y & = & 5 \\ 2x & + & y & = & 8 \\ x & - & 2y & = & -1 \end{array}$$



(a) **(b)**

Figure 1.3 *Linear systems with no solution.*

and short fat systems (more variables than equations) like the 2×4 system

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & - & x_3 & + & x_4 & = & 3 \\ 2x_1 & - & x_2 & & & + & x_4 & = & 4. \end{array}$$

Example 1.6

There are systems with one unique solution as illustrated in Figures 1.2a and 1.2b, systems with no solutions (inconsistent systems) as shown in Figures 1.3a and 1.3b, and systems with infinitely many solutions as shown in Figure 1.4 where one line lies on top of the other. One such example having infinitely many solutions is

$$\begin{array}{rcl} x + y & = & 3 \\ 2x + 2y & = & 6. \end{array}$$

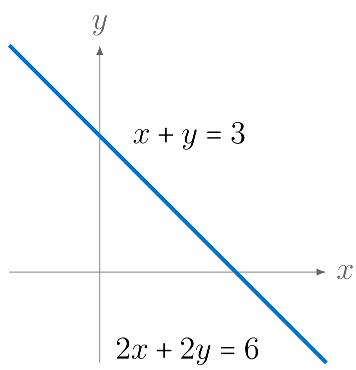


Figure 1.4 *A linear system with infinitely many solutions:* $x + y = 3, 2x + 2y = 6$.

There is one type of system that is easy to spot and always has at least one solution (never inconsistent). A **homogeneous** system is a system in which all equations have a constant term (right-hand side) of 0.

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ & & & & \vdots & & & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

Homogeneous systems play a key role in linear algebra theory and always have the **trivial solution** $x_1 = \cdots = x_n = 0$.

One of our early goals is to develop a systematic procedure for solving systems of linear equations that is carefully designed so as not to drop any solutions nor pick up any extraneous ones. You have already used the elimination process to solve small systems where you multiply equations by constants and add them to other equations to eliminate variables. This is the backbone of the process we are about to learn. Simple elimination gets unwieldy when the number of variables is more than two or three. The new process uses rectangular arrays called **matrices** to keep everything in order. To close the discussion in this section, we give an example showing two types of matrices we use to solve linear systems.

Example 1.7

The 3×3 system

$$\begin{array}{ccccccc} 2x_1 & + & x_2 & - & 2x_3 & = & -5 \\ x_1 & + & x_2 & - & x_3 & = & -2 \\ 3x_1 & + & x_2 & - & 2x_3 & = & -3 \end{array}$$

has associated with it a **coefficient matrix**

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

and an **augmented matrix**

$$\left[\begin{array}{ccc|c} 2 & 1 & -2 & -5 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right].$$

Note that the augmented matrix contains all of the information we really need from the underlying system without the clutter of variable names to slow us down. In the next section, we learn to manipulate the augmented matrix in a way that is equivalent to eliminating variables in the underlying sytem until the solution set is apparent.

Problem Set 1.1

1. Which of the following are linear equations? Put those that are linear in standard form.

(a) $3(x_1 - 2x_2 + 1) = 4x_3 + 7x_2 - 8$ **(b)** $x^2 + y^2 = 1$

(c) $xy + 4z = 3$ **(d)** $\sqrt{3}x_1 - (\sin\pi/5)x_2 = e^2$

(e) $e^{2x}e^{3y} = 5$

2. Determine the dimensions of each of the following systems and state which of the given ordered n-tuples are solutions to the linear system.

(a) $(6, 2)$, $(5, 1)$

$$\begin{array}{rcrcr} 3x & + & 4y & = & 26 \\ 2x & - & 5y & = & 2 \end{array}$$

(b) $(1, -2, 2)$, $(16, -11, 5)$

$$\begin{array}{rcrcrcr} x & + & 2y & + & z & = & -1 \\ 2x & + & 5y & + & 5z & = & 2 \\ 3x & + & 7y & + & 7z & = & 6 \end{array}$$

(c) $(0, 0, 0, 0)$, $(11, 4, 1, 1)$, $(3, 2, 1, -1)$, $(4, 1, 0, 2)$

$$\begin{array}{rrrrcr} x_1 & -2x_2 & -x_3 & -2x_4 & = & 0 \\ 2x_1 & -3x_2 & -5x_3 & -5x_4 & = & 0 \end{array}$$

(d) $(4, 1, 0)$, $(-2, -1, 2)$, $(1, 1, 1)$

$$\begin{array}{rcrcrcr} x & - & y & + & 2z & = & 3 \\ 3x & - & 2y & + & 7z & = & 10 \\ x & & & + & 4z & = & 6 \\ x & + & 2y & + & 6z & = & 8 \end{array}$$

3. Solve the linear systems using the method of elimination of variables. Though it is possible to solve all of these systems using elimination of variables, the point of this exercise is to introduce you to some of the difficulties and frustrations involved in using this method to solve linear systems. Our hope is that solving these systems by simply eliminating variables will show you why it is important to develop matrix techniques to streamline this process. Those matrix techniques are discussed over the next three sections.

(a)
$$\begin{aligned} 3x + 4y &= 10 \\ 2x + 5y &= 12 \end{aligned}$$

(b)
$$\begin{aligned} x + 2y - 5z &= 11 \\ 2x + 3y - 8z &= 18 \\ x - 2y + 3z &= -5 \end{aligned}$$

(c)
$$\begin{aligned} x + 3y + z &= 1 \\ 2x + 5y \quad\; &= 1 \\ 3x + 7y - z &= 0 \end{aligned}$$

(d)
$$\begin{aligned} x_1 + x_2 + 2x_3 - 3x_4 &= 1 \\ 2x_1 + x_2 + 5x_3 - 7x_4 &= 0 \\ x_1 + 3x_2 + x_3 - 3x_4 &= 4 \\ x_1 + 4x_2 \qquad\;\; - x_4 &= 11 \end{aligned}$$

4. Write the coefficient matrix and the augmented matrix for each of the following linear systems. Do not solve the systems.

(a)
$$\begin{aligned} 3(2x - 5y + 1) &= 2(x + 4z) \\ z + 2y - 3 &= 2x + y + z \\ 3z + 7y - x &= 0 \end{aligned}$$

(b)
$$\begin{aligned} 3x_1 + 4x_3 &= 5x_4 - 7x_2 \\ 2x_1 + x_2 &= 6x_3 - 8x_4 + 3x_5 \end{aligned}$$

5. Write the underlying linear system of equations in standard form for each of the following augmented matrices. Do not solve the systems.

(a) $\left[\begin{array}{cc|c} 4 & -15 & -6 \\ -2 & 1 & 0 \\ -1 & 7 & 3 \end{array}\right]$ **(b)** $\left[\begin{array}{cccc|c} 1 & 0 & 2 & 0 & 3 \\ 0 & 4 & 0 & 5 & 0 \\ 6 & 0 & 7 & 0 & 8 \end{array}\right]$

6. The general form for the equation of a parabola in the xy-plane is $y = ax^2 + bx + c$, and for a circle it is $x^2 + y^2 + ax + by = c$, where x and y are variables and a, b, and c are known constants (some restrictions apply). Notice that neither equation is linear in the variables x and y, but both are linear if x and y are considered as known constants and a, b, and c are considered to be the variables. Use systems of linear equations to answer each of the following questions.

(a) Find the equation of the parabola in the xy-plane that passes through the points $(-2,2)$, $(2,4)$, and $(5,3)$.

(b) Find the equation of the circle in the xy-plane that passes through the points $(-2,2)$, $(2,4)$, and $(5,3)$.

(c) Repeat parts (a) and (b) with the points $(1,5)$, $(3,4)$, and $(7,2)$. What is happening geometrically that explains why these systems behave so differently?

1.2 Elementary Row Operations

The primary method that we discuss for solving systems of linear equations is called **Gauss-Jordan elimination**. Theoretically, it amounts to the method of elimination you have used for years to solve simple 2×2 systems. One problem with the method of elimination is that when applied to larger systems is that it gets very messy and unwieldy when solving with pencil and paper. Gauss-Jordan streamlines the elimination process by retaining all the important information but nothing extra. It is a routine mechanical process. That is both a bad and a good thing. It is bad because it is very tedious and boring to perform by hand. It is a good thing because computing devices can and have been programmed to do the tedious calculations. Later we demonstrate how to use *Maple*, a computer algebra system, to carry out this process. You should also consider learning how to perform Gauss-Jordan elimination on your graphing calculator. Until you master the concept, you should perform the process by hand to make sure you understand it thoroughly. Use a calculator or *Maple* only to check your work. After you become fluent in performing the process, you should turn these calculations over to a computing device.

As we learn how Gauss-Jordan elimination works, we also are led to understand why it works. The process starts by putting a system of linear equations in standard form and constructing its augmented matrix as shown below.

$$\begin{array}{rcrcrcr} 2x & + & y & - & 2z & = & -5 \\ x & + & y & - & z & = & -2 \\ 3x & + & y & - & 2z & = & -3 \end{array} \qquad \left[\begin{array}{ccc|c} 2 & 1 & -2 & -5 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right]$$

Notice that the system can be recovered from the augmented matrix (up to the variable names used), so it contains all the necessary information held in the system.

The next step in Gauss-Jordan elimination is to manipulate the augmented matrix by performing a sequence of what are called **elementary row operations**. These elementary row operations change the augmented matrix, and hence the underlying system, in such a way that the solution set remains unchanged but the underlying system becomes so simple that you can discern the solution by inspection.

$$\text{Augmented matrix: } \left[\begin{array}{ccc|c} 2 & 1 & -2 & -5 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right] \to \text{New Augmented Matrix: } \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

This new augmented matrix represents a new system of equations

$$\begin{array}{cccl} x & & & = 2 \\ & y & & = 1 \\ & & z & = 5 \end{array}$$

It is obvious that $(2,1,5)$ is the only solution to this system.

The following elementary row operations constitute legal ways we can manipulate an augmented matrix so as not to change the solution set of the underlying system.

Elementary Row Operations

1. (Scaling) Multiply a row by a nonzero constant. $(r_i \to cr_i)$

2. (Interchange) Swap positions of two rows. $(r_i \leftrightarrow r_j)$

3. (Replacement) Replace a row by the sum of itself plus a constant multiple of another row. $(r_i \to r_i + cr_j)$

These row operations correspond to changing the underlying system (as in the method of elimination) as follows.

1. Multiply an equation in the system on both sides by a nonzero constant.

2. Interchange two equations in the system.

3. Add one equation to a constant multiple of another.

We now explain why elementary row operations do not change the solution sets of the underlying systems. The use of a little basic set theory will help us do this with less mess. To that end, suppose we have two sets A and B. When we write $A \subseteq B$ we mean that every element of A is also in B (i.e. $A \subseteq B$ means $x \in A$ implies $x \in B$). Think of $A \subseteq B$ as meaning A is a subset of B. Then $A = B$ means $A \subseteq B$ and $B \subseteq A$. That is, each set contains the other.

Solutions of a system are ordered n-tuples $(s_1, \cdots, s_n)$ that satisfy all the equations in the system. The solution set of a system is just the collection of all the solutions of that system. To show that elementary row operations do not change the solution set, we let A represent the solution set of the underlying system before the row operations and we let B represent the solution set of the underlying system after the row operations. We show $A = B$ by showing $A \subseteq B$ and $B \subseteq A$.

Consider the first elementary row operation (scaling).

$$\left[\begin{array}{c|c} * & * \\ a_1 \cdots a_n & b \\ * & * \end{array}\right] \rightarrow \left[\begin{array}{c|c} * & * \\ ca_1 \cdots ca_n & cb \\ * & * \end{array}\right]$$

This corresponds to changing the equation $a_1x_1 + \cdots + a_nx_x = b$ to $ca_1x_1 + \cdots + ca_nx_n = cb$ (via multiplication by $c \neq 0$). To show that $A \subseteq B$, we let $(s_1, \cdots, s_n) \in A$ and show that $(s_1, \cdots, s_n) \in B$. Remember that $(s_1, \cdots, s_n) \in A$ means that $(s_1, \cdots, s_n)$ satisfies the original system before scaling. Since only the one row changes, the solutions of all other equations in the system remain unchanged, so $(s_1, \cdots, s_n)$ satifies all other equations in the new system. We focus on the equation that changed. But,

$$\begin{aligned} ca_1s_1 + \cdots + ca_ns_n &= c(a_1s_1 + \cdots + a_ns_n) \\ &= cb \end{aligned}$$

since $(s_1, \cdots, s_n)$ satisfies $a_1x_1 + \cdots + a_nx_n = b$. So $(s_1, \cdots, s_n) \in B$. It follows that $A \subseteq B$. We show $B \subseteq A$ later.

Consider now the second elementary row operation (interchange). Swapping two rows only changes the order in which the equations are listed. It doesn't change the equations at all. Because of that, it doesn't change the solution sets at all. It follows that if the interchange row operation is applied, we have $A = B$.

Lastly, consider the third elementary row operation (replacement).

$$\left[\begin{array}{c|c} * & * \\ a_1 \cdots a_n & b \\ * & * \\ c_1 \cdots c_n & d \\ * & * \end{array}\right] \rightarrow \left[\begin{array}{c|c} * & * \\ a_1 \cdots a_n & b \\ * & * \\ (c_1 + ea_1) \cdots (c_n + ea_n) & (d + eb) \\ * & * \end{array}\right]$$

We suppose $(s_1, \cdots, s_n) \in A$ and proceed to show $(s_1, \cdots, s_n) \in B$ (i.e. we show $A \subseteq B$). Again, our only real concern is the underlying equation of the row that changes to $(c_1 + ea_1)x_1 + \cdots + (c_n + ea_n)x_n = d + eb$. But,

$$\begin{aligned} (c_1 + ea_1)s_1 + \cdots + (c_n + ea_n)s_n &= (c_1s_1 + ea_1s_1) + \cdots + (c_ns_n + ea_ns_n) \\ &= (c_1s_1 + \cdots + c_ns_n) + e(a_1s_1 + \cdots + a_ns_n) \\ &= d + eb \end{aligned}$$

Therefore $(s_1, \cdots, s_n) \in B$ and so $A \subseteq B$. Again, we show $B \subseteq A$ later.

So far we've shown that no solutions are lost through elementary row operations ($A \subseteq B$), but we have not yet shown that extraneous solutions are not picked up. This part is easy when we realize that all three elementary row operations are reversible.

Reversing Row Operations

1. Multiplying a row by a number $c \neq 0$ is reversed by multiplying the row again by $\frac{1}{c}$.

2. Swapping two rows is reversed by swapping the same two rows back.

3. Replacment $r_i \to r_i + cr_j$ is reversed by $r_i \to r_i - cr_j$.

Let's look at an example of reversing replacement.

Example 1.8

$$\left[\begin{array}{ccc|c} 2 & 1 & -2 & -5 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right] \xrightarrow{r_1 \to r_1 - 2r_2} \left[\begin{array}{ccc|c} 0 & -1 & 0 & -1 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right] \xrightarrow{r_1 \to r_1 + 2r_2} \left[\begin{array}{ccc|c} 2 & 1 & -2 & -5 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right]$$

Let

$$\begin{aligned} A &= \text{solution set of the original underlying system,} \\ B &= \text{solution set after one elementary row operation, and} \\ C &= \text{solution set after a second elementary row operation to reverse the first.} \end{aligned}$$

Because solutions are not lost through elementary row operations, $A \subseteq B$ and $B \subseteq C$. But since the second elementary row operation reversed the first, the first and third augmented matrices are identical making the underlying systems identical and the solutions sets A and C are equal (i.e. $A = C$). So $A \subseteq B$ and $B \subseteq A$ giving $A = B$. This gives us our first theorem.

Theorem 1.1. Elementary row operations do not change the solution sets to their underlying systems.

Before starting the exercises, we review again the notation we've introduced to describe elementary row operations.

$r_i \to cr_i$	means: Multiply row i by the nonzero constant c.
$r_i \leftrightarrow r_j$	means: Swap rows i and j.
$r_i \to r_i + cr_j$	means: Replace row i with itself plus the constant c times row j.

To perform a sequence of elementary row operation on a matrix, perform the first operation on the given matrix, the second operation on the matrix resulting from the first operation, the third operation on the result of the second operation, etc.

Problem Set 1.2

1. Perform the following sequence of elementary row operations on the given matrix: $r_1 \leftrightarrow r_2$, $r_2 \to r_2 - 2r_1$, $r_3 \to r_3 - 3r_1$, $r_2 \to -\frac{1}{5}r_2$, $r_3 \to r_3 + 10r_2$, $r_1 \to r_1 - 3r_2$.

$$\begin{bmatrix} 2 & 1 & 4 \\ 1 & 3 & 2 \\ 3 & -1 & 6 \end{bmatrix}$$

2. **(a)** Form the augmented matrix of the following linear system.

$$\begin{array}{rcrcrcr} 3x & - & y & + & z & = & 6 \\ 2x & + & y & + & 5z & = & 8 \\ x & - & y & - & 2z & = & 1 \end{array}$$

(b) Perform the following sequence of elementary row operations on the augmented matrix from (a): $r_1 \leftrightarrow r_3$, $r_2 \to r_2 - 2r_1$, $r_3 \to r_3 - 3r_1$, $r_2 \to \frac{1}{3}r_2$, $r_3 \to r_3 - 2r_2$, $r_1 \to r_1 + 2r_3$, $r_2 \to r_2 - 3r_3$, $r_1 \to r_1 + r_2$.

(c) Determine the underlying linear system associated with the final augmented matrix resulting from (b).

(d) What is the solution to the system in (c)?

(e) Check to see that the solution in (d) is also a solution to the linear system in (a).

3. The first elementary row operation requires that the constant c is nonzero, but in the third elementary row operation the constant c is not restricted. Explain the effect $c = 0$ would have on the the underlying system and its solution set under the first and third elementary row operations.

1.3 Row Reduction and Reduced Row-Echelon Form

There are several reasons for performing elementary row operations on a matrix. First among these is that of solving systems of linear equations using Gauss-Jordan elimination. To begin, we recall an example from the previous section.

Example 1.9

Consider a linear system that has already been put in standard form.

$$\begin{array}{rcrcrcr} 2x & + & y & - & 2z & = & -5 \\ x & + & y & - & z & = & -2 \\ 3x & + & y & - & 2z & = & -3 \end{array}$$

From here, we form an associated augmented matrix and then perform a sequence of elementary row operations on that matrix.

$$\left[\begin{array}{ccc|c} 2 & 1 & -2 & -5 \\ 1 & 1 & -1 & -2 \\ 3 & 1 & -2 & -3 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

From the new (improved!) augmented matrix, we recover a new but equivalent system of equations.

$$\begin{array}{ccccc} x & & & = & 2 \\ & y & & = & 1 \\ & & z & = & 5 \end{array}$$

From this point, the unique solution $(2,1,5)$ is obvious.

Not all systems reduce to systems that are this simple, so we want to develop a more general form that allows us to recognize the solution set of a system. Since this form is useful for other purposes as well, it is presented without reference to an underlying system.

Definition 1.3 (Reduced Row-Echelon Form (RREF)). A matrix is said to be in reduced row-echelon form (RREF) if it has the following four properties.

- All zero rows (rows of all zeroes) are below the nonzero rows.
- The first nonzero entry of a nonzero row is a one (leading one) and lies to the right of all leading ones above it.
- The entries below a leading one are all zero.
- The entries above a leading one are all zero.

Example 1.10

The following matrices are in reduced row-echelon form.

$$\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -7 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 4 & 8 \\ 0 & 1 & 6 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 4 & 0 & 9 \\ 0 & 0 & 1 & 2 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 2 & 4 & 6 \\ 0 & 1 & 3 & 5 & 7 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

These matrices are not in reduced row-echelon form.

$$\begin{bmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & 1 & 4 \end{bmatrix}, \begin{bmatrix} 2 & 0 & 0 & 2 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 4 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Which of these matrices is in reduced row-echelon form?

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 2 & 5 & 1 \\ 0 & 1 & 3 & 7 & 6 \end{bmatrix}$$

Recall that our goal is to perform elementary row operations on the augmented matrix formed from a system of linear equations in order to get the augmented matrix into reduced row-echelon form so that the solution of the system can be readily determined. Any sequence of elementary row operations that achieves this goal is fine, but the systematic process for achieving that goal is called Gauss-Jordan elimination. It is designed to transform a matrix into reduced row-echelon form with a minimum number of arithmetic operations. Matrices found at each stage of the Gauss-Jordan elimination process are related to each other through these elementary row operations. We give this relationship a name.

Definition 1.4. If one matrix A can be transformed into another B through a sequence of elementary row operations, we say that A and B are **row equivalent**.

From the discussion in section 1.2 it is clear that row-equivalent augmented matrices represent underlying systems of linear equations with identical solution sets. This fact is the key to the proof of the following theorem. The proof is in the appendix.

Theorem 1.2. Each matrix is row equivalent to exactly one matrix in reduced row-echelon form.

Though one generally can't initially tell where the leading ones will end up, the above theorem guarantees that for any given matrix, the positions of the leading ones are determined right from the start. That is, if Abe performed one sequence of elementary row operations on matrix A to get it into reduced row-echelon form and then Betty used another sequence to get it into reduced row-echelon form, Abe and Betty must end up with the same (reduced) matrix. The leading ones will be in exactly the same places for Abe and Betty and the other corresponding entries will match exactly as well.

Definition 1.5. Let A be a matrix and R its reduced row-echelon form. The positions in which the leading one of R appear are called the **pivot positions** of A. A column of A that contains a pivot position is called a **pivot column** of A.

Example 1.11

The pivot positions of the matrix

$$\begin{bmatrix} 1 & 4 & 0 & 9 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

are $(1,1)$ and $(2,3)$. The pivot columns are column 1 and column 3. The pivot positions and pivot columns are the same for all matrices that are row equivalent to this matrix.

We now illustrate Gauss-Jordan elimination with the following annotated example.

Example 1.12

Unless the first column is all 0, the first pivot position is (1,1) so we need to get any 0 out of that position.

$$\begin{bmatrix} 0 & 1 & 1 & 4 \\ -2 & 5 & -3 & -8 \\ 1 & -1 & 2 & 7 \end{bmatrix}$$

We could swap for the 1 or -2 in the first column. We want a leading 1 in (1,1) so we swap rows 1 and 3 ($r_1 \leftrightarrow r_3$).

$$\xrightarrow{r_1 \leftrightarrow r_3} \begin{bmatrix} 1 & -1 & 2 & 7 \\ -2 & 5 & -3 & -8 \\ 0 & 1 & 1 & 4 \end{bmatrix}$$

Now we need all 0's below the leading 1 in (1,1) so we perform row operation $r_2 \to r_2 + 2r_1$.

$$\xrightarrow{r_2 \to r_2 + 2r_1} \begin{bmatrix} 1 & -1 & 2 & 7 \\ 0 & 3 & 1 & 6 \\ 0 & 1 & 1 & 4 \end{bmatrix}$$

Next, ignore row 1 and column 1. Unless entries (2,2) and (3,2) in column 2 are both 0, (2,2) is a pivot position, so we need a leading 1 there. We can accomplish it in this example by swapping rows 2 and 3 ($r_2 \leftrightarrow r_3$).

$$\xrightarrow{r_2 \leftrightarrow r_3} \begin{bmatrix} 1 & -1 & 2 & 7 \\ 0 & 1 & 1 & 4 \\ 0 & 3 & 1 & 6 \end{bmatrix}$$

We now work to get a 0 below the second leading 1 (pivot position). We can do that by performing row operation $r_3 \to r_3 - 3r_2$.

$$\xrightarrow{r_3 \to r_3 - 3r_2} \begin{bmatrix} 1 & -1 & 2 & 7 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & -2 & -6 \end{bmatrix}$$

The (3,3) position is the last pivot position. This location must hold a leading 1. To do that, we scale. That is, we perform row operation $r_3 \to -\frac{1}{2}r_3$. Note: $r_3 \to r_3 + 3r_2$ or $r_2 \leftrightarrow r_3$ would get a 1 in the right position but would destroy some carefully placed 0's from our earlier work that we must preserve.

$$\xrightarrow{r_3 \to -\frac{1}{2}r_3} \begin{bmatrix} 1 & -1 & 2 & 7 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

So far working from left to right we have put the leading 1's in place and have 0's below them. Next we work from right to left to place 0's above the leading 1's. To start, we perform row operations $r_1 \to r_1 - 2r_3$ and $r_2 \to r_2 - r_3$.

$$\xrightarrow{r_1 \to r_1 - 2r_3,\ r_2 \to r_2 - r_3} \begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Finally, we place a 0 above the leading 1 in the pivot position $(2,2)$. To do so, perform row operation $r_1 \to r_1 + r_2$.

$$\xrightarrow{r_1 \to r_1 + r_2} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

This matrix is in reduced row-echelon form and thus completes the Gauss-Jordan elimination process.

We now do another example to see what else might happen in practice.

Example 1.13

Our goal in this example is perform Gauss-Jordan elimination on

$$\begin{bmatrix} 3 & 9 & -1 & 9 \\ 2 & 6 & 0 & 5 \\ 1 & 3 & -1 & 4 \end{bmatrix}.$$

To obtain a leading 1 in $(1,1)$, we swap rows 1 and 3 ($r_1 \leftrightarrow r_3$). Row operations $r_1 \to r_1 - r_2$ or $r_1 \to r_1 - 2r_3$ would also work, but the row swap seems easier.

$$\xrightarrow{r_1 \leftrightarrow r_3} \begin{bmatrix} 1 & 3 & -1 & 4 \\ 2 & 6 & 0 & 5 \\ 3 & 9 & -1 & 9 \end{bmatrix}$$

To get 0's below the leading 1 in position (1,1), we perform row operations $r_2 \to r_2 - 2r_1$ and $r_3 \to r_3 - 3r_1$.

$$\xrightarrow{r_2 \to r_2 - 2r_1,\ r_3 \to r_3 - 3r_1} \begin{bmatrix} 1 & 3 & -1 & 4 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 2 & -3 \end{bmatrix}$$

This time since both the $(2,2)$ and $(3,2)$ entries are 0, there is no pivot position in the second column. We move our attention to the third column where $(2,3)$ is a pivot position. We could do $r_2 \to \frac{1}{2}r_2$ to get a leading 1 in position $(2,3)$, but that introduces fractions unnescessarily early. Instead, $r_3 \to r_3 - r_2$ places a 0 below the $(2,3)$ pivot position.

$$\xrightarrow{r_3 \to r_3 - r_2} \begin{bmatrix} 1 & 3 & -1 & 4 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Now scale the second row.

$$\xrightarrow{r_2 \to \frac{1}{2}r_2} \begin{bmatrix} 1 & 3 & -1 & 4 \\ 0 & 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

To get a 0 above the second leading 1 (pivot position), perform $r_1 \to r_1 + r_2$.

$$\xrightarrow{r_1 \to r_1 + r_2} \begin{bmatrix} 1 & 3 & 0 & \frac{5}{2} \\ 0 & 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This matrix is in reduced row-echelon form.

Problem Set 1.3

1. Find the reduced row echelon form of each of the following matrices.

(a) $\begin{bmatrix} 3 & -5 & -9 \\ 1 & -2 & -4 \end{bmatrix}$

(b) $\begin{bmatrix} 3 & -9 & -7 \\ 2 & -6 & -3 \end{bmatrix}$

(c) $\begin{bmatrix} 3 & 9 & 7 & 11 \\ 1 & 3 & 2 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 5 & 14 \\ 4 & 3 & 15 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & -1 & -2 & -3 \\ 3 & -2 & -3 & -7 \\ 2 & 0 & 3 & -3 \end{bmatrix}$

(f) $\begin{bmatrix} 4 & 8 & -11 \\ 1 & 2 & -3 \\ -2 & -4 & 2 \end{bmatrix}$

(g) $\begin{bmatrix} 1 & 2 & 5 & 0 \\ -1 & 0 & 1 & 2 \\ 2 & 1 & 1 & -3 \end{bmatrix}$

(h) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & -3 & 2 & 1 \\ 1 & 5 & 2 & 3 \end{bmatrix}$

(i) $\begin{bmatrix} 1 & 2 & -2 \\ -2 & -3 & 1 \\ 2 & 6 & -10 \\ -1 & 1 & -7 \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 1 & 0 & 2 \\ 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & 6 \\ -1 & 1 & 2 & 2 \end{bmatrix}$

2. Each 2×2 matrix is row equivalent to a matrix in reduced row echelon form that falls into one of the following four categories by the number and location of the leading 1's:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & * \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The astrix ($*$) indicates an arbitrary number.

(a) List all such categories of 3×3 matrices in reduced row-echelon form.

(b) Do the same for 4×4 matrices.

(c) Based on these three examples, there seems to be a clear pattern as to how many such categories there are for $n \times n$ matrices in reduced row-echelon form. What is that pattern?

(d) Explain why this pattern holds in general.

1.4 Solutions of Systems of Linear Equations

Since the underlying systems of row-equivalent augmented matrices have the same solution sets, we can find the solution to our original system by reading off the solution to the underlying system of the reduced row-echelon form of its augmented matrix.

Example 1.14

The system

$$\begin{array}{rcrcrcr} & & y & + & z & = & 4 \\ -2x & + & 5y & - & 3z & = & -8 \\ x & - & y & + & 2z & = & 7 \end{array}$$

corresponds to an augmented matrix that can be reduced through a sequence of elementary row operations.

$$\left[\begin{array}{rrr|r} 0 & 1 & 1 & 4 \\ -2 & 5 & -3 & -8 \\ 1 & -1 & 2 & 7 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{rrr|r} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

From the reduced row-echelon form of the augmented matrix, we recover the equivalent system of equations

$$\begin{array}{ccccc} x & & & = & 2 \\ & y & & = & 1 \\ & & z & = & 3 \end{array}$$

having (obvious) unique solution $(2, 1, 3)$.

Example 1.15

The system

$$\begin{array}{rcrcrcr} x & - & y & - & z & = & -1 \\ 2x & - & y & + & z & = & 3 \\ 4x & - & 3y & - & z & = & 1 \end{array}$$

corresponds to an augmented matrix that can be reduced through a sequence of elementary row operations.

$$\left[\begin{array}{rrr|r} 1 & -1 & -1 & -1 \\ 2 & -1 & 1 & 3 \\ 4 & -3 & -1 & 1 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{rrr|r} 1 & 0 & 2 & 4 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

From the reduced row-echelon form of the augmented matrix, we recover the equivalent system of equations

$$\begin{array}{rcrcrcr} x & & & + & 2z & = & 4 \\ & & y & + & 3z & = & 5 \\ & & & & 0 & = & 0. \end{array}$$

Notice the last equation $0x + 0y + 0z = 0$ is true for all ordered triples. It places no restrictions on the solution set at all, so we can simply ignore it. Any ordered triple that satisfies the first two equations satisfies the whole system.

Note too that in this example the x and y columns are pivot columns (the reduced row-echelon form contains a leading 1) but the z column is not. That allows us to easily solve for x and y in terms of z. Now we see that z could be any number at all and x and y can be easily adjusted to produce a solution.

$$\begin{array}{rcrcr} x & = & 4 & - & 2z \\ y & = & 5 & - & 3z \end{array}$$

That is, if $z = 0$, then $x = 4$ and $y = 5$ so that $(4, 5, 0)$ is a solution. If $z = 1$, then $x = 3$ and $y = 2$ and $(3, 2, 1)$ is a solution. Etc.

Since z is seen as a free choice we illustrate that fact by setting it equal to a parameter t (i.e. we let $z = t$). What results is called a general solution to this system. A one-parameter family of solutions is given by

$$\begin{array}{rcrcr} x & = & 4 & - & 2t \\ y & = & 5 & - & 3t \\ z & = & & & t. \end{array}$$

Example 1.16

The system

$$\begin{array}{rcrcrcr} 3x & + & 9y & - & z & = & 9 \\ 2x & + & 6y & & & = & 5 \\ x & + & 3y & - & z & = & 4 \end{array}$$

has augmented matrix

$$\begin{bmatrix} 3 & 9 & -1 & 9 \\ 2 & 6 & 0 & 5 \\ 1 & 3 & -1 & 4 \end{bmatrix}.$$

This can be row reduced to

$$\begin{bmatrix} 1 & 3 & 0 & \frac{5}{2} \\ 0 & 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which represents the system

$$\begin{array}{rcrcrcr} x & + & 3y & & & = & \frac{5}{2} \\ & & & & z & = & -\frac{3}{2} \\ & & & & 0 & = & 0. \end{array}$$

This time the x and z columns are pivot columns and the y column is not, so y is a free choice. The general solution comes from setting y equal to a free parameter t.

$$\begin{array}{rcrcr} x & = & \frac{5}{2} & - & 3t \\ y & = & & & t \\ z & = & -\frac{3}{2} & & \end{array}$$

Note that z never changes but x and y depend on the parameter t.

Example 1.17

The system

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & -2 \\ x & + & 3y & + & 2z & = & -1 \\ x & + & 4y & + & 5z & = & 4 \end{array}$$

has augmented matrix

$$\begin{bmatrix} 1 & 2 & -1 & -2 \\ 1 & 3 & 2 & -1 \\ 1 & 4 & 5 & 4 \end{bmatrix}.$$

A few elementary row operations yield the matrix

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

which represents the system

$$\begin{array}{ccccccc} x & + & 2y & - & z & = & -2 \\ & & y & + & 3z & = & 1 \\ & & & & 0 & = & 4. \end{array}$$

Though it wouldn't hurt to continue Gauss-Jordan elimination there is no need since the solution set is already clear. The last equation $0x + 0y + 0z = 4$ is not satisfied by any ordered triple. This system has no solution because the third equation has no solution. Because the original system has the same solution set as this final one, the original system has no solution. It is an inconsistent system.

The lesson to be learned from this last example is that we can stop the row reduction process as soon as we realize that the last column - the column that represents the right hand side of the system of linear equations - is a pivot column of the augmented matrix.

Together, the last four examples illustrate the possible outcomes of Gauss-Jordan elimination quite well though not all systems are square and sometimes more than one parameter (free choice) is needed to describe the solution set. One last example illustrates this situation.

Example 1.18

Imagine that you are given a system that has the following matrix as its reduced row-echelon form.

$$\left[\begin{array}{ccccccc|c} 1 & 0 & 2 & 0 & 4 & 0 & 7 & 11 \\ 0 & 1 & 3 & 0 & 5 & 0 & 8 & 12 \\ 0 & 0 & 0 & 1 & 6 & 0 & 9 & 13 \\ 0 & 0 & 0 & 0 & 0 & 1 & 10 & 14 \end{array}\right]$$

Since the last column is not a pivot column, the system is consistent. Since columns 3, 5, and 7 are not pivot columns, we let $x_3 = r$, $x_5 = s$, and $x_7 = t$ be three parameters we use to describe the solution set to the system. To that end, we solve each variable $x_1, \ldots, x_7$ in terms of r, s, and t.

We write the general solution

$$\begin{array}{ccccccccc} x_1 & = & 11 & - & 2r & - & 4s & - & 7t \\ x_2 & = & 12 & - & 3r & - & 5s & - & 8t \\ x_3 & = & & & r & & & & \\ x_4 & = & 13 & & & - & 6s & - & 9t \\ x_5 & = & & & & & s & & \\ x_6 & = & 14 & & & & & - & 10t \\ x_7 & = & & & & & & & t \end{array}$$

To summarize,

- A system is **inconsistent** if and only if the last column of its augmented matrix is a pivot column.
- A system has one **unique solution** if and only if every column but the last of its augmented matrix is a pivot column.
- A system has a k**-parameter family of solutions** if and only if the last column of its augmented matrix is not a pivot column and there are k other columns that are not pivot columns.

The process you have just learned is Gauss-Jordan elimination. Though it is easier to write the solution from the reduced row-echelon form (RREF), there is another generally accepted method for solving systems that requires fewer elementary row operations. It is called **Gaussian elimination with back substitution**. The trade off is fewer row operations but the back substitution requires some work too.

Definition 1.6. A matrix is in **row-echelon form** (REF) if it satisfies the first three of the four properties that define reduced row-echelon form (RREF).

- All zero rows (rows of all zeroes) are below the nonzero rows.
- The first nonzero entry of a nonzero row is a one (leading one) and lies to the right of all leading ones above it.
- The entries below a leading one are all zero.

In row-echelon form, there must be zeroes below the leading ones (pivots) but not necessarily above.

Example 1.19

The system

$$\begin{array}{rcrcrcr} & & y & + & z & = & 4 \\ -2x & + & 5y & - & 3z & = & -8 \\ x & - & y & + & 2z & = & 7 \end{array}$$

has augmented matrix

$$\begin{bmatrix} 0 & 1 & 1 & 4 \\ -2 & 5 & -3 & -8 \\ 1 & -1 & 2 & 7 \end{bmatrix}.$$

This can be put in the row-echelon form (REF)

$$\begin{bmatrix} 1 & -1 & 2 & 7 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

which represents the system

$$\begin{aligned} x - y + 2z &= 7 \\ y + z &= 4 \\ z &= 3. \end{aligned}$$

We now perform the back substitution. First, we see $z = 3$ so we substitute 3 for z in the second equation above and solve for y. With y and z known, we substitute those values into the top equation and solve for x.

$$\begin{aligned} z &= 3 \\ y &= 4 - (3) = 1 \\ x &= 7 + (1) - 2(3) = 2 \end{aligned}$$

giving solution $(2, 1, 3)$. Back substitution can be used with parameters too. Just substitute the parameters in and solve algebraically.

It should be noted that a system has a unique solution if and only if every column but the last of the augmented matrix is a pivot column. For square systems, this means that the reduced row-echelon form of the coefficient matrix has 1's down the main diagonal and 0's elsewhere.

Definition 1.7. Let A be an $m \times n$ matrix. The **rank** of A, denoted *rank* A, equals the number of leading ones in its reduced row-echelon form (RREF). This is equal to the number of nonzero rows in a row-echelon form (REF) of A.

Many properties of matrices can be described in terms of rank so we will use this term often. For now we note that rank gives us a nice way to characterize consistent and inconsistent systems of linear equations. The following theorem is merely a reinterpretation of what we already know about solutions of systems of linear equations.

Theorem 1.3. A system of linear equations is consistent if and only if the rank of its augmented matrix equals the rank of its coefficient matrix.

Definition 1.8. Let A be an $m \times n$ matrix. The **nullity** of A, denoted *nullity* A, is $n - rank\ A$.

As noted in section 1.1, all homogeneous systems are consistent. The nullity of the coefficient matrix equals the number of free parameters in the solution set of a homogeneous system. The next theorem extends this idea to nonhomogeneous linear systems.

Theorem 1.4. If the ranks of the coefficient matrix and the augmented matrix of a linear system are equal, then the nullity of the coefficient matrix equals the number of free parameters in the system's solution.

Problem Set 1.4

Find the general solution of the underlying system of the augmented matrices in 1 - 8 in reduced row-echelon form.

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

2. $\begin{bmatrix} 1 & 0 & 4 & 2 \\ 0 & 1 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 3 & 0 & 5 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 6 & 5 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

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

6. $\begin{bmatrix} 1 & 0 & 2 & 5 & 4 \\ 0 & 1 & 3 & 1 & 6 \end{bmatrix}$

7. $\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

8. $\begin{bmatrix} 1 & 0 & 2 & 0 & 4 & 8 & 0 & 9 \\ 0 & 1 & 3 & 0 & 5 & 1 & 0 & 7 \\ 0 & 0 & 0 & 1 & 6 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 \end{bmatrix}$

Solve the systems in 9 - 13 using Gauss-Jordan elimination.

9. $x - y = 1$
$3x - 2y = 7$

10. $x + 2y + 3z = 7$
$x - y - 6z = -8$
$x + y - z = 0$

11. $x - y + 3z = -2$
$2x - y + 5z = -1$
$3x - y + 7z = 0$

12. $x_1 - x_2 + 3x_3 + 3x_4 = -3$
$3x_1 - 2x_2 + 4x_3 + 3x_4 = -5$

13. $x_1 + x_2 + 2x_3 + 2x_4 = -5$
$2x_1 + 3x_2 + 3x_3 + 3x_4 = -4$
$x_1 - x_2 + 5x_3 + 2x_4 = -8$
$2x_1 + x_2 + 6x_3 + 4x_4 = -11$

Solve the homogeneous systems in 14 - 16 using Gauss-Jordan elimination.

14. $3x + 4y = 0$
$5x + 8y = 0$

15. $2x - y - 6z = 0$
$3x + 7y - 9z = 0$
$x + 2y - 3z = 0$

16. $3x_1 - 4x_2 - 5x_3 - 6x_4 = 0$

Solve the systems in 17 and 18 using Gaussian elimination with back substitution.

17. $x + 3y + z = 2$
$2x + 5y - 2z = -1$
$x + y - z = -2$

18. $x + 3y + 2z = 4$
$2x + 7y + 5z = 11$
$x + 2y + z = 1$

19. A cubic curve in the x, y plane is the graph of an equation of the form $y = ax^3 + bx^2 + cx + d$, where a, b, c, and d are constants and $a \neq 0$. Find the equation of the cubic curve that passes through the points $(-1, -10), (0, -4), (1, -2)$, and $(2, 2)$.

Solve the systems in 20 and 21 for x and y in terms of a, and b.

20. $x + y = a$
$x + 2y = b$

21. $x + 2y = a$
$2x + 4y = b$

1.5 Matrix Operations

Matrices are used throughout mathematics and its applications in many ways other than representing systems of linear equations. Matrices can be added, subtracted, and multiplied (this is called **matrix arithmetic**). These matrix operations have many properties (like associativity and distributivity) that are similar to, though sometimes more complicated than, their counterparts in the real number system. We study matrix operations and their properties in the next two sections.

We have already seen several matrices. They consist of numbers lined up in rows and columns (an array). In some settings, we want to put other objects, like polynomials, in matrices rather than numbers. The objects that populate matrices in linear algebra are called **scalars**. From here on, assume that a scalar is a real number unless suggested otherwise. Later, we will briefly discuss scalars consisting of complex numbers.

Definition 1.9. A matrix is a rectangular array of scalars. A matrix with m rows and n columns is called an $m \times n$ matrix (m and n are called the **dimensions** of the matrix). A matrix with just one row is called a **row vector** and a matrix with just one column is called a **column vector**. Matrices with the same number of rows and columns are called **square matrices**.

Upper case letters like A, B, M, and P are usually used to indicate a matrix except when they are row or column vectors. In that case, bold-faced lowercase letters (e.g. $\mathbf{v}$ or $\mathbf{w}$) are used in print. When handwritten, arrows are usually placed over the top to indicate a vector (e.g. $\vec{v}$ or $\vec{w}$).

Example 1.20

In this example, A is a 2×3 matrix, $\mathbf{u}$ is a column vector called a 2-vector because it has two entries, $\mathbf{v}$ is a row vector (a 4-vector), and B is a 2×2 (and hence square) matrix.

$$A=\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix},\quad \mathbf{u}=\begin{bmatrix}3\\-5\end{bmatrix},\quad \mathbf{v}=\begin{bmatrix}1&0&5&-2\end{bmatrix},\quad B=\begin{bmatrix}1&2\\5&-3\end{bmatrix}$$

Double subscripts are used to indicate the position (row and column) of a scalar within a matrix. The entry in row i and column j of the matrix A is indicated by $a_{i,j}$ or $(A)_{i,j}$. Only one subscript is used to indicate the position of a scalar in a row or column vector because the other subscript is known to be 1. In the example above, $a_{2,1}=4$, $u_1=3$, $v_2=0$, and $(B)_{1,2}=2$.

Matrix Arithmetic

Definitions 1.10 and 1.11 indicate when and how we may compare or combine (add) matrices.

Definition 1.10 (Equality of Matrices)**.** Two matrices A and B are **equal** if they have the same dimensions and $a_{i,j}=b_{i,j}$ for all i and j. That is, $A=B$ if they have the same shape and corresponding entries are equal.

If two matrices have the same dimensions, they can be added and subtracted.

Example 1.21

Addition is performed by simply adding corresponding entries together.

$$\begin{bmatrix}2&1\\4&-3\end{bmatrix}+\begin{bmatrix}3&-4\\0&1\end{bmatrix}=\begin{bmatrix}5&-3\\4&-2\end{bmatrix}$$

Subtraction is similar.

$$\begin{bmatrix}2&1\\4&-3\end{bmatrix}-\begin{bmatrix}3&-4\\0&1\end{bmatrix}=\begin{bmatrix}-1&5\\4&-4\end{bmatrix}$$

Definition 1.11 (Matrix Addition)**.** If A and B are both $m\times n$ matrices, then $A+B$ is $m\times n$ and $(A+B)_{i,j}=a_{i,j}+b_{i,j}$.

Example 1.22

A matrix A can also be multiplied by a scalar. To do this, each entry of A gets multiplied by the scalar.

$$3\begin{bmatrix} 2 & 1 \\ 4 & -3 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ 12 & -9 \end{bmatrix}$$

Definition 1.12 (Scalar Multiplication)**.** If A is an $m \times n$ matrix and c is a scalar, then cA is $m \times n$ and $(cA)_{i,j} = ca_{i,j}$.

If we have a collection of matrices of the same dimensions, we can form a **linear combination** of them by multiplying each one by some scalar and adding them together.

Example 1.23

Since

$$\begin{aligned} 2\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} + (-1)\begin{bmatrix} 3 & 1 \\ -1 & 0 \end{bmatrix} + 3\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} 2 & 0 \\ 4 & 2 \end{bmatrix} + \begin{bmatrix} -3 & -1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 2 \\ 8 & 2 \end{bmatrix} \end{aligned}$$

we say that

$$\begin{bmatrix} -1 & 2 \\ 8 & 2 \end{bmatrix}$$

is a linear combination of

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ -1 & 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Of course, these three matrices have many other linear combinations formed by choosing different scalar multipliers.

We will frequently have opportunities to find linear combinations of sets of column vectors.

Vector Equations

Example 1.24

To determine whether $\begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, we look for scalars x and y such that $x\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + y\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}$. To solve this vector equation, the top component on the left side of the equation $2x+1y$ must equal 5. Likewise, the middle and bottom components must equal 1 and 8 respectively. That is, there must be scalars x and y that satisfy this system of equations.

$$\begin{array}{rcrcr} 2x & + & y & = & 5 \\ x & + & 2y & = & 1 \\ 3x & + & y & = & 8 \end{array}$$

Row reduction leads to the following sequence of augmented matrices.

$$\left[\begin{array}{rr|r} 2 & 1 & 5 \\ 1 & 2 & 1 \\ 3 & 1 & 8 \end{array}\right] \to \left[\begin{array}{rr|r} 1 & 2 & 1 \\ 0 & -3 & 3 \\ 0 & -5 & 5 \end{array}\right] \to \left[\begin{array}{rr|r} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{rr|r} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right]$$

The solution to the underlying system is thus $x = 3, y = -1$. So the answer is yes, $\begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. The vector equation below demonstrates this fact.

$$3\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + (-1)\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}$$

We see in this last example that solving a vector equation for unknown scalars x and y is equivalent to solving a system of linear equations. We will do this type of thing so often that we will skip over the middle step of writing down the system of linear equations as this next example illustrates.

Example 1.25

Is $\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$?

$$\left[\begin{array}{rr|r} 2 & 1 & 3 \\ 1 & 2 & 1 \\ 3 & 1 & 2 \end{array}\right] \to \left[\begin{array}{rr|r} 1 & 2 & 1 \\ 0 & -3 & 1 \\ 0 & -5 & 1 \end{array}\right] \to \left[\begin{array}{rr|r} 1 & 2 & 1 \\ 0 & -3 & 1 \\ 0 & 15 & -3 \end{array}\right] \to \left[\begin{array}{rr|r} 1 & 2 & 1 \\ 0 & -3 & 1 \\ 0 & 0 & 2 \end{array}\right]$$

Since the third column is a pivot column, we see that the underlying system is inconsistent. This means that $\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ is not a linear combination of $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

Matrix multiplication is the most complicated of the matrix operations. It distinguishes itself from addition and subtraction in that in order to multiply two matrices, they need not have the same dimensions. Strangely, in order to multiply two matrices, we require that the number of columns of the first matrix (on the left) equals the number of rows of the second matrix (on the right). We begin by considering the special case in which the second matrix is a column vector.

Definition 1.13. Suppose A is an $m \times n$ matrix and $\mathbf{v}$ is an $n \times 1$ column vector. The **product** $A\mathbf{v}$ is the $m \times 1$ column vector defined by

$$(A\mathbf{v})_i = \sum_{j=1}^{n} a_{ij} v_j.$$

Example 1.26

Let

$$A = \begin{bmatrix} 4 & -2 & 0 \\ 0 & -4 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}.$$

Since A has three columns and $\mathbf{v}$ has three rows, they can be multiplied. The result is a 2×1 column vector. Following the definition, we get

$$\begin{bmatrix} 4 & -2 & 0 \\ 0 & -4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \cdot 1 + (-2) \cdot (-3) + 0 \cdot 5 \\ 0 \cdot 1 + (-4) \cdot (-3) + 2 \cdot 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 22 \end{bmatrix}.$$

To compute the first entry $(A\mathbf{v})_1$ in the product we take the first row of the matrix on the left and line it up with the column vector on the right. We then multiply componentwise and add. The second entry is calculated in the same way except that the second row of A was used instead of the first.

This process of multiplying componentwise and adding is called a **dot product** and will be studied in more detail later. We find the first entry of the product $A\mathbf{v}$ by finding the dot product by 'dotting' the first row of A with $\mathbf{v}$. The second entry of the product is the second row of A 'dotted' with $\mathbf{v}$.

Example 1.27

There is another important way of viewing matrix multiplication that involves linear combinations. It is illustrated well in Example 1.26 except we take a detour through linear combinations of column vectors before we finish.

$$\begin{aligned} \begin{bmatrix} 4 & -2 & 0 \\ 0 & -4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix} &= \begin{bmatrix} 4 \cdot 1 + (-2) \cdot (-3) + 0 \cdot 5 \\ 0 \cdot 1 + (-4) \cdot (-3) + 2 \cdot 5 \end{bmatrix} \\ &= 1 \begin{bmatrix} 4 \\ 0 \end{bmatrix} + (-3) \begin{bmatrix} -2 \\ -4 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 10 \\ 22 \end{bmatrix}. \end{aligned}$$

Notice that the three column vectors in the linear combination come from the columns of A, and the scalar multipliers are the entries of $\mathbf{v}$. This is not an accident. As long as the matrix A can be multiplied by the vector $\mathbf{v}$, the result is a linear combination of the columns of A with the scalar multipliers coming from the entries of $\mathbf{v}$.

Matrix Equations

Is there a column vector $\begin{bmatrix} x \\ y \end{bmatrix}$ that satisfies the **matrix equation**

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}?$$

We can view this question in three different ways. The first is the matrix equation itself. Solutions are 2×1 column vectors that if substituted in for $\begin{bmatrix} x \\ y \end{bmatrix}$ would satisfy the matrix equation. But, if we carry out the matrix multiplication, we see that we seek x and y such that

$$\begin{bmatrix} 2x + y \\ x + 2y \\ 3x + y \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}.$$

That is, we seek a solution to the system

$$\begin{array}{ccccc} 2x & + & y & = & 5 \\ x & + & 2y & = & 1 \\ 3x & + & y & = & 8. \end{array}$$

This system of linear equations is the second interpretation of the matrix equation. We could recognize $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 1 \end{bmatrix}$ as the coefficient matrix of the system. Since solving a system of linear equations is equivalent to solving a matrix equation, we use convenient matrix notation $A\mathbf{x} = \mathbf{b}$ to discuss this particular system. Here A represents the coefficient matrix, the entries of vector $\mathbf{b}$ constitute the right-hand side of the system and the vector $\mathbf{x}$ contains the unknowns as entries.

A third way to view this problem is by taking a vector equation detour.

<u>Matrix Equation</u>	<u>System of Linear Equations</u>	<u>Vector Equation</u>
$\begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}$	$\begin{array}{ccccc} 2x & + & y & = & 5 \\ x & + & 2y & = & 1 \\ 3x & + & y & = & 8 \end{array}$	$x\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + y\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 8 \end{bmatrix}$

These are three equivalent ways of looking at the same thing. We solve all three with the same augmented matrix and using Gauss-Jordan elimination.

$$\left[\begin{array}{cc|c} 2 & 1 & 5 \\ 1 & 2 & 1 \\ 3 & 1 & 8 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right]$$

The reduced row-echelon form on the right above gives values of x and y (namely $x = 3$ and $y = -1$) so the column vector $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ satisfies the matrix equation.

In the matrix form $A\mathbf{x} = \mathbf{b}$ of a system of linear equations, the vector $\mathbf{x}$ is a column vector of unknowns. In light of this, it is natural to think of a solution to such a system as a column vector rather than an n-tuple. We introduce the **vector form for the solution of a linear system** here. If, for example, $(3, -1)$ is the only solution to a particular system, then the vector form for that solution is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

If, on the other hand, a linear system has a general solution of

$$\begin{array}{ccccccc} x & = & 2 & + & 3s & + & 4t \\ y & = & & & s & & \\ z & = & & & & & t \end{array}$$

then the vector form of the solution is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}.$$

The vector form is not the only acceptable way to describe the solution to a linear system, but it is used frequently.

We now move on to define matrix multiplication where the matrix on the right has more than one column.

Definition 1.14. If A is an $m \times p$ matrix and B is a $p \times n$ matrix, then the product AB is defined as an $m \times n$ matrix in which

$$(AB)_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

Chasing the double subscripts around can be very confusing so we do it as seldom as possible, but note that this formula tells us that the first column of the product AB (when $j = 1$) is obtained by dotting each row of A with the first column of B, just as earlier when B was just a column vector. Now that we consider B with more columns, we note that the second column of AB is obtained by dotting each row of A with the second column of B, etc. So, to find the product AB, we multiply A by the first column of B to get the first column of AB, then we repeat this process with each column of B to get the remaining columns of AB.

Thinking of $B = [\mathbf{b}_1, \cdots, \mathbf{b}_n]$ as columns augmented together, then $AB = A[\mathbf{b}_1, \cdots, \mathbf{b}_n] = [A\mathbf{b}_1, \cdots, A\mathbf{b}_n]$. Keep this thought in mind as it will be helpful to think this way later.

We can take this last idea one step further by breaking $A = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$ into m row vectors and $B = [\mathbf{b}_1, \cdots, \mathbf{b}_n]$ into n column vectors. Then, $(AB)_{ij} = \mathbf{a}_i \cdot \mathbf{b}_j$ (vector $\mathbf{a}_i$ dotted with vector $\mathbf{b}_j$). Visually,

$$AB = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} [\mathbf{b}_1, \cdots, \mathbf{b}_n] = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_n \\ \vdots & & \vdots \\ \mathbf{a}_m \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_m \cdot \mathbf{b}_n \end{bmatrix}.$$

Example 1.28

Can we find a 3×3 matrix X such that $AX = B$ where

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 5 & 2 \\ 1 & 7 & 3 \\ 6 & 6 & 5 \end{bmatrix}?$$

Solution We can find each column of X as a solution to each of these matrix equations respectively.

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 6 \end{bmatrix},$$

and

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

We could solve all three with Gauss-Jordan elimination. Since all three have A as the coefficient matrix, there would be a great deal of redundant calculation by solving each one separately. In fact, the calculations would only differ in the last column. To save work, we solve all three at once by augmenting all three last columns to A at once. We separate the three solutions once Gauss-Jordan elimination is complete.

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 3 & 5 & 2 \\ 1 & 2 & 1 & 1 & 7 & 3 \\ 3 & 1 & 1 & 6 & 6 & 5 \end{array}\right] \xrightarrow{r_1 \leftrightarrow r_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 7 & 3 \\ 2 & 1 & 0 & 3 & 5 & 2 \\ 3 & 1 & 1 & 6 & 6 & 5 \end{array}\right]$$

$$\xrightarrow{r_2 \to r_2 - 2r_1, r_3 \to r_3 - 3r_1} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 7 & 3 \\ 0 & -3 & -2 & 1 & -9 & -4 \\ 0 & -5 & -2 & 3 & -15 & -4 \end{array}\right]$$

$$\cdots \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 1 & 0 & 2 \end{array}\right]$$

This shows that if done separately, the final augmented matrices would be

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

The three columns of X are thus

$$\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

making

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

In Example 1.28, it turned out that there was just one possibility for each column of X resulting in one unique X that satisfied the equation $AX = B$. Other matrix equations could result in different kinds of solutions.

Definition 1.15 (Transpose of a Matrix)**.** The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix A^T formed by interchanging the rows and columns of A. That is $(A^T)_{ij} = a_{ji}$.

Example 1.29

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}, \text{ and } \begin{bmatrix} a & b & c \end{bmatrix}^T = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Definition 1.16 (Zero Matrix)**.** For each pair of dimensions m and n, there is an $m \times n$ **zero matrix**, denoted $0_{m,n}$, that has 0 for all of its entries.

When the dimensions are clear or unimportant the zero matrix may be denoted by 0.

Example 1.30

$$0_{2,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } 0_{2,3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The zero matrices serve as the additive identities for matrices because if A is $m \times n$, then $A + 0_{m,n} = 0_{m,n} + A = A$.

Definition 1.17 (Negative of a Matrix)**.** If A is $m \times n$, the negative of A, denoted $-A$, is $m \times n$ and $(-A)_{ij} = -a_{ij}$.

Definition 1.18 (Identity Matrix)**.** For each positive integer n, there is an $n \times n$ **identity matrix**, denoted I_n, that has 1's in the diagonal entries and 0's elsewhere. That is,

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Example 1.31

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The identity matrices serve as the multiplicative identities for matrices because if A is $m \times n$, then $AI_n = A$ and $I_m A = A$. The reader is invited to verify these facts in the exercises.

Matrix operations enjoy many algebraic properties. The following theorem lists many of them.

Theorem 1.5. Let A, B, and C represent matrices and a, b, and c scalars. Assuming the dimensions of the matrices allow the indicated operations, we have the following.

(a) $A + B = B + A$

(b) $A + (B + C) = (A + B) + C$

(c) $a(B + C) = aB + aC$

(d) $a(B - C) = aB - aC$

(e) $(a + b)C = aC + bC$

(f) $(a - b)C = aC - bC$

(g) $A(B + C) = AB + AC$

(h) $(A + B)C = AC + BC$

(i) $A(B - C) = AB - AC$

(j) $(A - B)C = AC - BC$

(k) $a(bC) = (ab)C$

(l) $a(BC) = (aB)C = B(aC)$

(m) $A(BC) = (AB)C$

(n) $A0 = 0$ and $0A = 0$

(o) $(-1)A = -A$

(p) $A + (-1)A = A - A = 0$

(q) $(A + B)^T = A^T + B^T$

(r) $(AB)^T = B^T A^T$ (note the order switch)

It is clear that most of these properties are true because they correspond to properties that you already know are true for real numbers and matrix operations are performed componentwise. They can all be proved using the componentwise definitions provided earlier. Two proofs of medium difficulty are shown below. Many are easier, but property 1.5(m) is more challenging. You are asked to prove others in the exercises. You should be familiar enough with these properties to recognize them when they are used and for you to use them yourself when needed, but you need not memorize them.

Proof (g) Suppose A is $m \times p$ and B and C are $p \times n$. Since B and C are $p \times n$, the sum $B + C$ is also $p \times n$. Since A is $m \times p$ and $(B + C)$ is $p \times n$, $A(B + C)$ is $m \times n$. Similarly, since A is $m \times p$ and B and C are both $p \times n$, we see that products AB and AC are $m \times n$. It follows that $AB + AC$ is $m \times n$. Therefore $A(B + C)$ and $AB + AC$ have the same dimensions. We need only show that their corresponding entries are equal. But

$$\begin{aligned}
(A(B+C))_{ij} &= \sum_{k=1}^{p}(A)_{ik}(B+C)_{kj} \text{ by the definition of matrix multiplication}\\
&= \sum_{k=1}^{p}(A)_{ik}\left[(B)_{kj}+(C)_{kj}\right] \text{ by the definition of matrix addition}\\
&= \sum_{k=1}^{p}\left[(A)_{ik}(B)_{kj}+(A)_{ik}(C)_{kj}\right] \text{ by the distributive property of real numbers}\\
&= \sum_{k=1}^{p}\left[(A)_{ik}(B)_{kj}\right]+\sum_{k=1}^{p}\left[(A)_{ik}(C)_{kj}\right] \text{ by rearranging the terms}\\
&= (AB)_{ij}+(AC)_{ij} \text{ by the definition of matrix multiplication}\\
&= (AB+AC)_{ij} \text{ by the definition of matrix addition.}
\end{aligned}$$

But if the $(i,j)^{th}$ elements of two matrices are equal for all i and j, the matrices are equal. That is, $A(B+C) = AB+AC$.

(r) We now show that $(AB)^T = B^TA^T$. To that end, suppose A is $m\times p$ and B is $p\times n$. The product AB is then $m\times n$ so that $(AB)^T$ is $n\times m$. Similarly, B^T is $n\times p$ and A^T is $p\times m$. It follows that B^TA^T is also $n\times m$. So we need only show that the corresponding entries of the matrices on each side are the same. But by the definition of a matrix transpose,

$$\begin{aligned}
\left((AB)^T\right)_{ij} &= (AB)_{ji}\\
&= \sum_{k=1}^{p}(A)_{jk}(B)_{ki}
\end{aligned}$$

by the definition of matrix multiplication. But these same two definitions together with the commutative property of multiplication in the real numbers gives

$$\begin{aligned}
(B^TA^T)_{i,j} &= \sum_{k=1}^{p}(B^T)_{ik}(A^T)_{kj}\\
&= \sum_{k=1}^{p}(B)_{ki}(A)_{jk}\\
&= \sum_{k=1}^{p}(A)_{jk}(B)_{ki}.
\end{aligned}$$

Therefore, $(AB)^T = B^TA^T$. ■

We have seen that if $A = [\mathbf{a}_1,\cdots,\mathbf{a}_n]$ is $m\times n$ with columns $\mathbf{a}_1,\cdots,\mathbf{a}_n$ and $\mathbf{b} = \begin{bmatrix} b_1\\ \vdots\\ b_n\end{bmatrix}$ is $n\times 1$, then the product $A\mathbf{b} = b_1\mathbf{a}_1+\cdots+b_n\mathbf{a}_n$ is a linear combination of the columns of A. In the same manner, if $A = \begin{bmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_m\end{bmatrix}$ is partitioned into rows and $\mathbf{b} = [b_1\cdots b_n]$ is a row vector then $\mathbf{b}A = b_1\mathbf{a}_1+\cdots+b_n\mathbf{a}_n$ is a linear combination of the rows of A.

From this, we see that every column of the product AB is a linear combination of the columns of A. Similarly, every row of the product BA is a linear combination of the rows of A. This can be seen directly from the definition of matrix multiplication or as an application of $(AB)^T = B^T A^T$. This will be important in section 4.4.

Though the list of algebraic properties of matrices is long, there is one property that is conspicuous by its absence. That property is the commutative property of multiplication. This property is missing because matrix multiplication is *not* commutative as the following example illustrates.

Example 1.32

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 6 & 3 \end{bmatrix}$$

but

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 2 & 7 \end{bmatrix}.$$

There are some other familiar properties of real numbers that do not carry over to matrices. For example, to solve $x^2 - x - 2 = 0$ we first factor to find $(x+1)(x-2) = 0$. This suggests that either $x + 1 = 0$ or $x - 2 = 0$ giving $x = -1$ or $x = 2$. To solve this equation, we used the fact that if a and b are real numbers such that $ab = 0$, then either $a = 0$ or $b = 0$. We say that the real numbers do not have **zero divisors**. But there are zero divisors among real matrices as the next example proves.

Example 1.33

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

Neither of the matrices on the left-hand side of the equation is the 2×2 zero matrix but their product (on the right) is a zero matrix.

Another familiar property of the real number system $\mathbb{R}$ is the cancelling law. We know that if $ab = ac$ and $a \neq 0$, we can 'cancel' the a's and conclude that $b = c$. Can the same be said about matrices? That is, if matrices $AB = AC$ and A is not a zero matrix, can we conclude that $B = C$? No!

Example 1.34

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 5 & 3 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 5 & 3 \end{bmatrix}$$

so that

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

But clearly we also have

$$\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \neq \begin{bmatrix} 4 & 2 \\ 1 & 1 \end{bmatrix}.$$

Problem Set 1.5

In 1 - 6, perform the following matrix operations, compare results of the parts of each problem, and explain.

1. (a) $3\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} + 2\begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix} - 4\begin{bmatrix} -1 \\ 2 \\ -2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 3 & -1 \\ 2 & -1 & 2 \\ 5 & 4 & -2 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$

2. (a) $\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\left(\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}\right)$

(b) $\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$

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

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

4. (a) $\begin{bmatrix} 2 & 1 & 4 \\ -1 & 3 & 5 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 5 & -2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 5 & -2 \end{bmatrix}\begin{bmatrix} 2 & 1 & 4 \\ -1 & 3 & 5 \end{bmatrix}$

5. (a) $\begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 4 \end{bmatrix}$

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

6. (a) $\left(\begin{bmatrix} 4 & 1 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}\right)^T$

(b) $\begin{bmatrix} 4 & 1 \\ 5 & 2 \end{bmatrix}^T\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}^T$

(c) $\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}^T\begin{bmatrix} 4 & 1 \\ 5 & 2 \end{bmatrix}^T$

Each system of linear equations in 7 - 9 can be rewritten as a vector equation or as a matrix equation. Rewrite each of the following in the other two forms and solve.

7. $x - y + 3z = 1$
$2x - y + 6z = 3$
$3x - y + 9z = 5$

8. $x\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ -5 \\ 2 \end{bmatrix}$

9. $\begin{bmatrix} 1 & 1 & 3 & 4 \\ 2 & 3 & 8 & 11 \\ -1 & 1 & 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix}$

10. Determine whether either of the two vectors listed below is a linear combination of the three vectors in the set. If so, write it as a linear combination of the three.
$$\begin{bmatrix} 1 \\ -4 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ -9 \\ 15 \\ 8 \end{bmatrix}, \left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 7 \\ 2 \end{bmatrix} \right\}$$

11. Find the 3×3 matrix X that satisfies the matrix equation $XA = B$, where
$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ 1 & 2 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 3 & 5 & 4 \\ 4 & 7 & 7 \\ 3 & 6 & 4 \end{bmatrix}.$$

12. Describe all 2×2 matrices X that satisfy the matrix equation $AX = B$, where
$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 5 & 3 \\ 10 & 6 \end{bmatrix}.$$

13. Prove parts (e) and (h) of Theorem 1.5.

14. Prove that for all $m \times n$ matrices A, $AI_n = A$ and $I_mA = A$.

15. A square matrix A is **symmetric** if $A^T = A$ and **skew-symmetric** if $A^T = -A$. Let B be a square matrix.

- **(a)** Prove BB^T is symmetric.
- **(b)** Prove $B + B^T$ is symmetric.
- **(c)** Prove $B - B^T$ is skew-symmetric.
- **(d)** Prove that any square matrix is the sum of a symmetric matrix and a skew-symmetric matrix. Hint: Write B as a linear combination of two of the three matrices above.

16. A **diagonal matrix** is a square matrix in which all entries off the main diagonal are 0. That is, a square matrix A is a diagonal matrix if $a_{i,j} = 0$ whenever $i \neq j$. An **upper triangular** matrix is a square matrix in which all entries below the main diagonal are 0. That is, A is an upper triangular matrix if $a_{i,j} = 0$ whenever $i > j$. A **lower triangular** matrix is a square matrix in which all entries above the main diagonal are 0. That is, A is a lower triangular matrix if $a_{i,j} = 0$ whenever $i < j$.

- **(a)** Prove that the sum of two diagonal matrices is diagonal.
- **(b)** Prove that the product of two diagonal matrices is diagonal.
- **(c)** Prove that the product of two upper triangular matrices is upper triangular.
- **(d)** Must the sum and product of two lower triangular matrices be lower triangular?

1.6 Matrix Inverses

In the set of real numbers $\mathbb{R}$ we say that 5 and $\frac{1}{5}$ are **multiplicative inverses** because $5(\frac{1}{5}) = (\frac{1}{5})5 = 1$, the multiplicative identity. In section 1.5, we saw that square matrices with 1's down the main diagonal and 0's elsewhere serve as the identities for matrix multiplication.

Since

$$\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

the two matrices on the left-hand side of the equation appear to be inverses of each other. However, the situation with matrices is more complicated than with real numbers in part because matrix multiplication is not commutative. So we must define multiplicative inverses in matrix arithmetic carefully.

Definition 1.19. Suppose A is an $m \times n$ matrix. A matrix L is a **left inverse** of A if $LA = I_n$. Similarly, a matrix R is a **right inverse** of A if $AR = I_m$. If B is both a left and a right inverse of A, then B is called a **two-sided inverse** of A.

Example 1.35

The multiplication shown above demonstrates that $\begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}$ is a right inverse of $\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$. Reversing order and multiplying yields

$$\begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so $\begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}$ is a left inverse of $\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$ too. Thus, $\begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}$ is a two-sided inverse of $\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$.

Theorem 1.6. Suppose A is an $m \times n$ matrix that has both a left inverse L and a right inverse R. Then $L = R$.

Proof Since $LA = I_n$ and A is $m \times n$, L is $n \times m$. Similarly, $AR = I_m$ and R is $n \times m$. So

$$L = LI_m = L(AR) = (LA)R = I_nR = R.$$

■

If A has both a left and a right inverse, then, in fact, it has a two-sided inverse since they are equal.

Corollary 1.7. Two-sided inverses are unique.

Proof Suppose B and C are two-sided inverses of a matrix A. We show that $B = C$. But, since B is a two-sided inverse of A, B is a left inverse of A. And, since C is a two-sided inverse of A, C is a right inverse of A. By Theorem 1.6, $B = C$. Thus, two-sided inverses are unique. ■

Matrices that are not square can have one-sided inverses. They are interesting and important, but are not usually studied in an introductory linear algebra course. Rather, when dealing with inverses, we restrict our consideration to square matrices where things work out particularly nicely. This is primarily because of the following theorem which we prove at the end of this section.

Theorem 1.8. If A is an $n \times n$ matrix and B is a left inverse of A, then B is a two-sided inverse of A. Similarly, if B is a right inverse of A, then B is a two-sided inverse of A.

Theorem 1.8 allows us to stop discussing left and right inverses when dealing with square matrices except where necessary for clarity and only consider two-sided inverses. Since this is the only type of inverse we consider, we refer to them simply as inverses rather than as two-sided inverses. Since two-sided inverses are unique, we can talk about *the* inverse of A rather than *an* inverse of A and we denote that inverse by A^{-1}.

Every real number except 0 has a multiplicative inverse (its reciprocal). That is not true, however, for every nonzero square matrix.

Example 1.36

The matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ has no inverse. In order for it to have an inverse, there would need to exist a matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

But,

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ a+c & b+d \end{bmatrix},$$

so in order for A to have an inverse we must have two real numbers a and c satisfying $a + c = 1$ and $a + c = 0$. Clearly that is impossible. So A has no inverse.

Definition 1.20. A square matrix that has an inverse is called an **invertible matrix** or a **nonsingular matrix**. A square matrix that has no inverse is called a **singular matrix** or a matrix that is not invertible.

Example 1.37

The matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is not invertible (see Example 1.36). It is singular. On the other hand, we see that $B = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$ does have an inverse so it is an invertible matrix. That is, B is nonsingular and

$$B^{-1} = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}.$$

Theorem 1.9 shows an important connection between solutions of square systems of linear equations and the invertibility of their coefficient matrices.

Theorem 1.9. If A is an invertible $n \times n$ matrix and $\mathbf{b}$ is any n-vector, then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof Substituting $A^{-1}\mathbf{b}$ for $\mathbf{x}$ in the equation $A\mathbf{x} = \mathbf{b}$ we see

$$A(A^{-1})\mathbf{b} = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}.$$

So, $A^{-1}\mathbf{b}$ satisfies the system and is a solution to the system. To show that it is the only solution, we suppose $\mathbf{u}$ is any solution to the system and show that $\mathbf{u} = A^{-1}\mathbf{b}$. Since $\mathbf{u}$ is a solution to the system, it satisfies $A\mathbf{x} = \mathbf{b}$. Thus, $A\mathbf{u} = \mathbf{b}$. Multiplying both sides of this equation on the left by A^{-1} yields the following equalities which complete the proof.

$$\begin{aligned} A^{-1}(A\mathbf{u}) &= A^{-1}\mathbf{b} \\ (A^{-1}A)\mathbf{u} &= A^{-1}\mathbf{b} \\ I_n\mathbf{u} &= A^{-1}\mathbf{b} \\ \mathbf{u} &= A^{-1}\mathbf{b} \end{aligned}$$

■

Finding inverses is particularly simple for 2×2 matrices.

Theorem 1.10. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

If $ad - bc = 0$, then A is not invertible. The quantity $ad - bc$ is called the **determinant** of A.

Proof Suppose that $ad - bc \neq 0$. Note that

$$\left(\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \right) \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} ad - bc & bd - bd \\ ac - ac & ad - bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

It follows that $\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ is a left inverse of A. Using Theorem 1.8 we conclude that it is the inverse of A (i.e. a two-sided inverse of A) so

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Suppose now that $ad - bc = 0$. We prove that A is not invertible. We do this by showing $A\mathbf{x} = \mathbf{0}$ has more than one solution and then we use Theorem 1.9 to conclude A is not invertible. We consider three cases.

Case I. $a = c = 0$. Here, $ad - bc = 0$ and the system

$$\begin{bmatrix} 0 & b \\ 0 & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

has distinct solutions $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Case II. $a \neq 0$, but $ad - bc = 0$. The system

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

has distinct solutions $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -b \\ a \end{bmatrix}$ since

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} -b \\ a \end{bmatrix} = \begin{bmatrix} ab - ab \\ ad - bc \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Case III. $c \neq 0$, but $ad - bc = 0$. Here, the distinct solutions are $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} d \\ -c \end{bmatrix}$.

From these three cases, it follows that if $ad - bc = 0$, A is not invertible. ■

Note that this proof referenced Theorem 1.8, a theorem we have not yet proved. We use Theorem 1.8 in one more proof in this section. We must be careful that we do not use either of these theorems in our proof of Theorem 1.8 so that our logic is sound. We cannot assume what we want to prove. We present this material in this rather convoluted order so that we can present the clearer, more applied results first and save the more theoretical results for later.

Example 1.38

For $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, the determinant of A is $(1)(4) - (2)(3) = -2 \neq 0$. So A is invertible and

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

It follows that the system

$$\begin{array}{rcrcl} x & + & 2y & = & 5 \\ 3x & + & 4y & = & 6 \end{array}$$

has one unique solution

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} -4 \\ \frac{9}{2} \end{bmatrix}.$$

For $B = \begin{bmatrix} 2 & 5 \\ 6 & 15 \end{bmatrix}$, the determinant of B is $(2)(15) - (5)(6) = 0$. So B is singular.

Recall from the last section that if we have more than one system of equations with the same coefficient matrix, we can solve them all at once as illustrated in Example 1.39.

Example 1.39

To solve the three linear systems

$$\begin{matrix} x - y = 1 \\ 2x - y = 3 \end{matrix}, \quad \begin{matrix} x - y = 1 \\ 2x - y = 5 \end{matrix}, \quad \text{and} \quad \begin{matrix} x - y = 1 \\ 2x - y = 7 \end{matrix},$$

form the augmented matrix

$$\left[\begin{array}{cc|ccc} 1 & -1 & 1 & 1 & 1 \\ 2 & -1 & 3 & 5 & 7 \end{array}\right]$$

which can be row reduced to

$$\left[\begin{array}{cc|ccc} 1 & 0 & 2 & 4 & 6 \\ 0 & 1 & 1 & 3 & 5 \end{array}\right].$$

The solutions to the three systems in vector form are

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \text{ and } \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

respectively.

We use the method from Example 1.39 to find inverses of square matrices that are larger than 2×2. But first a little notation.

Definition 1.21. For any positive integer n, define $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ to be the $1^{st}, 2^{nd}, \cdots, n^{th}$ columns of the $n \times n$ identity matrix I_n.

Example 1.40

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ giving } \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ giving } \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

In general the $n \times n$ identity can be partitioned into columns.

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n] \text{ giving } \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$ are sometimes called the **standard basis elements** or **standard unit n-vectors**. This terminology will make more sense later.

To find the inverse of an $n \times n$ matrix A, we seek an $n \times n$ matrix X with the property that $AX = I_n$. If we partition X and I_n into columns, the matrix equation $AX = I_n$ becomes

$$A[\mathbf{x}_1, \cdots, \mathbf{x}_n] = [A\mathbf{x}_1, \cdots, A\mathbf{x}_n] = [\mathbf{e}_1, \cdots, \mathbf{e}_n].$$

Thinking of the columns of this matrix equation separately, we get n different systems of equations with the same coefficient matrix A:

$$A\mathbf{x}_1 = \mathbf{e}_1, A\mathbf{x}_2 = \mathbf{e}_2, \cdots, A\mathbf{x}_n = \mathbf{e}_n.$$

The solution to the first system give us the first column of A^{-1}. The solution to the second system gives us the second column of A^{-1}, etc. But we can solve all of these at once by augmenting the whole identity I_n to A and reducing. In summary, to find A^{-1} we do the following.

1. Augment I_n to A. That is, form matrix $[A|I_n]$.
2. Perform elementary row operations on the augmented matrix to get it into reduced row-echelon form.
3. If I_n appears on the left-hand side of the reduced augmented matrix, then A^{-1} appears on the right-hand side. That is,

$$[A|I_n] \longrightarrow \cdots \longrightarrow [I_n|A^{-1}].$$

Example 1.41

If it exists, we find the inverse of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & -1 \\ -2 & -5 & -1 \end{bmatrix}.$$

To that end,

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 3 & -1 & 0 & 1 & 0 \\ -2 & -5 & -1 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -4 & -1 & 1 & 0 \\ 0 & -1 & 5 & 2 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -4 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -2 & -3 & -3 \\ 0 & 1 & 0 & 3 & 5 & 4 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -8 & -13 & -11 \\ 0 & 1 & 0 & 3 & 5 & 4 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

So,

$$A^{-1} = \begin{bmatrix} -8 & -13 & -11 \\ 3 & 5 & 4 \\ 1 & 1 & 1 \end{bmatrix}.$$

It is easy to check this via matrix multiplication:

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & -1 \\ -2 & -5 & -1 \end{bmatrix}\begin{bmatrix} -8 & -13 & -11 \\ 3 & 5 & 4 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

What happens if A is not invertible? Then there is no matrix X such that $AX = I_n$. In that case, at least one of the n systems we solve is inconsistent and that can only happen if the reduced row-echelon form of A has a row of zeroes.

Example 1.42

The matrix

$$B = \begin{bmatrix} 1 & 3 & 2 \\ 1 & 4 & 1 \\ 2 & 7 & 3 \end{bmatrix}$$

is not invertible. We can see this through row reduction:

$$\left[\begin{array}{ccc|ccc} 1 & 3 & 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 & 1 & 0 \\ 2 & 7 & 3 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -2 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right]$$

We summarize this process with Theorem 1.11.

Theorem 1.11. Let A be an $n \times n$ matrix. To determine whether A is invertible and if so to find A^{-1}, augment the identity matrix I_n to A and reduce to reduced row-echelon form.

- If the reduced row-echelon form of A is I_n, then we get

  $$[A|I_n] \longrightarrow \cdots \longrightarrow [I_n|A^{-1}]$$

  where A^{-1} appears on the right.

- If the reduced row-echelon form of A is not I_n, then the reduced row-echelon form of A has a row of zeroes and we get

  $$[A|I_n] \longrightarrow \cdots \longrightarrow \left[\begin{array}{c|c} * & * \\ 0\cdots0 & * \end{array}\right]$$

  so that A is singular.

Of course when we can find a matrix X such that $AX = I_n$, we have really just found a right inverse of A. We use Theorem 1.8 then to know the solution to the matrix equation is indeed the two-sided inverse. The proof of Theorem 1.11 also depends on Theorem 1.8. We must not use Theorem 1.11 in our proof of Theorem 1.8.

Theorem 1.12 (Properties of Inverses)**.** If A and B are invertible square matrices,

(a) $(A^{-1})^{-1} = A$

(b) $(AB)^{-1} = B^{-1}A^{-1}$ (socks - shoes)

(c) $(A^T)^{-1} = (A^{-1})^T$

Proof

(a) Because A^{-1} is defined as the two-sided inverse of A we know that $A^{-1}A = I_n$ and $AA^{-1} = I_n$. The first equation also tells us that A is a right inverse of A^{-1}, and the second equation tells us that A is a left inverse of A^{-1}. Together this tells us that A is the two-sided inverse of A^{-1}. Therefore, $(A^{-1})^{-1} = A$.

(b) Using properties of matrix multiplication we see that

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1}\\ &= AI_nA^{-1}\\ &= AA^{-1}\\ &= I_n\end{aligned}$$

and

$$\begin{aligned}(B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}A)B\\ &= B^{-1}I_nB\\ &= B^{-1}B\\ &= I_n\end{aligned}$$

This shows that $B^{-1}A^{-1}$ is the two-sided inverse of AB and $(AB)^{-1} = B^{-1}A^{-1}$.

(c) Using properties of matrix multiplication and transpose we see that

$$\begin{aligned}(A^T)(A^{-1})^T &= (A^{-1}A)^T\\ &= (I_n)^T\\ &= I_n\end{aligned}$$

and

$$\begin{aligned}(A^{-1})^T(A^T) &= (AA^{-1})^T\\ &= (I_n)^T\\ &= I_n\end{aligned}$$

This shows that $(A^{-1})^T$ is the two-sided inverse of A^T and $(A^T)^{-1} = (A^{-1})^T$. ■

The proofs of these parts of Theorem 1.12 could have been shortened by the use of Theorem 1.8. Theorem 1.8 was purposely avoided in these proofs so that these properties could be used in the proof of Theorem 1.8.

Corollary 1.13. If $A_1, \cdots, A_k$ are invertible $n \times n$ matrices, then $(A_1 \cdot \cdots \cdot A_k)^{-1} = A_k^{-1} \cdot \cdots \cdot A_1^{-1}$.

Proof Repeatedly use the properties of inverses from Theorem 1.12. More precisely, use induction on k and apply Theorem 1.12. ■

Elementary Matrices

Recall the elementary row operations:

1. (Scaling) Multiply a row by a nonzero scalar.
2. (Interchange) Swap positions of two rows.
3. (Replacement) Replace a row by the sum of itself plus a scalar multiple of another row.

Recall also that the elementary row operations can be undone by other elementary row operations.

Example 1.43

If we scale row two by 5, we can undo that operation by scaling row two by $\frac{1}{5}$.

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix} \xrightarrow{r_2 \to 5r_2} \begin{bmatrix} 1 & 2 & 0 \\ 10 & -5 & 15 \\ 3 & 1 & 2 \end{bmatrix} \xrightarrow{r_2 \to (1/5)r_2} \begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix}$$

If we swap rows two and three, we can undo the operation by swapping rows two and three again.

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 2 \\ 2 & -1 & 3 \end{bmatrix} \xrightarrow{r_3 \leftrightarrow r_2} \begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix}$$

If we replace row one with itself plus 2 times row two (i.e. $r_1 \to r_1 + 2r_2$), we can undo this operation by replacing row one with itself plus -2 times row two (i.e. $r_1 \to r_1 - 2r_2$).

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix} \xrightarrow{r_1 \to r_1 + 2r_2} \begin{bmatrix} 5 & 0 & 6 \\ 10 & 2 & -1 \\ 3 & 1 & 2 \end{bmatrix} \xrightarrow{r_1 \to r_1 - 2r_2} \begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix}$$

Notice too in each case above that not only does the second elementary row operation undo the first, but the first also undoes the second.

Definition 1.22. An **elementary matrix** is a matrix that is obtained from an identity matrix by performing a single elementary row operation.

Example 1.44

Applying a single row operation (as below) to the identity matrix forms elementary matrices.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{r_2 \to 5r_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{r_1 \to r_1 + 2r_2} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The three muliplications below illustrate the fact that multiplication on the left by an elementary matrix has the same effect as performing the corresponding elementary row operation on the matrix. Thus, every time we perform an elementary row operation on a matrix we can think of it as multiplication on the left by an elementary matrix.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 10 & -5 & 15 \\ 3 & 1 & 2 \end{bmatrix}$$

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

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 6 \\ 2 & -1 & 3 \\ 3 & 1 & 2 \end{bmatrix}$$

It should come as no surprise that all elementary matrices are invertible. In fact, the inverse of one is the elementary matrix that corresponds to the elementary row operation that undoes what the first does.

Example 1.45

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

So,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Similarly,

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Theorem 1.14. Elementary matrices are invertible and if E is an elementary matrix obtained from I_n by performing a certain elementary row operation, then E^{-1} is the elementary matrix obtained from I_n by performing the reverse elementary row operation.

Proof Let E by an elementary matrix obtained from I_n by performing a certain elementary row operation and let F by the elementary matrix obtained from I_n by performing the reverse elementary row operation. Start with the identity I_n and perform the first elementary row operation followed by its reverse elementary row operation.

$$I_n \longrightarrow E \longrightarrow I_n$$

Performing elementary row operations are equivalent to multiplication on the left by the corresponding elementary matrix, so the diagram above can be rewritten.

$$I_n \longrightarrow EI_n \longrightarrow F(EI_n)$$

So $F(EI_n) = I_n$. But $EI_n = E$ tells us that $FE = I_n$. That is, F is a left inverse of E. To show that F is a right inverse of E too we repeat the process but this time we start with the reverse elementary row operation corresponding to F.

$$\left\{ \begin{array}{c} I_n \longrightarrow F \longrightarrow I_n \\ I_n \longrightarrow FI_n \longrightarrow E(FI_n) \end{array} \right\} \Longrightarrow EF = I_n$$

Therefore F is the (two-sided) inverse of E and E is invertible with $E^{-1} = F$. ■

Corollary 1.15. A product of elementary matrices is invertible and its inverse is also a product of elementary matrices.

Proof Suppose $E_1, \cdots, E_k$ are elementary matrices. By Theorem 1.14, each E_i $(1 \le i \le k)$ is invertible and each inverse is elementary. By Corollary 1.13, the product $E_k \cdot \cdots \cdot E_1$ is invertible and $(E_k \cdot \cdots \cdot E_1)^{-1} = E_1^{-1} \cdot \cdots \cdot E_k^{-1}$, a product of elementary matrices. ■

Theorem 1.16. Let A be an $n \times n$ matrix. The following are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row-echelon form of A is the identity matrix I_n.

(d) A is a product of elementary matrices.

Proof ((a) $\implies$ (b)) Since A is invertible, by Theorem 1.9 the only solution to $A\mathbf{x} = \mathbf{0}$ is $A^{-1}(\mathbf{0}) = \mathbf{0}$.
((b) $\implies$ (c)) Since $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, every column of A must be a pivot column. This means that every column of the reduced row-echelon form of A contains a leading 1. But since A is square, the reduced row-echelon form of A is I_n.
((c) $\implies$ (d)) Since the reduced row-echelon form of A is I_n there is a sequence of elementary row operations that transforms A into I_n.

$$A \longrightarrow \cdots \longrightarrow I_n$$

Since elementary row operations are reversible, there is a sequence of elementary row operations that transforms I_n to A.

$$I_n \longrightarrow \cdots \longrightarrow A$$

But an elementary row operation has the same effect on a matrix as multiplication on the left by an elementary matrix. So, there is a sequence $E_1, \cdots, E_k$ of elementary matrices such that $E_k \cdot \cdots \cdot E_1 I_n = A$. Therefore $A = E_k \cdot \cdots \cdot E_1$, a product a elementary matrices.
((d) $\implies$ (a)) This is Corollary 1.15. ■

Corollary 1.17. If A and B are $n \times n$ matrices and AB is invertible, then A and B are both invertible.

Proof If the conclusions were false, then at least one of A or B would be singular.
Case I. We suppose B is singular and show that the product AB is singular. Now B singular implies that there exists a nonzero vector $\mathbf{v}$ such that $B\mathbf{v} = \mathbf{0}$. But $(AB)\mathbf{v} = A(B\mathbf{v}) = A(\mathbf{0}) = \mathbf{0}$ shows that the product AB is singular.
Case II. We suppose A is singular and show that AB is singular. If A is singular, there exists $\mathbf{w} \neq \mathbf{0}$ such that $A\mathbf{w} = \mathbf{0}$. By case I, we can assume that B is not singular (i.e. that B is invertible). But then $B^{-1}\mathbf{w} \neq \mathbf{0}$ and $(AB)(B^{-1}\mathbf{w}) = A(BB^{-1})\mathbf{w} = A\mathbf{w} = \mathbf{0}$. It follows that AB is singular. ■

Theorem 1.8. If B is a left inverse of an $n \times n$ matrix A, then in fact B is a two-sided inverse of A. Similarly, if B is a right inverse of A, then B is a two-sided inverse of A.

Proof Suppose B is a left inverse of A. We claim that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Clearly $\mathbf{0}$ is a solution since $A(\mathbf{0}) = \mathbf{0}$. We now suppose that $\mathbf{u}$ is any solution and attempt to show that $\mathbf{u} = \mathbf{0}$. But, since $\mathbf{u}$ is a solution, we have $A\mathbf{u} = \mathbf{0}$. Multiplying both sides on the left by B, we see that $B(A\mathbf{u}) = B(\mathbf{0}) = \mathbf{0}$. On the other hand, $B(A\mathbf{u}) = (BA)\mathbf{u} = I_n\mathbf{u} = \mathbf{u}$. So $\mathbf{u} = \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. By Theorem 1.16, A is invertible and has a two-sided inverse A^{-1}.

Since A^{-1} is a two-sided inverse of A, A^{-1} is a right inverse and B is a left inverse of A so $B = A^{-1}$ by Theorem 1.6 and B is in fact a two-sided inverse.

Suppose B is a right inverse of A. Then $AB = I_n$. So $(AB)^T = I_n^T$ and $B^TA^T = I_n$. This says that B^T is a left inverse of A^T. By the first part of this proof, B^T is in fact the two-sided inverse of A^T (i.e. $(A^T)^{-1} = B^T$). This implies A^T is invertible and so $(A^T)^T = A$ is invertible with $A^{-1} = \left((A^T)^T\right)^{-1} = ((A^T)^{-1})^T = (B^T)^T = B$ by Theorem 1.12(c). ∎

This fills in the theoretical gap we introduced at the beginning of this section. We can be sure now that for *square* matrices you cannot have a left inverse that is not also a right inverse and vice-versa. Square matrices are either invertible and have unique two-sided inverses or else they are singular and have no inverse (neither right nor left).

There are many equivalent ways of saying a square matrix is invertible. Theorem 1.16 gives four ways that are equivalent. Here is a list that we have already proved. We will continue to add to this list in the future.

Theorem 1.18. Let A is an $n \times n$ matrix. The following are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row-echelon form of A is the identity matrix I_n.

(d) A is a product of elementary matrices.

(e) A has n pivot columns.

(f) A has a left inverse.

(g) A has a right inverse.

(h) For all $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(i) Every n-vector $\mathbf{b}$ is a linear combination of the columns of A.

(j) A^T is invertible.

(k) $\mathit{rank}\ A = n$.

(l) $\mathit{nullity}\ A = 0$.

Problem Set 1.6

Use Theorem 1.10 to determine which of the following matrices are invertible and to find the inverse of those that are.

1. $\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}$ **2.** $\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$ **3.** $\begin{bmatrix} 4 & 6 \\ 10 & 15 \end{bmatrix}$

4. $\begin{bmatrix} \sqrt{6} & \sqrt{3} \\ 1 & \sqrt{2} \end{bmatrix}$ **5.** $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ **6.** $\begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix}$

Use Theorem 1.11 to determine which of the following matrices are invertible and to find the inverse of those that are.

7. $\begin{bmatrix} 1 & 2 & 3 \\ -1 & -1 & 1 \\ 2 & 3 & 3 \end{bmatrix}$ **8.** $\begin{bmatrix} 1 & -1 & -1 \\ 2 & -1 & 1 \\ 1 & 1 & 5 \end{bmatrix}$ **9.** $\begin{bmatrix} 2 & 1 & 3 \\ 5 & 4 & 1 \\ 1 & 6 & 0 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ **11.** $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ **12.** $\begin{bmatrix} p & 0 & 0 & 0 \\ 1 & q & 0 & 0 \\ 0 & 1 & r & 0 \\ 0 & 0 & 1 & s \end{bmatrix}, pqrs \neq 0$

13. Solve the following system for x and y in terms of b_1 and b_2. Use the inverse of the coefficient matrix in the process.

$$\begin{aligned} 4x - 7y &= b_1 \\ x + 2y &= b_2 \end{aligned}$$

14. Consider the following sequence of matrices obtained by performing elementary row operations.

$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 1 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & -2 \\ 2 & 5 & 2 \\ 3 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & -2 \\ 0 & 3 & 6 \\ 3 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

From left to right label these four matrices A, B, C, and D.

(a) Find three elementary matrices E_1, E_2, and E_3 such that $B = E_1A$, $C = E_2B$, and $D = E_3C$.

(b) Write D as a product of three elementary matrices times A.

(c) Write A as a product of three elementary matrices times D.

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

(a) Find the reduced row-echelon form of A. Do this by hand.

(b) Write A as a product of elementary matrices.

(c) Write A^{-1} as a product of elementary matrices.

16. Let A be a square matrix with a row of zeros. Prove that A is singular.

17. Suppose that A is invertible and $AB = AC$. Prove that $B = C$.

18. Suppose that A is an $m \times n$ matrix. Prove that there exists an invertible matrix C such that CA is the reduced row-echelon form of A.

19. Suppose A is symmetric and invertible. Prove that A^{-1} is symmetric.

20. Suppose A is upper triangular. Prove that A is invertible if and only if all diagonal entries of A are nonzero.

21. Suppose A is upper triangular and invertible. Prove A^{-1} is upper triangular.

2 Euclidean 2-Space and 3-Space

2.1 Vectors in the Plane and in Space

In chapter 2 we give geometric significance to vectors. The Cartesian (rectangular) coordinate systems provide us with a connection between ordered pairs and points on the plane and between ordered triples and points in space.

Definition 2.1. Let

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x, y \in \mathbb{R} \right\} \text{ and } \mathbb{R}^3 = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x, y, .z \in \mathbb{R} \right\}.$$

In general, let

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{R} \text{ for } i = 1, \cdots, n \right\}.$$

The above sets are read as "R-two," "R-three," and "R-n." We now give three geometric interpretations of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

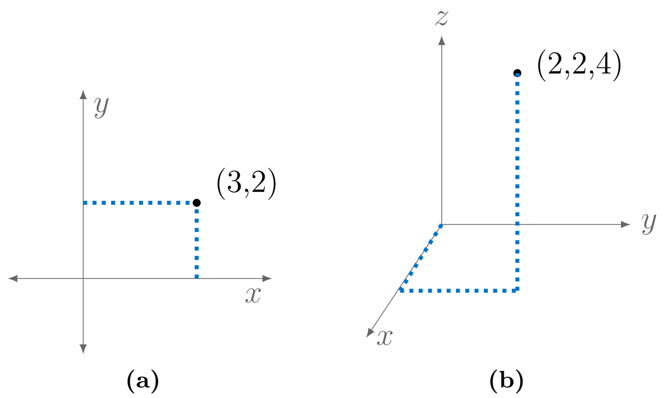
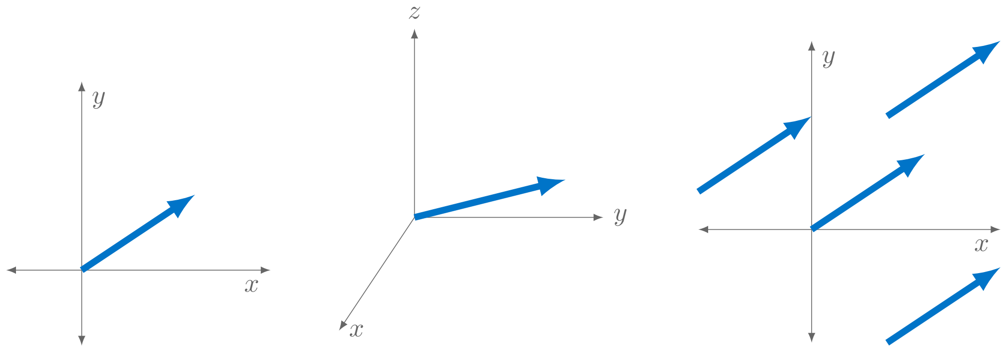


Figure 2.1 *Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ correspond with points in the plane and space.*

1. It is clear that the same connection can be made between the vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ with the points in the plane and space respectively.

The vector sets $\mathbb{R}^2$ and $\mathbb{R}^3$ could be defined as row vectors rather than as column vectors with the analogous connection with points in the plane and space. This is oftentimes simply a matter of convenience. For our purposes column vectors are usually more convenient so we defined $\mathbb{R}^2$ and $\mathbb{R}^3$ that way.

2. Instead of thinking of a vector as representing just a point, think of a vector as representing the directed line segment (arrow) that starts at the origin and ends at that point (see Figures 2.2a and 2.2b).



(a) Vectors in $\mathbb{R}^2$ can be thought of as arrows starting at the origin.

(b) Vectors in $\mathbb{R}^3$ can be thought of as arrows starting at the origin.

(c) Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ can be thought of as a number of different arrows.

Figure 2.2

3. Instead of just one arrow out of the origin, we let any directed line segment that is the same length and points in the same direction as the one in the previous interpretation represent a given vector (see Figure 2.2c).

It is very important to know and work with all three of these geometric interpretations. Sometimes we will use more than one interpretation in the same problem or even in the same equation.

We know how to perform scalar multiplication and vector addition on column vectors. Example 2.1 now illustrates how to interpret these notions geometrically.

Scalar Multiplication

Example 2.1

In Figure 2.3, we illustrate three scalar multiplications:

$$2\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, (-1)\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \text{ and } \frac{1}{2}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}.$$

It is both clear from Example 2.1 and easy to prove using similar triangles that the scalar multiple $c\mathbf{v}$ of the vector $\mathbf{v}$ points in the same direction as $\mathbf{v}$ but is stretched by a factor of c if $c \geq 1$. It shrinks $\mathbf{v}$ by a factor of c if $0 \leq c \leq 1$ and does the same if $c < 0$ except the direction is also reversed. Vectors that are scalar multiples of each other are called **parallel vectors**.

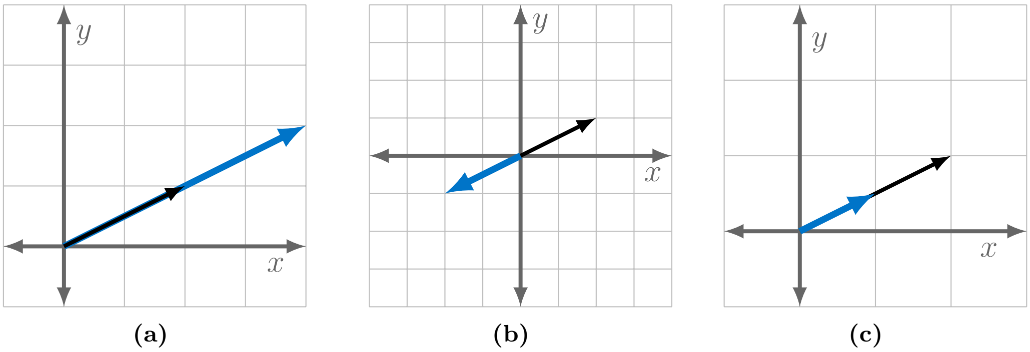


Figure 2.3 *Scalar multiplication by 2, -1, and* $\frac{1}{2}$.

Vector Addition

Example 2.2

In the parallelogram rule for adding, the directed line segments are placed tail to tail and a parallelogram is completed. The diagonal is the sum. Figure 2.4 illustrates the sum

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

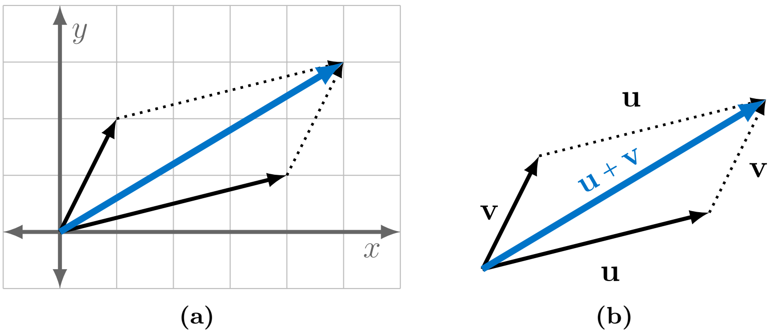


Figure 2.4 *The parallelogram method of vector addition.*

Another equivalent geometric interpretation is the tip-to-tail method. Here the arrow for the second vector starts where the first vector ends. The sum starts where the first one starts and ends where the second one ends.

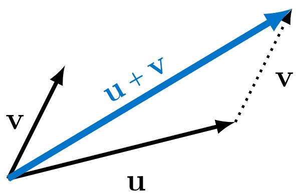
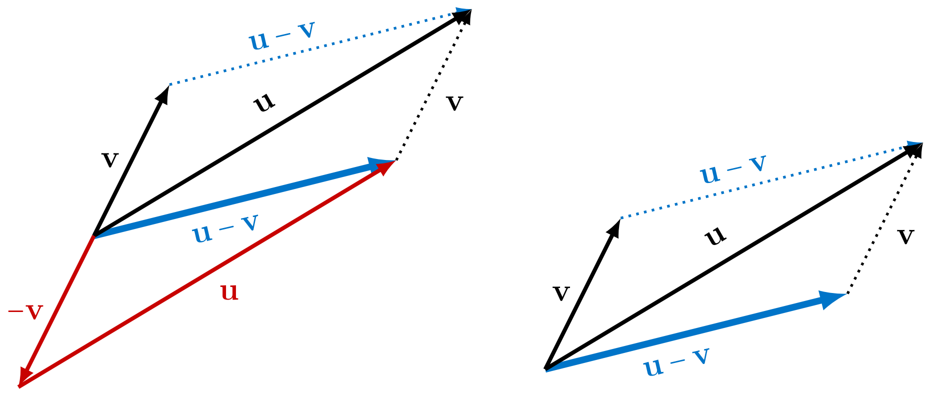


Figure 2.5 *The tip-to-tail method of vector addition.*

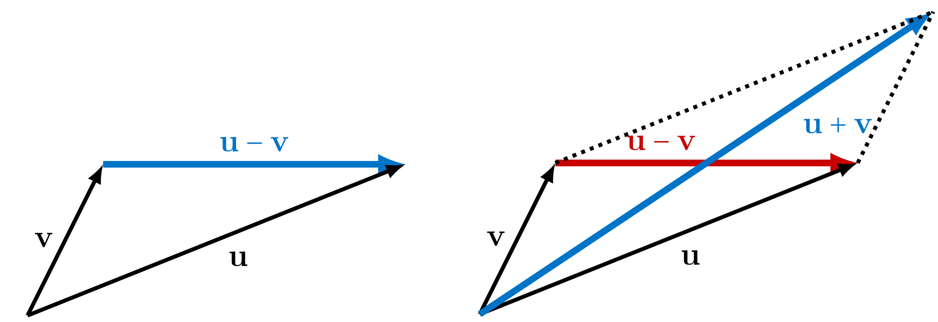
Vector subtraction can be broken down into scalar multiplication and vector addition because $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$. This interpretation is clear (see Figure 2.6a) but a little clumsy. Another way to view $\mathbf{u} - \mathbf{v}$ is by adding it to $\mathbf{v}$ by using the parallelogram method of vector addition to get $\mathbf{u}$ (see Figure 2.6b).



(a) Adding $(-1)\mathbf{v}$ to $\mathbf{u}$. **(b)** Adding $\mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ gives $\mathbf{u}$.

Figure 2.6 *Vector subtraction.*

Stripping away some of the excess information in these previous illustrations, we find a good way to view the difference $\mathbf{u} - \mathbf{v}$. This is called the tip-to-tip interpretation of vector subtraction and is illustrated in Figure 2.7a. Figure 2.7b shows $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ to be the two diagonals in the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$.



(a) Tip-to-tip interpretation of vector subtraction. **(b)** $\mathbf{u} - \mathbf{v}$ is the other diagonal of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$.

Figure 2.7

Linear Combinations

Linear combinations have interesting and very important geometric interpretations.

Example 2.3

Suppose $\mathbf{u}$ and $\mathbf{v}$ are two nonzero, nonparallel vectors in $\mathbb{R}^3$. A linear combination of $\mathbf{u}$ and $\mathbf{v}$, $a\mathbf{u} + b\mathbf{v}$, is just a sum of scaled versions of $\mathbf{u}$ and $\mathbf{v}$ (see Figure 2.8).

There are many possible linear combinations that can be made. To understand those that are possible, look at the plane that contains the *points* $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{v}$ (see Figure 2.9).

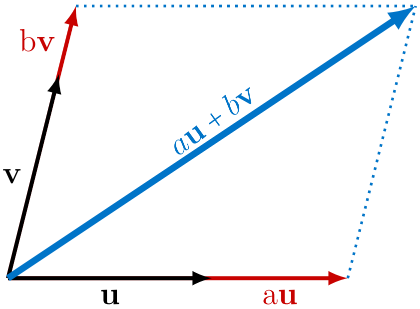


Figure 2.8 *A linear combination of* **u** *and* **v***.*

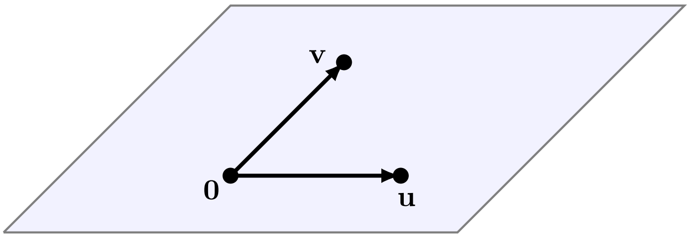


Figure 2.9 *A plane containing points* **0***,* **u** *and* **v***.*

It is easy enough to see that any point on that plane is a linear combination of **u** and **v** (see Figure 2.10).

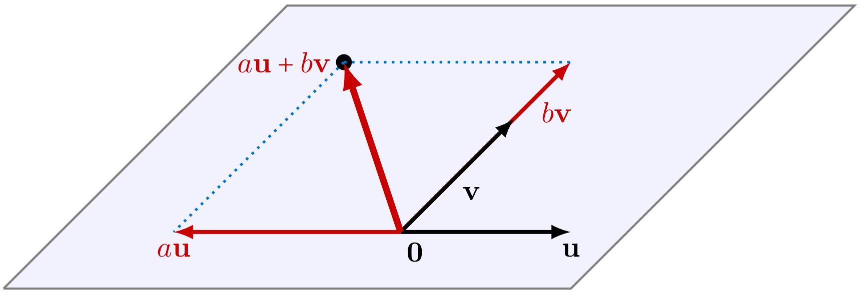


Figure 2.10 *Any point on this plane is a linear combination* $a\mathbf{u} + b\mathbf{v}$ *of* **u** *and* **v***.*

And any vector that does not lie on that plane through **0** is not a linear combination of **u** and **v** because the parallelograms lie entirely on that plane (see Figure 2.11).

Example 2.4

Do the points $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}$ lie on the plane that passes through the points $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$?

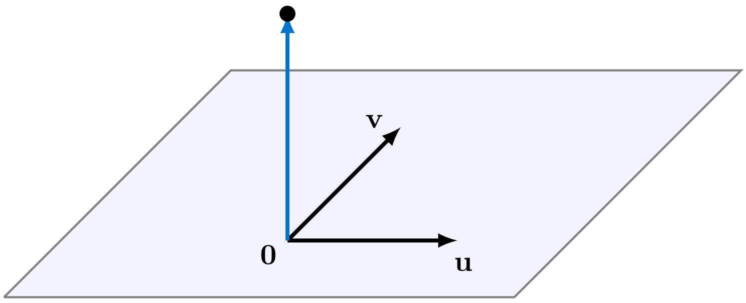


Figure 2.11 *Any point not on this plane is not a linear combination of* **u** *and* **v***.*

Solution To determine whether $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ lies on the plane, we seek to determine whether $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$. That is, are there scalars x and y such that

$$x \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}?$$

This is equivalent to solving the system

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

Similarly, for $\begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}$ we must solve

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}.$$

We can solve these together since they have the same coefficient matrix.

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 3 \\ 2 & 3 & 3 & 5 \\ 1 & 2 & 1 & 4 \end{array}\right] \longrightarrow \left[\begin{array}{cc|cc} 1 & 2 & 1 & 3 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|cc} 1 & 2 & 1 & 3 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|cc} 1 & 0 & 3 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

So $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ lies in that plane through the origin but $\begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}$ doesn't since

$$3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 1 \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

and the second system is inconsistent.

In applications we often think of a vector as something with a magnitude and a direction like wind velocity, which has magnitude (speed) and direction. In $\mathbb{R}^2$ and $\mathbb{R}^3$ we define the magnitude of a vector as the length of its directed line segment. We call this the **norm of a vector**.

Definition 2.2. If $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$, we define the **norm** of $\mathbf{v}$ to be $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$. If $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3$, then $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$. In general, if $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$, then $\|\mathbf{v}\| = \sqrt{v_1^2 + \cdots + v_n^2}$.

Example 2.5

Let $\mathbf{v} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. Then, $\|\mathbf{v}\| = \sqrt{3^2 + (-4)^2} = \sqrt{25} = 5$ and $\|\mathbf{w}\| = \sqrt{1^2 + 2^2 + 1^2} = \sqrt{6}$.

Because multiplication of a vector by a scalar c changes the length of its directed line segment by a factor of c, the following theorem is intuitively clear geometrically. The proof is presented for a vector in $\mathbb{R}^3$. The proof in $\mathbb{R}^2$ or indeed in $\mathbb{R}^n$ in general is analogous.

Theorem 2.1. For any vector $\mathbf{v} \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$, $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.

Proof Suppose $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. Then $c\mathbf{v} = \begin{bmatrix} cv_1 \\ cv_2 \\ cv_3 \end{bmatrix}$ so that

$$\begin{aligned} \|c\mathbf{v}\| &= \sqrt{(cv_1)^2 + (cv_2)^2 + (cv_3)^2} \\ &= \sqrt{c^2(v_1^2 + v_2^2 + v_3^2)} \\ &= \sqrt{c^2}\sqrt{v_1^2 + v_2^2 + v_3^2} \\ &= |c|\|\mathbf{v}\|. \end{aligned}$$

■

Definition 2.3. A **unit vector** is a vector with a norm equal to 1.

In $\mathbb{R}^2$, the standard unit vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are unit vectors in this sense because their norms equal 1. These vectors are usually called $\mathbf{i}$ and $\mathbf{j}$ respectively in multivariable calculus and physics. Of course there are many other unit vectors.

In fact, geometrically we see that any vector represented by an arrow that starts at the origin and ends on the unit circle $x^2 + y^2 = 1$ is a unit vector in $\mathbb{R}^2$ (see Figure 2.12).

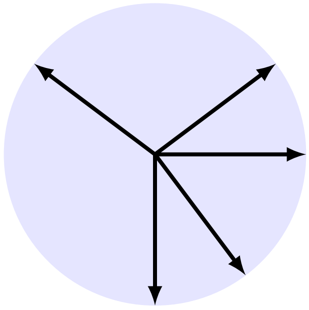


Figure 2.12 *Some unit vectors in* $\mathbb{R}^2$.

In $\mathbb{R}^3$, the standard unit vectors are

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

These are also called $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ in multivariable calculus and physics. Any vector that starts at the origin and ends on the unit sphere $x^2 + y^2 + z^2 = 1$ is a unit vector in $\mathbb{R}^3$.

If $\mathbf{v}$ is any nonzero vector, the **unit vector in the direction of v** is the vector that points in the same direction as $\mathbf{v}$ but has a norm equal to 1.

Let $\mathbf{u}$ be the unit vector in the direction of a nonzero vector $\mathbf{v}$. Since $\mathbf{u}$ points in the same direction as $\mathbf{v}$, $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$. So, there is a scalar $c \in \mathbb{R}$ such that $\mathbf{u} = c\mathbf{v}$. To find $\mathbf{u}$, we need only find this value of c.

From Theorem 2.1, $1 = \|\mathbf{u}\| = \|c\mathbf{v}\| = |c|\|\mathbf{v}\|$. And, since $\mathbf{v}$ is nonzero, $\|\mathbf{v}\| \neq 0$, so we can divide by $\|\mathbf{v}\|$ yielding $|c| = \frac{1}{\|\mathbf{v}\|}$. Solving for c gives $c = \pm\frac{1}{\|\mathbf{v}\|}$. This gives two possible solutions:

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} \text{ and } \mathbf{u} = -\frac{1}{\|\mathbf{v}\|}\mathbf{v}.$$

Since $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ points in the same direction as $\mathbf{v}$ and $-\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ points in the opposite direction, $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ is the unit vector in the direction of $\mathbf{v}$.

Example 2.6

To find the unit vector in the direction of $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$, we calculate $\|\mathbf{v}\| = \sqrt{4+1+4} = 3$ giving us $\mathbf{u} = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$.

Distance Between Vectors

Since we are already familiar with the distance formula between points in the plane and between points in space, we use these formulas to motivate the definition of distance between two vectors. To that end, recall that if P and Q are points in the plane with coordinates (x_1, y_1) and (x_2, y_2), then the distance between them is

$$d(P,Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

If P and Q are points in space with coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2), then the distance between them is

$$d(P,Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

Using points as our geometric interpretation of vectors in $\mathbb{R}^2$, it is natural to define the **distance between two vectors** $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ as

$$d(\mathbf{u},\mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}.$$

If $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ in $\mathbb{R}^3$, we define

$$d(\mathbf{u},\mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}.$$

Note too that by using the tip-to-tip interpretation of vector subtraction, the directed line segment that begins at the point $\mathbf{v}$ and ends at the point $\mathbf{u}$ represents the vector $\mathbf{u} - \mathbf{v}$ so the length of $\mathbf{u} - \mathbf{v}$ equals the distance between $\mathbf{u}$ and $\mathbf{v}$ (see Figure 2.7a). That is,

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

whether $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ or $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. This formula is verified in the exercises.

Problem Set 2.1

1. Let $\mathbf{v}$ be the vector in $\mathbb{R}^3$ represented by the directed line segment $\overrightarrow{PQ}$ that begins at the point P and ends at the point Q, where P and Q have coordinates $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 4 \\ -3 \end{bmatrix}$ respectively.

(a) Write $\mathbf{v}$ as a column vector.

(b) What is the norm of $\mathbf{v}$?

(c) Write the vector represented by the directed line segment $\overrightarrow{QP}$ as a column vector.

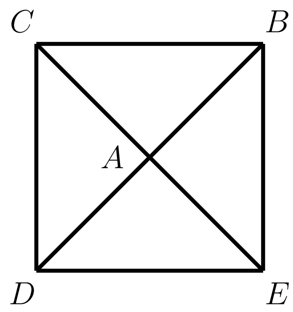


Figure 2.13

2. Let the directed line segments from A to B and from A to C represent the vectors $\mathbf{u}$ and $\mathbf{v}$ respectively ($\mathbf{u} = \overrightarrow{AB}$ and $\mathbf{v} = \overrightarrow{AC}$) from Figure 2.13. Describe the vectors $\overrightarrow{AD}$, $\overrightarrow{AE}$, $\overrightarrow{BC}$, $\overrightarrow{CD}$, $\overrightarrow{DE}$, $\overrightarrow{EB}$ in terms of $\mathbf{u}$ and $\mathbf{v}$. Which of the vectors are parallel?

3. Do the points $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 4 \\ 5 \end{bmatrix}$ lie on the plane in $\mathbb{R}^3$ that passes through the origin and the points $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$?

4. Write the following vectors from $\mathbb{R}^3$ as column vectors.

(a) $3\mathbf{e_1} - 2\mathbf{e_2} + 4\mathbf{e_3}$ **(b)** $4\mathbf{k} - \mathbf{i}$ **(c)** $r\mathbf{j} + s\mathbf{k} - t\mathbf{i}$

5. Write the following vectors from $\mathbb{R}^3$ as linear combinations of the standard unit vectors.

(a) $\begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}$ **(b)** $\begin{bmatrix} 6 \\ 5 \\ 0 \end{bmatrix}$ **(c)** $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

6. Find the norms of the following vectors.

(a) $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ **(b)** $\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$ **(c)** $\begin{bmatrix} 3t \\ 4t \end{bmatrix}$

7. Find the vectors in $\mathbb{R}^2$ with a norm of 2 that make an angle of $\pi/6$ with

(a) the positive x-axis **(b)** the positive y-axis

8. Find the unit vectors in the direction of each of the following vectors.

(a) $\begin{bmatrix} 5 \\ -12 \end{bmatrix}$ **(b)** $\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ **(c)** $\begin{bmatrix} 2t \\ -t \\ 2t \end{bmatrix}$

9. Give a geometric description of the set of all vectors $\mathbf{x}$ in $\mathbb{R}^2$ that satisfy the equation $\|\mathbf{x}\| = 4$. Interpret the vectors geometrically as points.

10. Use the definition of norm of a vector and the definition of distance between two vectors in $\mathbb{R}^2$ to prove that $d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|$.

11. Let $\mathbf{x_0}$ be a fixed point in $\mathbb{R}^3$. Give a geometric description of the set of all vectors $\mathbf{x}$ in $\mathbb{R}^3$ that satisfy the equation $\|\mathbf{x}-\mathbf{x_0}\| = 3$. Interpret the vectors geometrically as points.

2.2 The Dot Product

To complete our geometric interpretation of vectors, we introduce the dot product. The dot product provides us with a means for determining the angle between two vectors.

Definition 2.4. For $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$, we define the **dot product $\mathbf{u}\cdot\mathbf{v}$** by

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2.$$

Similarly, for $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3$,

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3.$$

In general, for $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$,

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + \cdots + u_nv_n.$$

Example 2.7

If $\mathbf{u} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, then $\mathbf{u}\cdot\mathbf{v} = (3)(2) + (1)(-1) = 5$.

Observe that the dot product of two vectors is a scalar. It is not immediately clear how the dot product can be used to answer geometric questions about vectors. However, as we shall soon see, it provides us with a means for determining the angle between two vectors.

It turns out that there are many different ways of defining angles and norms of vectors using things called **inner products**. The dot product is one example of an inner product and is sometimes called the **Euclidean inner product**. It is the most commonly used inner product and the one that matches our geometric intuition best. The vector sets $\mathbb{R}^2$ and $\mathbb{R}^3$ together with vector addition and scalar multiplication as defined earlier and this dot product are called **Euclidean 2-space** E^2 and **Euclidean 3-space** E^3. Euclidean n-space is defined analogously using vectors from $\mathbb{R}^n$.

We begin with some obvious algebraic properties. The proofs of these are all relatively simple. Some appear in the exercises.

Theorem 2.2. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(a). $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

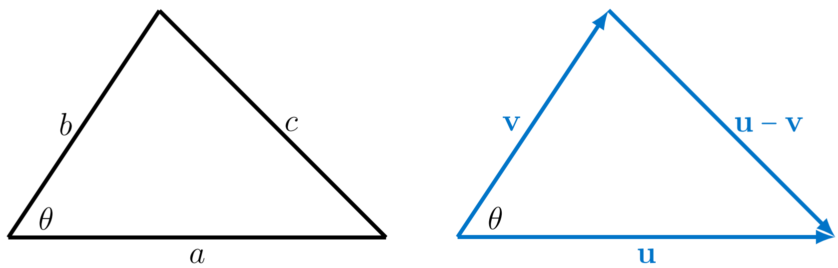
(b). $(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$

(c). $\mathbf{u} \cdot (\mathbf{v} \pm \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} \pm \mathbf{u} \cdot \mathbf{w}$

(d). $(\mathbf{u} \pm \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} \pm \mathbf{v} \cdot \mathbf{w}$

(e). $\|\mathbf{u}\| = (\mathbf{u} \cdot \mathbf{u})^{\frac{1}{2}}$

(f). $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$



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

(b) $\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$

Figure 2.14 *The law of cosines.*

Recall the law of cosines illustrated in Figure 2.14a. Applying the law of cosines to the vectors in Figure 2.14b, we obtain

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta.$$

But, the algebraic rules in Theorem 2.2 yield

$$\begin{aligned}
\|\mathbf{u} - \mathbf{v}\|^2 &= (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) \\
&= (\mathbf{u} - \mathbf{v}) \cdot \mathbf{u} - (\mathbf{u} - \mathbf{v}) \cdot \mathbf{v} \\
&= \mathbf{u} \cdot \mathbf{u} - \mathbf{v} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\
&= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v}
\end{aligned}$$

It follows that

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

which simplifies to the relationship between the dot product and angle θ between the vectors that we seek.

$$\boxed{\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta}$$

Definition 2.5. The angle between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as the smaller but positive angle θ $(0 \le \theta \le \pi)$ satisfying

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}.$$

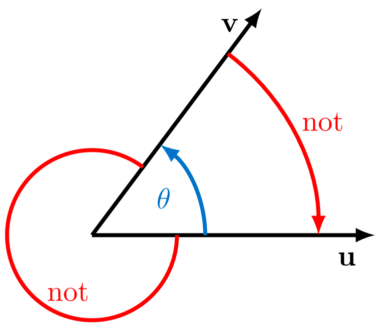


Figure 2.15 *The angle θ between two vectors.*

Example 2.8

The angle between $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ -3 \\ 1 \end{bmatrix}$ is determined by

$$\cos\theta = \frac{(2)(3) + (1)(-3) + (3)(1)}{\sqrt{4+1+9}\sqrt{9+9+1}} = \frac{6}{\sqrt{14}\sqrt{19}}.$$

So, $\theta = \cos^{-1}\left(\frac{6}{\sqrt{14}\sqrt{19}}\right) \approx 68.41^\circ$ or 1.194 radians.

For $\mathbf{u}, \mathbf{v} \ne \mathbf{0}$, we have that $\|\mathbf{u}\| > 0$, $\|\mathbf{v}\| > 0$ and $\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$. So (as in Figure 2.16),

$$\mathbf{u}\cdot\mathbf{v} > 0 \iff \cos\theta > 0 \iff \theta \text{ is acute,}$$

$$\mathbf{u}\cdot\mathbf{v} < 0 \iff \cos\theta < 0 \iff \theta \text{ is obtuse,}$$

$$\mathbf{u}\cdot\mathbf{v} = 0 \iff \cos\theta = 0 \iff \theta = \frac{\pi}{2}.$$

Definition 2.6. We say that $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{u}\cdot\mathbf{v} = 0$.

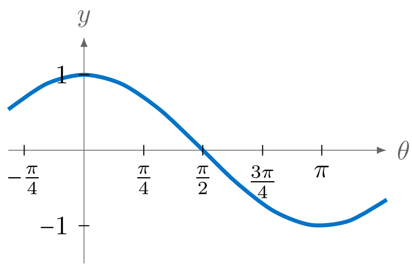


Figure 2.16 $y = \cos\theta$ *for* $0 \le \theta \le \pi$.

By the definition, $\mathbf{0}$ is orthogonal to every vector (including itself) and if $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$, then $\mathbf{u}$ and $\mathbf{v}$ orthogonal means that $\mathbf{u}$ and $\mathbf{v}$ are perpendicular.

Definition 2.7. The **direction angles** α, β, γ of a vector $\mathbf{u}$ are the angles $\mathbf{u}$ makes with the positive x, y, and z axes respectively (see Figures 2.17a, 2.17b, and 2.17c).

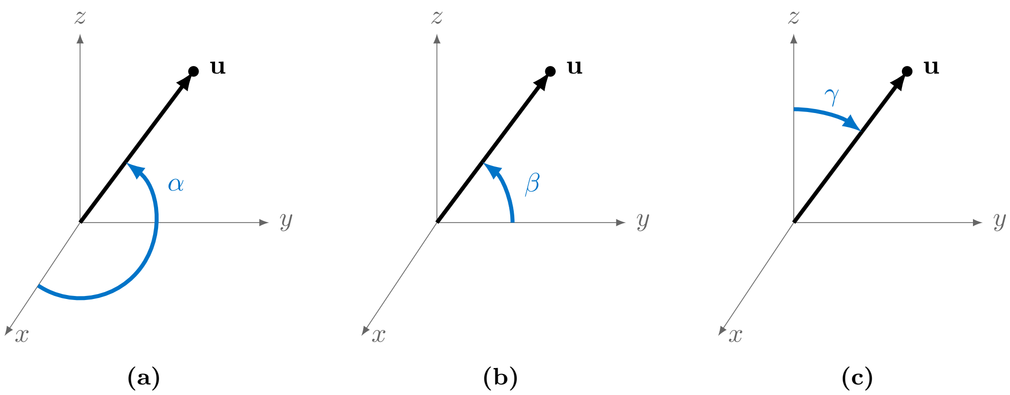


Figure 2.17 *The direction angles* $\alpha, \beta,$ *and* γ *of a vector* $\mathbf{u}$.

For unit vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$, the direction angle α is the angle between $\mathbf{u}$ and the standard unit vector $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. So

$$\cos\alpha = \frac{\mathbf{u} \cdot \mathbf{e}_1}{\|\mathbf{u}\|\|\mathbf{e}_1\|} = \frac{(u_1)(1) + (u_2)(0) + (u_3)(0)}{(1)(1)} = u_1.$$

Similarly, $\cos\beta = u_2$ and $\cos\gamma = u_3$. These are called the **direction cosines** of $\mathbf{u}$. If $\mathbf{v}$ is any nonzero vector in $\mathbb{R}^3$, then by normalizing $\mathbf{v}$ (i.e. finding the unit vector in the direction of $\mathbf{v}$), we find its direction cosines

$$\frac{1}{\|\mathbf{v}\|}\mathbf{v} = \begin{bmatrix} \cos\alpha \\ \cos\beta \\ \cos\gamma \end{bmatrix}.$$

Orthogonal Projection

For $\mathbf{a} \neq \mathbf{0}$, the **orthogonal projection** of a vector $\mathbf{v}$ onto $\mathbf{a}$ is illustrated in Figure 2.18. To calculate the orthogonal projection, note that it is a multiple of $\mathbf{a}$ so that $proj_{\mathbf{a}}\mathbf{v} = k\mathbf{a}$ and we just need to determine k.

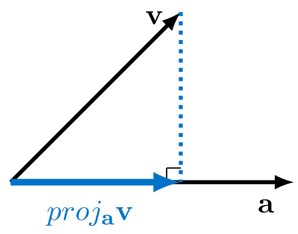


Figure 2.18 *The orthogonal projection of* $\mathbf{v}$ *onto* $\mathbf{a}$*.*

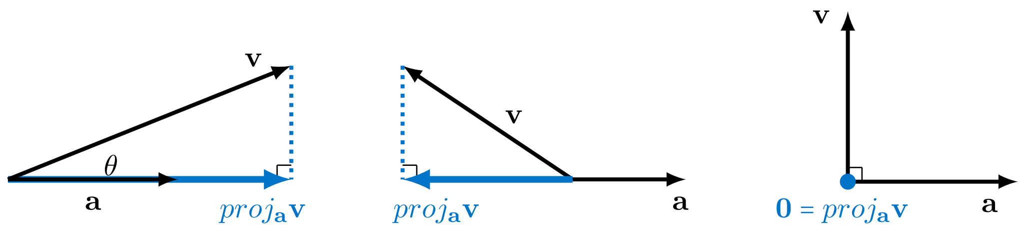
Observe that $\mathbf{v} - proj_{\mathbf{a}}\mathbf{v}$ is orthogonal to $\mathbf{a}$ so that $\mathbf{a} \cdot (\mathbf{v} - k\mathbf{a}) = 0$. This equation can be solved for k:

$$\mathbf{a} \cdot \mathbf{v} - k(\mathbf{a} \cdot \mathbf{a}) = 0$$
$$\frac{\mathbf{a} \cdot \mathbf{v}}{\mathbf{a} \cdot \mathbf{a}} = k$$

So,

$$\boxed{proj_{\mathbf{a}}\mathbf{v} = \frac{\mathbf{a} \cdot \mathbf{v}}{\mathbf{a} \cdot \mathbf{a}}\mathbf{a}.}$$

Figure 2.18 illustrates the orthogonal projection well when $0 < k < 1$. There are similar diagrams that illustrate the orthogonal projection for other values of k (see Figures 2.19a, 2.19b, and 2.19c).



(a) $k > 1$ **(b)** $k < 0$ **(c)** $k = 0$

Figure 2.19 *Visualizing the orthogonal projection.*

Example 2.9

Find the orthogonal projection of $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ onto $\mathbf{a} = \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}$.

We compute:

$$proj_{\mathbf{a}}\mathbf{v} = \frac{\mathbf{a} \cdot \mathbf{v}}{\mathbf{a} \cdot \mathbf{a}}\mathbf{a} = \frac{(1)(2) - (3)(5) + (2)(1)}{4 + 25 + 1}\begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix} = -\frac{11}{30}\begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix} = \begin{bmatrix} -11/15 \\ 11/6 \\ -11/30 \end{bmatrix}$$

Given a vector $\mathbf{v}$ and a nonzero vector $\mathbf{a}$, it is often desireable to write $\mathbf{v}$ as the sum of two component vectors, one parallel to $\mathbf{a}$ and the other orthogonal to $\mathbf{a}$ (see Figure 2.20). Then, $proj_{\mathbf{a}}\mathbf{v}$ serves as the component parallel to $\mathbf{a}$ and $\mathbf{v} - proj_{\mathbf{a}}\mathbf{v}$ serves as the component orthogonal to $\mathbf{a}$. In Example 2.9,

$$\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} -11/15 \\ 11/6 \\ -11/30 \end{bmatrix} = \begin{bmatrix} 26/15 \\ 7/6 \\ 71/30 \end{bmatrix}$$

so

$$\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} -11/15 \\ 11/6 \\ -11/30 \end{bmatrix} + \begin{bmatrix} 26/15 \\ 7/6 \\ 71/30 \end{bmatrix}$$

is the desired decomposition.

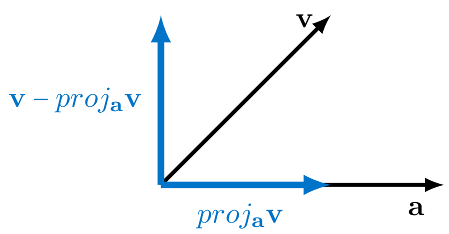


Figure 2.20 *Orthogonal decomposition of* $\mathbf{v}$ *into vectors parallel and orthogonal to* $\mathbf{a}$.

Problem Set 2.2

1. In each case find $\mathbf{u} \cdot \mathbf{v}$.

(a) $\mathbf{u} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ 3 \end{bmatrix}$

(b) $\mathbf{u} = \mathbf{e_1} + 3\mathbf{e_2} - 5\mathbf{e_3}$,
$\mathbf{v} = 3\mathbf{e_1} - 2\mathbf{e_2} - 4\mathbf{e_3}$

(c) $\|\mathbf{u}\| = 3$, $\|\mathbf{v}\| = 5$, and the angle between them is $\theta = \pi/3$.

2. Find the angle between the following pairs of vectors. Give an expression for the exact angle and round to the nearest degree.

(a) $\begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 7 \end{bmatrix}$

(b) $\begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \\ 7 \end{bmatrix}$

(c) $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -14 \\ -2 \\ 5 \end{bmatrix}$

(d) $\begin{bmatrix} 4 \\ -6 \\ 10 \end{bmatrix}, \begin{bmatrix} -6 \\ 9 \\ -15 \end{bmatrix}$

3. In each case find the orthogonal projection of $\mathbf{v}$ onto $\mathbf{a}$.

(a) $\mathbf{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

(b) $\mathbf{v} = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$

(c) $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} -5 \\ 2 \\ 4 \end{bmatrix}$

(d) $\mathbf{v} = \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} 3 \\ -3 \\ 6 \end{bmatrix}$

4. Let $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ and $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$. Decompose $\mathbf{v}$ into the sum of two vectors, one parallel to $\mathbf{a}$ and the other orthogonal to $\mathbf{a}$.

5. Find the direction cosines and the direction angles of the vector $\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$. Give an exact expressions for the angle and round to the nearest degree.

6. Find the unit vectors in $\mathbb{R}^3$ that are orthogonal to both $\mathbf{i} + \mathbf{j}$ and $\mathbf{i} + \mathbf{k}$.

7. **(a)** Find the angle between an edge and a diagonal of a cube that emanate from the same vertex of the cube.

(b) Find the angle between a diagonal of a cube and a diagonal of a face of a cube that emanate from the same vertex of the cube.

8. Find the interior angles of the triangle with vertices at the points P, Q, and R with coordinates $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, and $\begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix}$ respectively.

9. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ t \end{bmatrix}$. Find the values of t that make the following statements true.

(a) $\mathbf{u}$ and $\mathbf{v}$ are parallel.

(b) $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

(c) $\mathbf{u}$ and $\mathbf{v}$ form an angle of $45°$.

10. Prove parts (b) and (c) of Theorem 2.2.

11. Draw two non-parallel directed line segments emanating from the same point and label them $\mathbf{u}$ and $\mathbf{v}$. Complete the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$. Next, draw and label the diagonals of the parallelogram $(\mathbf{u}+\mathbf{v})$ and $\mathbf{u} - \mathbf{v}$. The identity

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$

is known as the **parallelogram rule**.

(a) Interpret the parallelogram rule geometrically.

(b) Prove the parallelogram rule. *Hint:* $\|\mathbf{u} + \mathbf{v}\|^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v})$ and use Theorem 2.2.

12. Prove the identity $\mathbf{u} \cdot \mathbf{v} = \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2$.

2.3 Cross Product

The **cross product** is a different vector product. It is similar to the dot product in some regards, but it is different in some important ways.

- The dot product of two vectors is a scalar, but the cross product of two vectors is another vector.
- Unlike the dot product, the cross product is not an example of an inner product.
- For any positive integer n, if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have $\mathbf{u} \cdot \mathbf{v}$ defined. Unlike the dot product, the cross product only applies to vectors in $\mathbb{R}^3$.

Definition 2.8. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. We define the cross product $\mathbf{u} \times \mathbf{v}$ by

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}.$$

This definition can be difficult to remember. What follows is a mnemonic device that involves determinants of 3×3 matrices. If you have seen such determinants before, this should help you. If not, you might want to learn it anyway because the determinant is the topic of the next chapter. First, recall that

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} a_1 - \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} a_2 + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} a_3.$$

Then, formally, we write

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \\ &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{e}_3 \\ &= (u_2 v_3 - u_3 v_2) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - (u_1 v_3 - u_3 v_1) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (u_1 v_2 - u_2 v_1) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}. \end{aligned}$$

Theorem 2.3 tells us information about the geometry of the cross product.

Theorem 2.3. The vector $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Proof
$$\begin{aligned}\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) &= u_1(u_2v_3 - u_3v_2) + u_2(u_3v_1 - u_1v_3) + u_3(u_1v_2 - u_2v_1) \\ &= u_1u_2v_3 - u_1v_2u_3 + v_1u_2u_3 - u_1u_2v_3 + u_1v_2u_3 - v_1u_2u_3 \\ &= 0\end{aligned}$$
Similarly, $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$. ■

Some straightforward but messy algebraic properties are summarized in Theorem 2.4.

Theorem 2.4. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ and $c \in \mathbb{R}$.

(a) $\mathbf{v} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$

(b) $\mathbf{u} \times (\mathbf{v} \pm \mathbf{w}) = \mathbf{u} \times \mathbf{v} \pm \mathbf{u} \times \mathbf{w}$

(c) $(\mathbf{u} \pm \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} \pm \mathbf{v} \times \mathbf{w}$

(d) $(c\mathbf{u}) \times \mathbf{v} = c(\mathbf{u} \times \mathbf{v}) = \mathbf{u} \times (c\mathbf{v})$

(e) $\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$

(f) $\mathbf{u} \times \mathbf{u} = \mathbf{0}$

One property of the cross product that is helpful but not obvious is that for $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$,

$$\|\mathbf{u} \times \mathbf{v}\| = \sqrt{\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2} = \|\mathbf{u}\|\,\|\mathbf{v}\| \sin\theta$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Why? On the one hand, we know that $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta$ so

$$\begin{aligned}(\mathbf{u} \cdot \mathbf{v})^2 &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2 \cos^2\theta \\ &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2(1 - \sin^2\theta)\end{aligned}$$

This tells us that

$$\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \sin^2\theta = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$$

from which taking square roots gives

$$\|\mathbf{u}\|\,\|\mathbf{v}\| \sin\theta = \pm\sqrt{\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2}.$$

But $\|\mathbf{u}\|, \|\mathbf{v}\| > 0$ and $\sin\theta \geq 0$ for $0 \leq \theta \leq \pi$. So

$$\|\mathbf{u}\|\,\|\mathbf{v}\| \sin\theta = \sqrt{\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2}.$$

On the other hand,

$$\begin{aligned} \|\mathbf{u}\times\mathbf{v}\| &= \sqrt{(u_2v_3 - v_2u_3)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2} \\ &= \sqrt{u_2^2v_3^2 - 2u_2v_3v_2u_3 + v_2^2u_3^2 + \cdots} \end{aligned}$$

and

$$\begin{aligned} \sqrt{\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u}\cdot\mathbf{v})^2} &= \sqrt{(u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3)^2} \\ &= \sqrt{u_1^2v_1^2 + u_1^2v_2^2 + u_1^2v_3^2 + \cdots} \end{aligned}$$

Messy. But it turns out these two are the same. Therefore $\|\mathbf{u}\times\mathbf{v}\| = \sqrt{\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u}\cdot\mathbf{v})^2}$ and, more importantly, we have Theorem 2.5.

Theorem 2.5. For $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ in $\mathbb{R}^3$, $\|\mathbf{u}\times\mathbf{v}\| = \|\mathbf{u}\|\,\|\mathbf{v}\|\sin\theta$.

Theorem 2.5 tells us the norm (length) of $\mathbf{u}\times\mathbf{v}$ but nothing about the direction. Theorem 2.3 tells us the direction of $\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. This leaves two possibilities (see Figure 2.21). Theorem 2.4(a) tells us that one is $\mathbf{u}\times\mathbf{v}$ and the other is $\mathbf{v}\times\mathbf{u} = -(\mathbf{u}\times\mathbf{v})$. The cross product is designed to obey a right-hand rule, so in Figure 2.21 $\mathbf{u}\times\mathbf{v}$ is up and $\mathbf{v}\times\mathbf{u}$ is down.

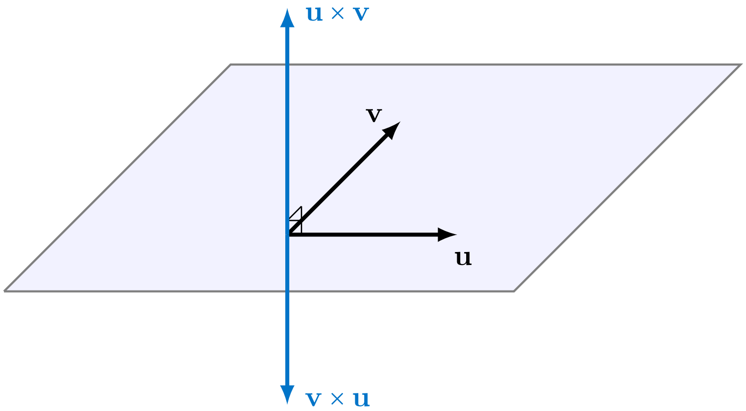


Figure 2.21 $\mathbf{u}\times\mathbf{v}$ *and* $\mathbf{v}\times\mathbf{u}$.

The parallelogram shown in Figure 2.22 gives us another interesting interpretation of the cross product. It says that the area of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$ is $\|\mathbf{u}\|h = \|\mathbf{u}\|\,\|\mathbf{v}\|\sin\theta = \|\mathbf{u}\times\mathbf{v}\|$.

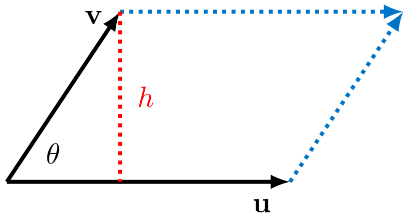


Figure 2.22 $\|\mathbf{u}\times\mathbf{v}\|$ *represents the area of a parallelogram determined by* $\mathbf{u}$ *and* $\mathbf{v}$.

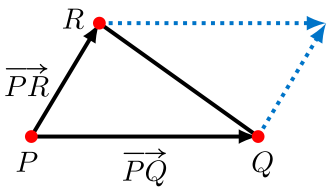


Figure 2.23 *Triangle PQR is half of a parallelogram.*

Example 2.10

Find the area of the triangle with vertices (corners) at the points $P(1,2,2)$, $Q(3,5,1)$, and $R(5,4,3)$.

Solution $\triangle PQR$ is half the parallelogram formed by vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ (see Figure 2.23). Since $\overrightarrow{PQ} = \begin{bmatrix}3\\5\\1\end{bmatrix} - \begin{bmatrix}1\\2\\2\end{bmatrix} = \begin{bmatrix}2\\3\\-1\end{bmatrix}$ and $\overrightarrow{PR} = \begin{bmatrix}5\\4\\3\end{bmatrix} - \begin{bmatrix}1\\2\\2\end{bmatrix} = \begin{bmatrix}4\\2\\1\end{bmatrix}$, we have

$$\overrightarrow{PQ}\times\overrightarrow{PR} = \begin{vmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3\\ 2 & 3 & -1\\ 4 & 2 & 1\end{vmatrix} = (3+2)\mathbf{e}_1 - (2+4)\mathbf{e}_2 + (4-12)\mathbf{e}_3 = \begin{bmatrix}5\\-6\\-8\end{bmatrix}.$$

The area of $\triangle PQR = \frac{1}{2}\|\overrightarrow{PQ}\times\overrightarrow{PR}\| = \frac{1}{2}\sqrt{25+36+64} = \frac{1}{2}\sqrt{125} = \frac{5\sqrt{5}}{2}$.

The Scalar Triple Product

The volume V of the **parallelepiped** determined by $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ can be found by multiplying the area of the base and the height h (see Figure 2.24). But since the base is a parallelogram determined by vectors $\mathbf{u}$ and $\mathbf{v}$, we have

$$\begin{aligned} V &= \|\mathbf{u}\times\mathbf{v}\|h \\ &= \|\mathbf{u}\times\mathbf{v}\|\|\mathbf{w}\|\cos\theta \\ &= (\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}. \end{aligned}$$

The product $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$ is called a **scalar triple product**. If $\mathbf{u}$ and $\mathbf{v}$ swapped positions, $\mathbf{u}\times\mathbf{v}$ would point in the opposite direction changing θ (the angle between $\mathbf{u}\times\mathbf{v}$ and $\mathbf{w}$) to its supplementary angle thus switching the sign of $\cos\theta$. That gives a negative V. Rather

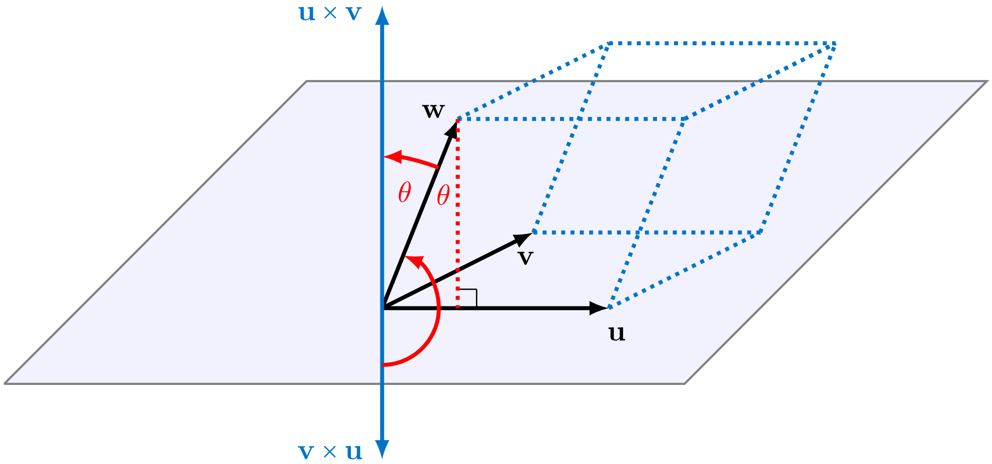


Figure 2.24 $\mathbf{u} \times \mathbf{v}$ *and* $\mathbf{v} \times \mathbf{u}$.

than switching and recalculating, we just take the absolute value when determining the volume:

$$\boxed{V = |(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}|}$$

Example 2.11

Find the volume of the tetrahedron with vertices at the points P, Q, R, and S (see Figure 2.25).

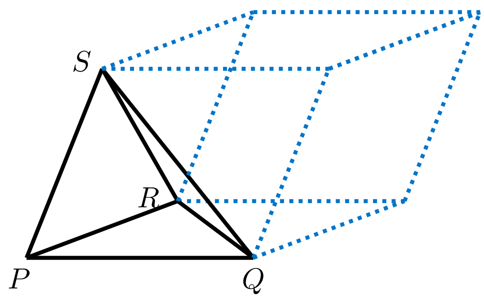


Figure 2.25 *Tetrahedron with vertices at* P, Q, R, *and* S.

The tetrahedron is just one corner of a parallelepiped. In fact, it is a cone. You've studied generalized cones before – objects having cross sections similar to a base that shrink proportionally down to a point at a height h from the base. You learned that the volume is given by

$$V = \frac{1}{3}(\text{area of the base})(\text{height}).$$

The base of our tetrahedron ($\triangle PQR$) is half a parallelogram. So,

$$\begin{aligned} V &= \frac{1}{3}\left(\frac{1}{2}\|\overrightarrow{PQ}\times\overrightarrow{PR}\|\right)(\text{height}) \\ &= \frac{1}{6}\|\overrightarrow{PQ}\times\overrightarrow{PR}\|h \\ &= \frac{1}{6}\left|\|\overrightarrow{PQ}\times\overrightarrow{PR}\|\|\overrightarrow{PS}\|\cos\theta\right| \\ &= \frac{1}{6}\left|(\overrightarrow{PQ}\times\overrightarrow{PR})\cdot\overrightarrow{PS}\right|. \end{aligned}$$

Problem Set 2.3

1. In each case find $\mathbf{u}\times\mathbf{v}$.

(a) $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$

(b) $\mathbf{u} = \mathbf{i}+\mathbf{j}-\mathbf{k}$, $\mathbf{v} = 2\mathbf{i}-3\mathbf{j}$.

2. Find two unit vectors that are orthogonal to both $\begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$.

3. Find the area of the parallelogram that has the vectors $\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$ as two of its sides.

4. Find the area of the triangle with vertices at the points $\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix}$, and $\begin{bmatrix} 6 \\ 1 \\ 4 \end{bmatrix}$.

5. Find the volume of the parallelepiped that has the vectors $\begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$, and $\begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$ as three of its sides.

6. Find the volume of the tetrahedron with vertices at the four points listed below. $\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$, and $\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$.

7. Simplify $(\mathbf{u}+\mathbf{v})\times(\mathbf{u}-\mathbf{v})$.

8. Prove the identity $(\mathbf{u}+k\mathbf{v})\times\mathbf{v} = \mathbf{u}\times\mathbf{v}$.

9. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in $\mathbb{R}^3$, $\mathbf{u}\cdot\mathbf{v}\neq 0$, and θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, then

$$\tan\theta = \frac{\|\mathbf{u}\times\mathbf{v}\|}{\mathbf{u}\cdot\mathbf{v}}.$$

10. Prove that the cross product is not an associative operation by producing a counterexample the demonstrates it is possible that $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \neq \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$.

2.4 Lines in Space

Lines and planes in space are particular types of sets of points. In this section, we develop equations of lines. To understand this development you must understand all three geometric interpretations of vectors (see Section 2.1).

A line in space can be described by a point it passes through and a direction vector. Suppose we wish to describe the line l that passes through the point with coordinates (x_0, y_0, z_0) and that moves in the direction of (i.e. is parallel to) the vector $\mathbf{v} \neq \mathbf{0}$. Our line is thus described by vectors $\mathbf{x}_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

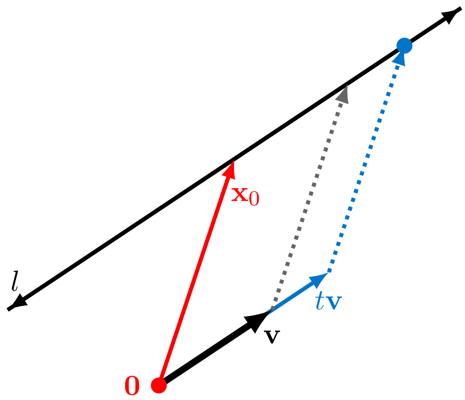


Figure 2.26 *Points on line l are described parametrically by* $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$.

The three variables x, y, and z describe the coordinates of points in space, but a fourth quantity (a parameter) t is also used to help describe a line in space. In addition to $\mathbf{x}_0$, notice that by the parallelogram rule, the sum $\mathbf{x}_0 + \mathbf{v}$ takes us to another point on line l (see Figure 2.26). In fact, *any* point on l can be reached by scaling $\mathbf{v}$ appropriately with some value of the parameter t. That is, the various points on l are all described by $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$ for the various values of the parameter t.

Definition 2.9. Given a point $\mathbf{x}_0$ and a direction vector $\mathbf{v} \neq \mathbf{0}$, the line through the point $\mathbf{x}_0$ and in the direction of $\mathbf{v}$ is described parametrically by the equation

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

This equation of the line is said to be in **vector form** or in **point-parallel form**.

Example 2.12

An equation of the line that passes through the point $(2, 3, -1)$ and is parallel to the vector $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$ in point-parallel form is given by

$$\mathbf{x}(t) \quad = \quad \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} + t \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}. \tag{2.1}$$

A nice physical way to interpret this equation is to think of t as representing time and to think of $\mathbf{x}(t)$ as the position at time t of a particle that is moving along the line. In Example 2.12, the particle is at point $\mathbf{x}(0) = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$ at time $t = 0$ and at time $t = 1$ the particle is at position $\mathbf{x}(1) = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ -3 \end{bmatrix}$.

It is important to realize that different vector equations can describe the same line. For example, the equation

$$\mathbf{x}(t) \quad = \quad \begin{bmatrix} 5 \\ 7 \\ -3 \end{bmatrix} + t \begin{bmatrix} 6 \\ 8 \\ -4 \end{bmatrix} \tag{2.2}$$

looks different but describes the same line as the equation in Example 2.12. To understand this, note that $(5, 7, -3)$ is a different point on the same line. Also note that the direction vectors in the two equations

$$\begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix} \text{ and } \begin{bmatrix} 6 \\ 8 \\ -4 \end{bmatrix} = 2 \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$$

are parallel. So the second equation (2.2) describes a particle starting at a different point on the line when $t = 0$ and the second particle moves in the same direction as the first, but it moves twice as fast.

A single vector equation $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$ has three component parts (one for each coordinate). We can break that down to three separate scalar equations. For example,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} + t \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$$

can be written as

$$\begin{array}{rcrcr} x & = & 2 & + & 3t \\ y & = & 3 & + & 4t \\ z & = & -1 & - & 2t \end{array} \tag{2.3}$$

and more generally

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + t \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

can be written as

$$\begin{array}{rcrcr} x & = & x_0 & + & at \\ y & = & y_0 & + & bt \\ z & = & z_0 & + & ct \end{array} \tag{2.4}$$

The above gives an example and the general form for **parametric equations** of a line. The components of $\mathbf{v}$, the scalars a, b, and c, are called the **direction numbers** of the line l. We can eliminate the parameters t by solving all three equations in (2.3) and (2.4) for t and equating them. For example, (2.3) gives

$$\frac{x-2}{3} = t, \quad \frac{y-3}{4} = t, \quad \frac{z+1}{-2} = t$$

or

$$\frac{x-2}{3} = \frac{y-3}{4} = \frac{z+1}{-2}.$$

More generally, (2.4) gives

$$\frac{x-x_0}{a} = t, \quad \frac{y-y_0}{b} = t, \quad \frac{z-z_0}{c} = t$$

or

$$\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}. \tag{2.5}$$

Equation 2.5 gives another form for describing a line with equations called **symmetric equations**. It is helpful for finding the other coordinates of a point on a line when you only know one.

Example 2.13

For a line l described by the vector equation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} + t \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix},$$

find the other coordinates of the point on l that has an x-coordinate of 8.

Solution The symmetric equations

$$\frac{x-2}{3} = \frac{y-3}{4} = \frac{z+1}{-2}$$

are really three equations. Substituting the value 8 for x and solving we get

$$\frac{8-2}{3} = \frac{y-3}{4} \text{ and } \frac{8-2}{3} = \frac{z+1}{-2}$$

which give $y = 11$ and $z = -5$. So the point on l with an x-coordinate of 8 is $(8, 11, -5)$.

Looking at the general form for symmetric equations in 2.5, it is easy to pick out a point on the line (x_0, y_0, z_0) and in the denominators the direction numbers a, b, and c. We see that there is a problem with the symmetric equations we have presented if one of the direction numbers is 0. Ths next example illustrates and resolves this issue.

Example 2.14

Find symmetric equations for the line through the point $(5, 4, 3)$ and parallel to the vector $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

Solution Substituting into the form presented yields

$$\frac{x-5}{2} = \frac{y-4}{1} = \frac{z-3}{0}$$

and a problem of division by 0. This is not a problem in the vector form

$$\mathbf{x}(t) = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix} + t \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

and the parametric equations

$$\begin{aligned} x &= 5 + 2t \\ y &= 4 + t \\ z &= 3 \end{aligned}$$

show that because the direction number $c = 0$, the line is horizontal. That is, the line has no movement in the z-direction as $z = 3$ always. We describe this with symmetric equations as

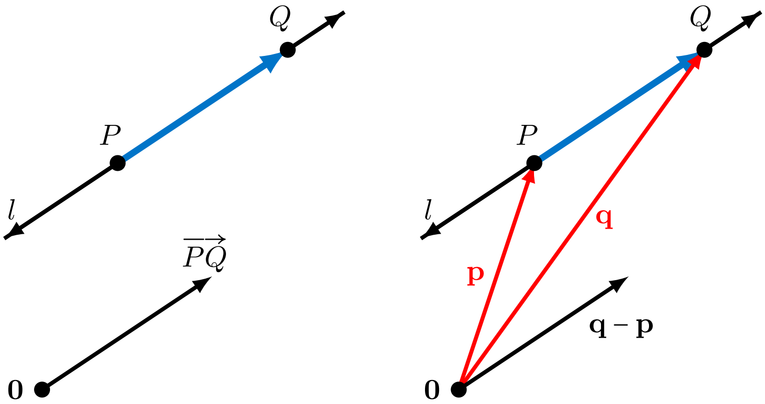
$$\frac{x-5}{2} = \frac{y-4}{1}; z = 3.$$

Example 2.15

Find an equation of the line l through the points $P(3, 2, 7)$ and $Q(1, 10, -2)$ in point-parallel form.

Solution For the point-parallel form, we need a point on the line (we have two of them) and we need a parallel vector. If we let $\mathbf{p} = \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix}$ and $\mathbf{q} = \begin{bmatrix} 1 \\ 10 \\ -2 \end{bmatrix}$, the tip-to-tip interpretation of vector subtraction provides us with the parallel vector

$$\overrightarrow{PQ} = \mathbf{q} - \mathbf{p} = \begin{bmatrix} 1 \\ 10 \\ -2 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} -2 \\ 8 \\ -9 \end{bmatrix}.$$



(a) $\overrightarrow{PQ}$ is parallel to line l.

(b) $\mathbf{q}-\mathbf{p}$ is a vector parallel to line l.

Figure 2.27 *Finding the point-parallel form of a line through points P and Q.*

So

$$\mathbf{x}(t)=\begin{bmatrix}3\\2\\7\end{bmatrix}+t\begin{bmatrix}-2\\8\\-9\end{bmatrix}$$

describes the line in point-parallel form.

There is another form to describe a line called the **two-point form**. Given two points $\mathbf{p}$ and $\mathbf{q}$, one can substitute right into this without the need to calculate a direction vector. It is derived from the point-parallel form having direction vector $\mathbf{q}-\mathbf{p}$.

$$\begin{aligned}\mathbf{x}(t) &= \mathbf{p}+t(\mathbf{q}-\mathbf{p})\\ &= \mathbf{p}+t\mathbf{q}-t\mathbf{p}\\ &= (\mathbf{p}-t\mathbf{p})+t\mathbf{q}\\ &= (1-t)\mathbf{p}+t\mathbf{q}\end{aligned}$$

The two point form places a particle at $\mathbf{p}$ when $t=0$ and at $\mathbf{q}$ when $t=1$. For $0<t<1$, the particle moves in a straight line from $\mathbf{p}$ to $\mathbf{q}$. For $t>1$, the particle is beyond $\mathbf{q}$ and for $t<0$ the particle has not yet reached $\mathbf{p}$.

Example 2.16

In two-point form, the line from Example 2.15 is described by

$$\mathbf{x}(t)=(1-t)\begin{bmatrix}3\\2\\7\end{bmatrix}+t\begin{bmatrix}1\\10\\-2\end{bmatrix}.$$

Example 2.17

Do the lines

$$\mathbf{x}(t)=\begin{bmatrix}4\\4\\5\end{bmatrix}+t\begin{bmatrix}2\\-1\\4\end{bmatrix}\ \text{and}\ \mathbf{x}(t)=\begin{bmatrix}8\\7\\5\end{bmatrix}+t\begin{bmatrix}3\\1\\2\end{bmatrix}$$

intersect?

Solution The parametric equations of these lines are

$$\begin{array}{rcrcr} x &=& 4 &+& 2t \\ y &=& 4 &-& t \\ z &=& 5 &+& 4t \end{array} \quad \text{and} \quad \begin{array}{rcrcr} x &=& 8 &+& 3t \\ y &=& 7 &+& t \\ z &=& 5 &+& 2t \end{array}$$

For the lines to intersect, there has to be a point on the two lines that has the same x-coordinate, y-coordinate, and z-coordinate. This suggests that we should equate the x's, equate the y's, and equate the z's and solve for t. Trying this we get

$$\begin{array}{rcl} 4+2t &=& 8+3t \\ -4 &=& t \end{array} \qquad \begin{array}{rcl} 4-t &=& 7+t \\ -3 &=& 2t \\ -\frac{3}{2} &=& t \end{array} \quad \text{and} \quad \begin{array}{rcl} 5+4t &=& 5+2t \\ 2t &=& 0 \\ t &=& 0 \end{array}$$

We see that the values of t do not agree, so we can't have the x, y, and z coordinates on these lines equal at the same time. Does this mean the lines do not intersect? It is clear that the lines are not parallel because their direction vectors are not parallel. Perhaps the lines are skew.

There is a problem with this reasoning, but it is presented here because it represents a common error. To understand, think of the moving-particle interpretation of these equations and realize that in order for the two lines to intersect it is not necessary for the two particles to be at the same point at the same time. It's just like two cars need not crash just because they are driving on intersecting roads. They may cross the intersection at different times.

To determine whether the lines intersect we seek two possibly different times s and t when all three coordinates are equal. Changing the parameter to s in the first equation (now labeled as $\mathbf{y}(s)$) we get $x = 4+2s$, $y = 4-s$, and $z = 5+4s$. The second equation can still be described with parameter t as $x = 8+3t$, $y = 7+t$, and $z = 5+2t$. Equating these gives

$$\begin{array}{rcl} 4+2s &=& 8+3t \\ 4-s &=& 7+t \\ 5+4s &=& 5+2t. \end{array}$$

The linear system with unknowns s and t in standard form is

$$\begin{array}{rcrcr} 2s &-& 3t &=& 4 \\ -s &-& t &=& 3 \\ 4s &-& 2t &=& 0 \end{array}.$$

Using Gauss-Jordan elimination, we get

$$\left[\begin{array}{cc|c} 2 & -3 & 4 \\ -1 & -1 & 3 \\ 4 & -2 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 1 & -3 \\ 2 & -3 & 4 \\ 2 & -1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 1 & -3 \\ 0 & -5 & 10 \\ 0 & -3 & 6 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 1 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{array}\right].$$

So $s = -1$ and $t = -2$. Checking, we see

$$\mathbf{y}(-1) = \begin{bmatrix} 4 \\ 4 \\ 5 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$

and

$$\mathbf{x}(-2) = \begin{bmatrix} 8 \\ 7 \\ 5 \end{bmatrix} + (-2)\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}.$$

So the two lines indeed *do* intersect at $(2,5,1)$.

Distance Between a Point and a Line

The distance between a point and a line is the shortest distance between the point and all of the points on the line. It is the perpendicular distance.

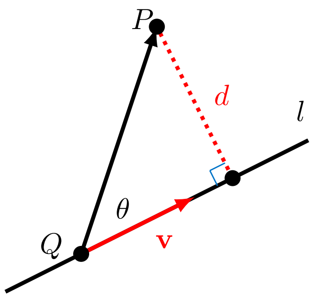


Figure 2.28 *Distance between a point P and a line l.*

Let $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$ be a vector-form equation of a line l and P a point presumably not on the line. Let Q be any point on l (for example, $\mathbf{x}_0$ would work) and let $\mathbf{v}$ be a direction vector for l. We can determine the distance from P to l using simple trigonometry (see Figure 2.28).

$$\begin{aligned} d &= \|\overrightarrow{QP}\| \sin\theta \\ &= \frac{\|\overrightarrow{QP}\|\,\|\mathbf{v}\| \sin\theta}{\|\mathbf{v}\|} \\ &= \frac{\|\overrightarrow{QP} \times \mathbf{v}\|}{\|\mathbf{v}\|} \end{aligned}$$

Example 2.18

Find the distance from the point $P(1,-1,2)$ to the line

$$\mathbf{x}(t) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + t\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}.$$

Solution Let Q be the point $(1,2,3)$. Then

$$\overrightarrow{QP} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ -1 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}.$$

We calculate

$$\overrightarrow{QP} \times \mathbf{v} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 0 & -3 & -1 \\ 2 & 3 & 1 \end{vmatrix} = (-3+3)\mathbf{e}_1 - (0+2)\mathbf{e}_2 + (0+6)\mathbf{e}_3 = \begin{bmatrix} 0 \\ -2 \\ 6 \end{bmatrix}.$$

So

$$d = \frac{\|\overrightarrow{QP} \times \mathbf{v}\|}{\|\mathbf{v}\|} = \frac{\sqrt{0+4+36}}{\sqrt{4+9+1}} = \frac{\sqrt{40}}{\sqrt{14}} = \sqrt{\frac{20}{7}} = \frac{2\sqrt{35}}{7}.$$

Problem Set 2.4

1. Describe each of the following lines in the following ways (i) by writing a vector equation of the line in point-parallel form, (ii) by writing parametric equations for the line, and (iii) by writing symmetric equations for the line.

(a) The line that passes through the point $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and is parallel to the vector $\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$.

(b) The line that passes through the points $\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}$.

(c) The line that passes through the points $\begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}$.

(d) The line that passes through the point $\begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$ and is parallel to the line

$$\mathbf{x}(t) = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}.$$

2. Determine whether the following pairs of lines intersect at a point, are skew, are parallel, or are coincident (the same line). If they intersect, find their intersection.

(a) $\mathbf{x}(t) = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$, $\mathbf{y}(t) = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} + t\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$.

(b) $\begin{array}{rcrcr} x & = & 3 & + & t \\ y & = & 1 & - & t \\ z & = & 1 & + & 3t \end{array}$, $\frac{x-4}{-2} = y - 2 = \frac{z-3}{-4}$

(c) $\frac{x-1}{2} = \frac{y}{-2} = \frac{z-3}{6}$, $\mathbf{x}(t) = (1-t)\begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} + t\begin{bmatrix} -1 \\ 4 \\ -5 \end{bmatrix}$

(d) $\mathbf{x}(t) = (1-t)\begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} + t\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$, $\begin{array}{rcrcr} x & = & 2 & - & 6t \\ y & = & -1 & - & 3t \\ z & = & 1 & + & 9t \end{array}$

3. In each case find the distance from the point to the line.

(a) $\begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{x}(t) = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

(b) $\begin{bmatrix} 2 \\ 5 \\ 4 \end{bmatrix}$, $\begin{array}{rcrcr} x & = & 5 & + & 3t \\ y & = & 3 & + & t \\ z & = & 4 & + & 2t \end{array}$

(c) $\begin{bmatrix} 4 \\ 2 \\ -7 \end{bmatrix}$, $\frac{x-8}{2} = y - 4 = \frac{z-1}{4}$

4. Write (a) a vector equation, (b) parametric equations, and (c) symmetric equations of the line that passes through the point $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and is parallel to the y axis.

5. Find an equation of the line that intersects and is perpendicular to both $\mathbf{x}(s) = \begin{bmatrix} 4 \\ 1 \\ -3 \end{bmatrix} + s\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{y}(t) = \begin{bmatrix} 6 \\ -3 \\ 5 \end{bmatrix} + t\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

2.5 Planes in Space

The basic information needed to write an equation of a line in space is a point on the line and a nonzero direction vector (i.e. a vector parallel to the line). The basic information for an equation of a plane in space is a point on the plane and a nonzero normal vector, that is, a nonzero vector that is perpendicular to the plane.

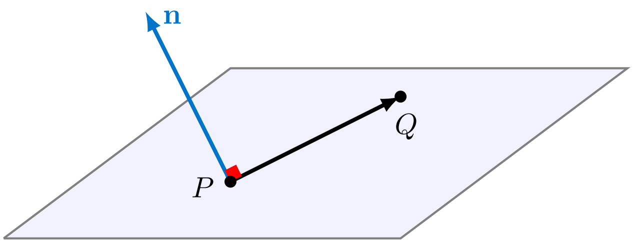


Figure 2.29 *Vector* **n** *is perpendicular to a plane passing through P and Q.*

Let $P(x_0, y_0, z_0)$ be the known point on the plane and $Q(x, y, z)$ any point. Notice that if Q is on the plane, then the vector $\overrightarrow{PQ}$ is orthogonal to $\mathbf{n}$, so $\mathbf{n} \cdot \overrightarrow{PQ} = 0$ and if Q is not on the plane, then $\overrightarrow{PQ}$ is not orthogonal to $\mathbf{n}$, so $\mathbf{n} \cdot \overrightarrow{PQ} \neq 0$. Thus, by determining whether $\mathbf{n} \cdot \overrightarrow{PQ} = 0$ tells us precisely whether Q is on the plane.

Let $\mathbf{x}_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$ be the vector form of P, $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ the vector form of Q, and $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ a nonzero vector normal to the plane. The **point-normal form** for an equation of the plane is

$$\boxed{\mathbf{n} \cdot (\mathbf{x} - \mathbf{x}_0) = 0.}$$

Example 2.19

Find an equation of the plane through $P(3,1,5)$ and normal to $\mathbf{n} = (7,2,4)$.

Solution In point-normal form, we have

$$\begin{bmatrix} 7 \\ 2 \\ 4 \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} \right) = 0.$$

A bit of algebra leads to a linear equation

$$7(x-3) + 2(y-1) + 4(z-5) = 0$$

describing the plane which can be rewritten as

$$7x + 2y - 4z = 43.$$

The same algebra that was performed in Example 2.19, can be done to a general plane in point-normal form

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) = 0.$$

Algebraically, we have

$$a(x-x_0) + b(y-y_0) + c(z-z_0) = 0$$

which can be rewritten as

$$ax + by + cz - (ax_0 + by_0 + cz_0) = 0.$$

If we define $d = ax_0 + by_0 + cz_0$, we can describe this plane with the linear equation

$$\boxed{ax + by + cz = d.}$$

So the graph of a linear equation in three variables from chapter 1 is in fact a plane (not a line) in space.

We have now developed enough machinery to allow us several ways to solve many problems. We present two different methods for solving each of the next two examples.

Example 2.20

Find an equation of the plane that passes through the points $P(1,-3,2)$, $Q(4,2,1)$, and $R(2,1,3)$.

Solution 1 We have

$$\overrightarrow{PQ} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix} \text{ and } \overrightarrow{PR} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}.$$

The perpendicular vector $\mathbf{n}$ can be found from a cross product:

$$\mathbf{n} = \overrightarrow{PQ} \times \overrightarrow{PR} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 3 & 5 & -1 \\ 1 & 4 & 1 \end{vmatrix} = (5+4)\mathbf{e}_1 - (3+1)\mathbf{e}_2 + (12-5)\mathbf{e}_3 = \begin{bmatrix} 9 \\ -4 \\ 7 \end{bmatrix}.$$

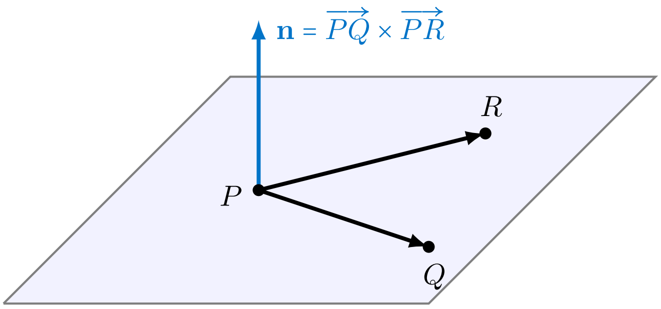


Figure 2.30 *Using a cross product to find an equation of a plane passing through points P, Q and R.*

We then use P for a point $\mathbf{x}_0$ on the plane (though Q or R would work fine too). The equation for this plane is then a simple calculation:

$$\begin{bmatrix} 9 \\ -4 \\ 7 \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} \right) = 0$$

$$\begin{aligned} 9(x-1) - 4(y+3) + 7(z-2) &= 0 \\ 9x - 4y + 7z &= 35 \end{aligned}$$

Solution 2 The plane has a standard form equation $ax + by + cz = d$. For the proper a, b, c, and d, the three points satisfy the equation so we can bring d to the left-hand side (writing $ax + by + cz - d = 0$) and get three linear equations in a, b, c, and d. We can then solve the system

$$\begin{array}{rcrcrcrcl} a & - & 3b & + & 2c & - & d & = & 0 \\ 4a & + & 2b & + & c & - & d & = & 0 \\ 2a & + & b & + & 3c & - & d & = & 0 \end{array}$$

using Gauss-Jordan elimination:

$$\left[\begin{array}{rrrr|r} 1 & -3 & 2 & -1 & 0 \\ 4 & 2 & 1 & -1 & 0 \\ 2 & 1 & 3 & -1 & 0 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{rrrr|r} 1 & 0 & 0 & -9/35 & 0 \\ 0 & 1 & 0 & 4/35 & 0 \\ 0 & 0 & 1 & -1/5 & 0 \end{array}\right]$$

If we let d be represented by a parameter t, we get $a = \frac{9}{35}t$, $b = -\frac{4}{35}t$, $c = \frac{1}{5}t$, and $d = t$. We have (and should expect) infinitely many solutions to this system since any nonzero scalar multiple of an equation of that plane is another equation of the same plane. If we choose $t = 35$ (to avoid fractions), we arrive at the equation $9x - 4y + 7z = 35$ as in solution 1.

Example 2.21

Find an equation for the line of intersection of the two planes $x+y+z = 10$ and $2x+y-z = 6$.

Solution 1 We need a point on the line of intersection and a direction vector for that line. There are many points on that line, but we need only one. We simplify our search by trying to find a point where the x-coordinate is 0. When $x = 0$, the equations simplify to $y + z = 10$ and $y - z = 6$. Solving, we get $y = 8$ and $z = 2$, so $(0, 8, 2)$ lies on the line of intersection. Since a direction vector for the line of intersection is parallel to the line, it is perpendicular to both normal vectors $\mathbf{n}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{n}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$. The cross product $\mathbf{n}_1 \times \mathbf{n}_2$ gives us one.

$$\mathbf{n}_1 \times \mathbf{n}_2 = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 1 & 1 & 1 \\ 2 & 1 & -1 \end{vmatrix} = (-1-1)\mathbf{e}_1 - (-1-2)\mathbf{e}_2 + (1-2)\mathbf{e}_3 = \begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix}$$

So,

$$\mathbf{x}(t) = \begin{bmatrix} 0 \\ 8 \\ 2 \end{bmatrix} + t \begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} \tag{2.6}$$

is the line we seek.

Solution 2 The points on the line of intersection are precisely the solution set to the system

$$\begin{aligned} x + y + z &= 10 \\ 2x + y - z &= 6. \end{aligned}$$

We solve this system through row reduction.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 2 & 1 & -1 & 6 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -2 & -4 \\ 0 & 1 & 3 & 14 \end{array}\right]$$

Setting $z = t$,

$$\begin{array}{rcrcr} x & = & -4 & + & 2t \\ y & = & 14 & - & 3t \\ z & = & & & t \end{array}$$

which can be written in parametric vector form as

$$\mathbf{x}(t) = \begin{bmatrix} -4 \\ 14 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}. \tag{2.7}$$

The equations of the line from solutions 1 and 2 appear to be different but note that their direction vectors are parallel and from equation 2.6

$$\mathbf{x}(2) = \begin{bmatrix} 0 \\ 8 \\ -2 \end{bmatrix} + 2\begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -4 \\ 14 \\ 0 \end{bmatrix},$$

the base point in equation 2.7. So (2.6) and (2.7) both represent the same line.

In many ways lines in the plane and planes in space are comparable. There are many different forms for equations of lines in the plane and a corresponding form for equations of planes in space. The following chart shows the analogs.

	Lines in the Plane	**Planes in Space**
Standard Form	$ax + by = c$ $(a, b$ not both $0)$	$ax + by + cz = d$ $(a, b, c$ not all $0)$
Slope-Intercept Form	$y = mx + b$	$z = m_1 x + m_2 y + b$
Point-Slope Form	$y - y_0 = m(x - x_0)$	$z - z_0 = m_1(x - x_0) + m_2(y - y_0)$
Parametric Equations	$x = x_0 + at$ $y = y_0 + bt$	$x = x_0 + a_1 s + a_2 t$ $y = y_0 + b_1 s + b_2 t$ $z = z_0 + c_1 s + c_2 t$
Vector-Form	$\mathbf{x}(t) = \mathbf{x_0} + t\mathbf{v}$ $\mathbf{v} \neq \mathbf{0}$	$\mathbf{x}(s, t) = \mathbf{x_0} + s\mathbf{u} + t\mathbf{v}$ $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ and not parallel
Point-Parallel Form	$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + t\begin{bmatrix} a \\ b \end{bmatrix}$	$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + s\begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} + t\begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$
Point-Normal Form	$\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) = 0$ $\mathbf{n} \neq \mathbf{0}$ $\begin{bmatrix} a \\ b \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right) = 0$	$\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) = 0$ $\mathbf{n} \neq \mathbf{0}$ $\begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) = 0$

A few comments are in order.

1. All lines in the plane and all planes in space have equations in all of these forms but for one type of exception. Vertical lines and vertical planes do not fit into the slope-intercept and the point-slope forms because the slopes are undefined.
2. The slopes m_1 and m_2 represent $\frac{\Delta z}{\Delta x}$ while y is held fixed and $\frac{\Delta z}{\Delta y}$ while x is held fixed.
3. The vector form or point-parallel form for a plane in space was touched on in section 2.1 where it was noted that the set of all linear combinations of two nonparallel vectors in $\mathbb{R}^3$ is a plane through the origin. Thus, for any s and t, $s\mathbf{u} + t\mathbf{v}$ produces a point on the plane through the origin determined by $\mathbf{u}$ and $\mathbf{v}$ (that $\mathbf{u}$ and $\mathbf{v}$ lie on) adding fixed point $\mathbf{x}_0$ to these shifts the plane, so the plane passes through $\mathbf{x}_0$ rather than the origin.

4. To find a point-normal form if given the point-parallel form in $\mathbb{R}^3$, simply note that we may let $\mathbf{n} = \mathbf{u} \times \mathbf{v}$.

5. To find a point-parallel form given a standard form, simply solve for one variable in terms of the other two as Example 2.22 illustrates.

Example 2.22

Write a point-parallel equation for the plane $2x + y - 3z = 5$.

Solution In this example we solve for y: $y = 5 - 2x + 3z$. Since this is just a short system with only one equation (and three unknowns), we let $x = s$ and $z = t$ to write the solution as parametric equations

$$\begin{array}{ccccccc} x & = & & & s & & \\ y & = & 5 & - & 2s & + & 3t \\ z & = & & & & & t \end{array}$$

which can also be placed in the vector form

$$\mathbf{x}(s,t) = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}. \tag{2.8}$$

Now the vector form (2.8) and the standard form $2x + y - 3z = 5$ are very different descriptions of the same plane. They are each helpful in their own way. In order to generate many different points on the plane, the vector form (2.8) is helpful. By substituting in many different values for s and t, we generate different points. To check whether a particular point of interest lies on the plane, use the equation $2x + y - 3z = 5$ and substitute in the coordinates of the point to see whether the equation is satisfied.

Distance from a Point to a Plane

Let π be a plane and P a point presumably not on the plane. Let $\mathbf{n}$ be a vector normal to π. We see from Figure 2.31 that given any point Q on π, the distance from P to π is

$$d = \|proj_{\mathbf{n}} \overrightarrow{PQ}\|.$$

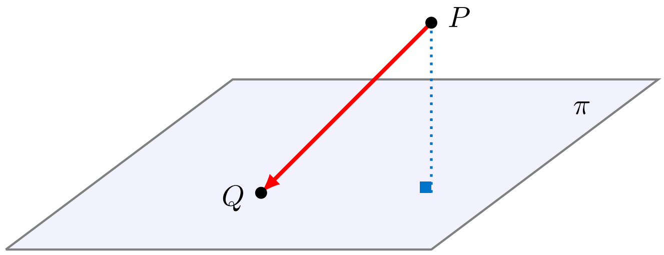


Figure 2.31 *The distance from point P to plane π is the length of a projection.*

Example 2.23

Find the distance from $P(1,1,3)$ to the plane $2x - y + z = 7$.

Solution Clearly the point $(0,0,7)$ is on the plane, so we choose that for Q. This gives

$$\overrightarrow{PQ} = \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix} \text{ and } \mathbf{n} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

so that

$$proj_{\mathbf{n}}\overrightarrow{PQ} = \frac{\overrightarrow{PQ}\cdot\mathbf{n}}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n} = \frac{(-1)(2)+(-1)(-1)+(4)(1)}{4+1+1}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \frac{3}{6}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

and

$$d = \left\| proj_{\mathbf{n}}\overrightarrow{PQ} \right\| = \left\| \frac{1}{2}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \right\| = \frac{1}{2}\sqrt{4+1+1} = \frac{\sqrt{6}}{2}.$$

Example 2.24

Determine whether the lines

$$\mathbf{x}(t) = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{x}(t) = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + t\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

coincided, intersect at a point, are parallel, or are skew.

Solution Clearly, these lines do not coincide and are not parallel because their direction vectors $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ are not parallel. We check to see whether they intersect as in Example 2.17 by solving

$$\begin{aligned} 1+s &= 2+3t \\ 2+2s &= 3+t \\ s &= 4+2t. \end{aligned}$$

This linear system is put into standard form that can then be row reduced.

$$\begin{aligned} s-3t &= 1 \\ 2s-t &= 1 \\ s-2t &= 4. \end{aligned}$$

$$\left[\begin{array}{cc|c} 1 & -3 & 1 \\ 2 & -1 & 1 \\ 1 & -2 & 4 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & -3 & 1 \\ 0 & 5 & -1 \\ 0 & 3 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & -3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & -16 \end{array}\right]$$

The system is inconsistent so the lines do not intersect. The lines are skew because they do not intersect and are not parallel.

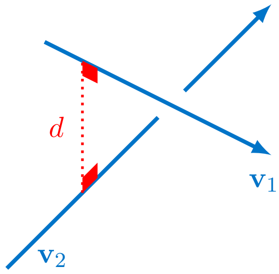


Figure 2.32 *Skew lines are not parallel and do not intersect.*

Example 2.25

Find the distance between the skew lines in Example 2.24.

Solution Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix},$$

the direction vectors of each of the lines. Then

$$\mathbf{v}_1 \times \mathbf{v}_2 = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 1 & 2 & 1 \\ 3 & 1 & 2 \end{vmatrix} = (4-1)\mathbf{e}_1 - (2-3)\mathbf{e}_2 + (1-6)\mathbf{e}_3 = \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}$$

is perpendicular to both $\mathbf{v}_1$ and $\mathbf{v}_2$, so the plane

$$\begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right) = 0$$

contains the first line and is parallel to the second. Hence the distance between the two lines equals the distance from $P(2,3,4)$ to the plane. We choose $(1,2,0)$ for Q and $\mathbf{n} = \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}$. It follows that

$$\overrightarrow{PQ} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -4 \end{bmatrix}$$

and

$$proj_{\mathbf{n}}\overrightarrow{PQ} = \frac{(-1)(3) + (-1)(1) + (-4)(-5)}{9+1+25} \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix} = \frac{16}{35} \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}.$$

So the distance between the lines is

$$d = \|proj_{\mathbf{n}}\overrightarrow{PQ}\| = \frac{16}{35}\sqrt{35}.$$

Problem Set 2.5

1. Find the equations in standard form of the following planes:

(a) through the point $(1,-3,0)$ with a normal vector of $\begin{bmatrix} 2 \\ 5 \\ -4 \end{bmatrix}$.

(b) through the points $(1,4,3)$, $(2,6,4)$, and $(6,2,-1)$.

(c) that contains the two intersecting lines $\mathbf{x}(t) = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} + t\begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$ and $\mathbf{x}(t) = \begin{bmatrix} 5 \\ 2 \\ -2 \end{bmatrix} + t\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$. *(Make sure the lines intersect.)*

(d) through the point $(2,1,-2)$ and perpendicular to the line $\frac{x-1}{3} = \frac{y+2}{4} = \frac{z-3}{5}$.

(e) through the point $(1,-1,3)$ and parallel to the plane $2x+3y-z=1$.

(f) that contains the line

$$\begin{array}{rcrcr} x & = & 1 & + & 3t \\ y & = & 2 & - & t \\ z & = & 4 & + & t \end{array}$$

and is perpendicular to the plane $x+3y-2z=4$. *(Planes are perpendicular if their normal vectors are perpendicular.)*

(g) through the point $(4,1,3)$ and perpendicular to the planes $x+2y-z=4$ and $3x-y+2z=1$.

2. Write the equation of the plane from Exercise 1(a) in the following six forms:

(a) standard form **(b)** point-normal form **(c)** slope-intercept form

(d) point-slope form **(e)** parametric equations **(f)** point-parallel form

3. Find the equation of the line of intersection of the planes $x+3y-2z=1$ and $2x+7y+z=4$.

4. Find the equation of the line on the plane $\mathbf{x}(s,t) = \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} + s\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}$, through the point $(2,6,-2)$ and perpendicular to the vector $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

5. Find the intersection of the planes $x+3y+2z=4$, $2x+7y+3z=9$, and $x+5y+z=9$.

6. Find the intersection of the plane $2x+3y-z=4$ and the line

$$\mathbf{x}(t) = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} + t\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}.$$

7. Find the distance between the point $(5,1,4)$ and the plane $2x+3y+4z=6$.

8. **(a)** Find the distance between the skew lines

$$\mathbf{x}(s) = \begin{bmatrix} 6 \\ -1 \\ -1 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{y}(t) = \begin{bmatrix} -7 \\ -6 \\ -1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

(b) Find the two points (one from each line in part (a)) such that the distance between them equals the distance between the two lines.

9. **(a)** Show that the line $\mathbf{x}(t) = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$ and the plane $2x + 3y - 3z = 2$ are parallel.

(b) Find the distance between the line and the plane in part (a).

10. **(a)** Show that the lines $\mathbf{x}(t) = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} + t \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix}$ and $\mathbf{x}(t) = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} + t \begin{bmatrix} -2 \\ 4 \\ -6 \end{bmatrix}$ are parallel.

(b) Find the distance between the parallel lines in part (a).

11. The standard form for the equation of a plane is $ax + by + cz = d$ (a, b, c not all 0). In each part below write a sentence or two to describe geometrically the planes with equations that have the following characteristics.

(a) $a = b = 0$ **(b)** $b = c = 0$ **(c)** $c = 0$ **(d)** $d = 0$

12. Recall the symmetric equations for a line, $\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}$, $(a, b, c \neq 0)$.

(a) Describe geometrically the set of points that satisfy the equation $\frac{x-x_0}{a} = \frac{y-y_0}{b}$, and describe the set of points that satisfy the equation $\frac{y-y_0}{b} = \frac{z-z_0}{c}$.

(b) What is the connection between your answers in part (a) and the line described by the symmetric equations above.

3 Determinants

In section 1.6, we learned that a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad-bc \neq 0$. We called $ad-bc$ the determinant of the matrix. In this chapter, we learn more properties of the determinant and how the definition extends to larger square matrices.

3.1 The Definition of Determinant

There are several equivalent ways to define the determinant of an $n \times n$ matrix. We take a standard (and probably the easiest) definition. It is a recursive definition, which means that the definition of the determinant of an $n \times n$ matrix is given in terms of determinants of $(n-1) \times (n-1)$ matrices. For example, the determinant of a 5×5 matrix is given in terms of determinants of 4×4 matrices which, in turn, are defined in terms of determinants of 3×3 matrices, etc. Of course, we already know how to calculate the determinant of a 2×2 matrix, but we are going to start even smaller.

Definition 3.1. If A is a 1×1 matrix, then there is a real number a such that $A = [a]$. We define the **determinant of** A to be $\det A = a$.

Definition 3.2. Let A be an $n \times n$ matrix with $n > 1$. For each i and j between 1 and n, define A_{ij} to be the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j from A. The matrix A_{ij} is called the (i, j) **submatrix of** A and its determinant

$$M_{ij} = \det A_{ij}$$

is called the (i, j) **minor of** A.

Example 3.1

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Then,

$$A_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} \text{ and } M_{11} = (5)(9) - (6)(8) = -3,$$

$$A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} \text{ and } M_{12} = (4)(9) - (6)(7) = -6,$$

and

$$A_{22} = \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix} \text{ and } M_{22} = (1)(9) - (3)(7) = -12.$$

Definition 3.3. Suppose A is an $n \times n$ matrix and $n \geq 2$. Define the **determinant of** A as

$$\begin{aligned} \det A &= a_{11}M_{11} - a_{12}M_{12} + \cdots + (-1)^{1+n}a_{1n}M_{1n} \\ &= \Sigma_{j=1}^{n}(-1)^{1+j}a_{1j}M_{1j} \end{aligned}$$

The notation $|A|$ is also used to denote the determinant of A.

Example 3.2

Compute the determinant of

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

Solution

$$\begin{aligned} \det A &= (1)\det\begin{bmatrix} 0 & 1 \\ 3 & 4 \end{bmatrix} - (2)\det\begin{bmatrix} 5 & 1 \\ -2 & 4 \end{bmatrix} + (-3)\det\begin{bmatrix} 5 & 0 \\ -2 & 3 \end{bmatrix} \\ &= (1)(0-3) - (2)(20+2) + (-3)(15-0) \\ &= -3 - 44 - 45 \\ &= -92 \end{aligned}$$

It is not clear at this point what this number, the determinant, indicates about the matrix.

Using the recursive definition to calculate the determinant of a general 2×2 matrix, we get

$$\begin{aligned} \det\begin{bmatrix} a & b \\ c & d \end{bmatrix} &= a\det[d] - b\det[c] \\ &= ad - bc. \end{aligned}$$

So this recursive definition agrees with the definition presented in section 1.6 for 2×2 matrices.

Definition 3.4. If A is an $n \times n$ matrix and i and j are between 1 and n, then we define the (i, j) **cofactor of** A to be

$$C_{ij} = (-1)^{i+j}\det A_{ij} = (-1)^{i+j}M_{ij}.$$

Combining Definition 3.4 with the recursive definition of determinant, we get

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} = \Sigma_{j=1}^{n} a_{1j}C_{1j}.$$

This is called the **cofactor expansion of** $\det A$ across the first row.

It turns out that we obtain the same result by expanding across any row or down any column of a matrix.

Theorem 3.1 (Laplace Expansion, Cofactor Expansion, or Expansion by Minors)**.** Let A be an $n \times n$ matrix. For any fixed i between 1 and n,

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \Sigma_{j=1}^{n} a_{ij}C_{ij}$$

and for any fixed j between 1 and n,

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \Sigma_{i=1}^{n} a_{ij}C_{ij}.$$

The proof of Theorem 3.1 is too time consuming to present here, but we illustrate it by expanding Example 3.2 across the second row and down the third column instead of across the first row but with the same result.

Example 3.3

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & -3 \\ 5 & 0 & 1 \\ -2 & 3 & 4 \end{vmatrix} &= (-1)^{2+1}(5)\begin{vmatrix} 2 & -3 \\ 3 & 4 \end{vmatrix} + (-1)^{2+2}(0)\begin{vmatrix} 1 & -3 \\ -2 & 4 \end{vmatrix} + (-1)^{2+3}(1)\begin{vmatrix} 1 & 2 \\ -2 & 3 \end{vmatrix} \\
&= -5(8+9) + 0(4-6) - 1(3+4) \\
&= -85 + 0 - 7 \\
&= -92
\end{aligned}$$

and

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & -3 \\ 5 & 0 & 1 \\ -2 & 3 & 4 \end{vmatrix} &= (-1)^{1+3}(-3)\begin{vmatrix} 5 & 0 \\ -2 & 3 \end{vmatrix} + (-1)^{2+3}(1)\begin{vmatrix} 1 & 2 \\ -2 & 3 \end{vmatrix} + (-1)^{3+3}(4)\begin{vmatrix} 1 & 2 \\ 5 & 0 \end{vmatrix} \\
&= -3(15-0) - 1(3+4) + 4(0-10) \\
&= -45 - 7 - 40 \\
&= -92
\end{aligned}$$

It follows immediately from Theorem 3.1 that if a square matrix has a row or column of zeros, then its determinant equals zero. By expanding across the row or column of zeros, a determinant of zero is produced regardless of the minors since the minors are multiplied by zero.

Corollary 3.2. If A is an $n \times n$ matrix with a row or column of zeros, then $\det A = 0$.

Another corollary of Theorem 3.1 is that the determinant of a square matrix and its transpose must be equal for clearly expansion across the first row of a matrix yields the same result as expansion down the first column of its transpose.

Corollary 3.3. If A is an $n \times n$ matrix, then $\det A = \det A^T$.

The sign of each term in the expansion depends on the factor $(-1)^{i+j}$. These powers of -1 form a $\pm$ checkerboard

$$\begin{bmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{bmatrix}$$

that starts with $+\left((-1)^{1+1}\right)$ in the upper left corner and alternates. Expansion across the first row or down the first column begins with a $+$, but expansion across the second row or down the second column begins with a $-$. Be sure to keep the signs straight when expanding.

As we have seen, the $\times$ pattern works for finding the terms of determinants of 2×2 matrices

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

For 3×3 matrices a basket weave pattern works. Repeat the first two columns of A to the right of the matrix. Then form the terms of the expansion by multiplying the three entries that line up along the echelons (down-right and up-right). Multiply the down-right terms by $+1$, the up-right terms by -1 and add.

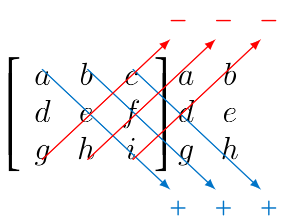


Figure 3.1 $det A = aei + bfg + cdh - gec - hfa - idb.$

These basket weave patterns do not generalize nicely to larger matrices.

Calculating a determinant by cofactor expansion can be a lot of work if the matrix is large. For many applications a 25×25 matrix is small, but even a very large computer that can calculate a trillion multiplications per second would require over 500,000 years to calculate that determinant by cofactor expansion. We need (and have) a faster way that involves elementary row operations. For 2×2 and 3×3 matrices, we don't bother with the elementary row operations because cofactor expansion, the × or the basket weave are at least as fast, but the elementary row operations help even with 4 × 4 matrices and help is more pronounced with larger matrices.

The first thing to note is that the determinant of a triangular matrix is particularly easy to calculate. When calculating the determinant of an upper triangular matrix by expanding down the first column we notice that it is not necessary to calculate most of the minors because they are multiplied by zero anyway. Example 3.4 illustrates.

Example 3.4

$$\begin{aligned}
\begin{vmatrix} 2 & 1 & 4 & 3 & 7 \\ 0 & 3 & -1 & 2 & 5 \\ 0 & 0 & 4 & 6 & 2 \\ 0 & 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 0 & -2 \end{vmatrix} &= (2)\begin{vmatrix} 3 & -1 & 2 & 5 \\ 0 & 4 & 6 & 2 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & -2 \end{vmatrix} - 0(*) + 0(*) - 0(*) + 0(*) \\
&= (2)(3)\begin{vmatrix} 4 & 6 & 2 \\ 0 & -1 & 4 \\ 0 & 0 & -2 \end{vmatrix} \\
&= (2)(3)(4)\begin{vmatrix} -1 & 4 \\ 0 & -2 \end{vmatrix} \\
&= (2)(3)(4)(-1)(-2) = 48
\end{aligned}$$

Though this is only an example, it is clear that this process works for any square upper triangular matrix. The determinant of an upper triangular matrix equals the product of the diagonal entries. This also works for square lower triangular matrices by expanding across the first row.

Theorem 3.4. If A is an $n \times n$ triangular matrix, then $\det A$ equals the product of the diagonal entries. That is, $\det A = a_{11}a_{22}\cdots a_{nn}$.

We have used elementary row operations to reduce a matrix into triangular forms. If we understood the effects of elementary row operations on the determinant, we could reduce the matrix to a triangular form and account for the effects of the elementary row operations as we go. Through this process we will be able to calculate determinants much more quickly than through cofactor expansion.

In the next section we study the effects of elementary row operations on the determinant.

Problem Set 3.1

1. Let $A = \begin{bmatrix} 1 & 8 & 9 \\ 7 & 0 & 4 \\ 2 & 6 & 3 \end{bmatrix}$.

(a) Find all 9 submatrices $A_{i,j}$.

(b) Find all 9 minors $M_{i,j}$.

(c) Find all 9 cofactors $C_{i,j}$.

2. Calculate the determinant of each of the following using the definition of determinant.

(a) $\begin{bmatrix} -7 \end{bmatrix}$

(b) $\begin{bmatrix} 7 & 8 \\ -3 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & -4 & -2 \\ 4 & -3 & 2 \\ 1 & 5 & 8 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & 0 & 5 \\ 0 & -2 & 4 & 0 \\ -1 & 0 & -3 & -2 \\ 5 & 2 & -3 & -1 \end{bmatrix}$

3. Use the basket-weave method to calculate the determinant of the following two matrices.

(a) $\begin{bmatrix} -3 & -2 & 2 \\ 3 & 2 & 1 \\ 1 & -1 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 3 & -2 & 0 \\ 2 & 4 & -3 \\ 5 & 3 & 1 \end{bmatrix}$

4. Verify that the determinants of the following matrix and its transpose are equal.

$$\begin{bmatrix} 1 & -2 & 0 \\ 2 & 3 & -3 \\ 5 & 4 & 1 \end{bmatrix}$$

5. Evaluate the determinants of the following matrices by inspection.

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

(b) $\begin{bmatrix} 3 & -2 & 0 \\ 2 & 4 & 0 \\ 5 & 3 & 0 \end{bmatrix}$

6. Use any of the theorems or corollaries of this section to calculate the determinants of the following. *None of these should be very much work if calculated correctly.*

(a) $\begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 4 & 4 & 2 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}$

(b) $\begin{bmatrix} 4 & 1 & 2 & 2 \\ 0 & -2 & -1 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} 4 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 \\ -2 & -1 & 5 & 0 \\ 0 & 3 & 0 & -2 \end{bmatrix}$

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

(e) $\begin{bmatrix} 1 & 1 & 0 & 5 & 0 & 3 \\ 0 & -2 & 4 & 0 & 0 & 2 \\ -1 & 0 & -3 & -2 & 0 & -4 \\ 5 & 2 & -3 & -1 & 0 & 8 \\ 3 & 6 & 5 & 2 & 0 & 8 \\ 3 & 2 & 8 & 5 & 0 & 3 \end{bmatrix}$ **(f)** $\begin{bmatrix} 1 & 1 & 0 & 5 & 0 & 3 \\ 0 & -2 & 4 & 0 & 3 & 0 \\ -1 & 0 & -3 & -2 & 0 & 0 \\ 5 & 2 & -3 & 0 & 0 & 0 \\ 3 & 6 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(g) $\begin{bmatrix} 0 & 3 & -2 & 1 & 3 & 5 \\ 0 & 4 & 0 & 0 & 0 & 5 \\ 0 & 2 & 0 & 6 & 0 & 3 \\ 0 & 1 & 5 & 7 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 1 & 3 & -2 & 3 & -3 & 2 \end{bmatrix}$

3.2 Elementary Row Operations and the Determinant

Below is a chart of the three elementary row operations with their effects on the determinant.

Elementary Row Operation	**Effect on the Determinant**
1. Multiply a row by a nonzero scalar r.	1. Changes the determinant by a factor of $r \neq 0$.
2. Swap two rows.	2. Switches the sign of the determinant.
3. Replace a row with itself plus a scalar multiple of another row.	3. Does not change the determinant.

We start the explanation with the first elementary row operation. Let A be an $n \times n$ matrix and let B be obtained from A by multiplying row i of A by a nonzero scalar r.

$$A = \begin{bmatrix} * \\ a_{i1} \cdots a_{in} \\ * \end{bmatrix}, \qquad B = \begin{bmatrix} * \\ ra_{i1} \cdots ra_{in} \\ * \end{bmatrix}$$

By expanding both determinants across row i we get

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \quad \text{and} \quad \det B = \sum_{j=1}^{n} (-1)^{i+j} b_{ij} \det B_{ij}.$$

But $b_{ij} = ra_{ij}$, and since A and B differ only on row i, for each j, $A_{ij} = B_{ij}$ because the row where A and B differ is deleted. So

$$\begin{aligned} \det B &= \sum_{j=1}^{n}(-1)^{i+j}(ra_{ij})\det A_{ij} \\ &= r\left(\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}\right) \\ &= r\det A. \end{aligned}$$

It follows that the first elementary row operation (multiplication by a scalar $r \neq 0$) changes the determinant by a factor of r.

For the second elementary row operation, we begin by assuming that rows to be swapped are *adjacent* to each other. Let A be the matrix before the swap and B the matrix after the swap. Assume rows i and $i+1$ are swapped.

$$A = \begin{bmatrix} * \\ a_{i1}\cdots a_{in} \\ a_{i+1,1}\cdots a_{i+1,n} \\ * \end{bmatrix}, \qquad B = \begin{bmatrix} * \\ a_{i+1,1}\cdots a_{i+1,n} \\ a_{i1}\cdots a_{in} \\ * \end{bmatrix}$$

Expansion across row i for $\det A$ yields

$$\det A = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}.$$

Expansion across row $i+1$ for $\det B$ yields

$$\det B = \sum_{j=1}^{n}(-1)^{i+1+j}b_{i+1,j}\det B_{i+1,j}.$$

Though rows i and $i+1$ are different rows, because of the swap the i^{th} row of A contains the same entries as the $i+1^{st}$ row of B and the i^{th} row of B contains the same entries as the $i+1^{st}$ row of A. We have $a_{ij} = b_{i+1,j}$ and the same entries are deleted when forming the minors $A_{i,j}$ and $B_{i+1,j}$. Also, the $i+1^{st}$ row of A and the i^{th} row of B (which are identical) move into the same positions in $A_{i,j}$ and $B_{i+1,j}$ so $B_{i+1,j} = A_{i,j}$. Therefore,

$$\begin{aligned} \det B &= \sum_{j=1}^{n}(-1)^{i+1+j}b_{i+1,j}\det B_{i+1,j} \\ &= \sum_{j=1}^{n}(-1)(-1)^{i+j}a_{ij}\det A_{i,j} \\ &= (-1)\left(\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{i,j}\right) \\ &= (-1)\det A \end{aligned}$$

So the second elementary row operation changes the sign of the determinant if the rows swapped are adjacent.

Next, suppose the rows to be swapped are *not* adjacent. Suppose B is obtained by swapping rows i and k with $k > i+1$. We can obtain B from A by doing several adjacent swaps. We must count how many adjacent swaps are done.

$$1\ 2\ \cdots\ i\ \cdots\ k\ \cdots\ n$$

Swap the entries in row i down 1 repeatedly until they get all the way down to row k. This is done in $k - i$ swaps. After that the entries originally in row i are down in row k and the entries in each row between i and k (including k) have moved up one row. To get the entries originally in row k up to row i we swap them with the row just above until they get up to row i. This requires one fewer swap since the swaps that took row i down left row k up one. So getting the row k entries up to row i requires $k - i - 1$ swaps. The rows above i and below k are unchanged. The rows i and k are swapped and the rows between i and k moves up one and then down one back to their original positions. The total number of swaps is $(k - i) + (k - i - 1) = 2k - 2i - 1 = 2(k - i) - 1$ which is an odd number. Each swap is an adjacent swap so each changes the sign of the determinant. Changing the sign an odd number of times leaves the sign the opposite of what it was originally. So, the second elementary row operation changes the sign of the determinant.

Before proceeding to the third elementary row operation, we stop to collect two corollaries of this result regarding the second elementary row operation.

Corollary 3.5. If A is a square matrix with two identical rows or columns, then $\det A = 0$.

Proof Suppose rows i and k of A are identical with $i \neq k$. Let $d = \det A$. Swapping rows i and k switches the sign of the determinant so after the switch the determiant is $-d$. But, because the rows are identical the matrix, hence its determiant, is unchanged by the swap so that $d = -d$. Since 0 is the only real number equal to its own negative, $d = 0$. Therefore, $\det A = 0$. ■

Corollary 3.6. If A is a square matrix in which one row is a scalar multiple of another row, then $\det A = 0$.

Proof Suppose A is an $n \times n$ matrix in which row k is a scalar, r, times row i of A.

$$\begin{aligned}
\det A &= \begin{vmatrix} * \\ a_{i1} \cdots a_{in} \\ * \\ r a_{i1} \cdots r a_{in} \\ * \end{vmatrix} \begin{array}{l} \\ \leftarrow \text{row } i \\ \\ \leftarrow \text{row } k \\ \\ \end{array} \\
&= r \begin{vmatrix} * \\ a_{i1} \cdots a_{in} \\ * \\ a_{i1} \cdots a_{in} \\ * \end{vmatrix} \begin{array}{l} \\ \leftarrow \text{row } i \\ \\ \leftarrow \text{row } k \\ \\ \end{array} \\
&= (r)(0) \quad \text{by Corollary 3.5} \\
&= 0
\end{aligned}$$

■

Now we move to the third elementary row operation. Suppose A is an $n\times n$ matrix and B is obtained from A by replacing row i with itself plus a scalar multiple r of row $k\neq i$.

$$A=\left[\begin{array}{c} * \\ a_{i1}\cdots a_{in} \\ * \\ a_{k1}\cdots a_{kn} \\ * \end{array}\right]\begin{array}{l} \\ \leftarrow \text{row } i \\ \\ \leftarrow \text{row } k \\ \\ \end{array}$$

$$B=\left[\begin{array}{c} * \\ a_{i1}+ra_{k1}\cdots a_{in}+ra_{kn} \\ * \\ a_{k1}\cdots a_{kn} \\ * \end{array}\right]\begin{array}{l} \\ \leftarrow \text{row } i \\ \\ \leftarrow \text{row } k \\ \\ \end{array}\qquad C=\left[\begin{array}{c} * \\ ra_{k1}\cdots ra_{kn} \\ * \\ a_{k1}\cdots a_{kn} \\ * \end{array}\right]\begin{array}{l} \\ \leftarrow \text{row } i \\ \\ \leftarrow \text{row } k \\ \\ \end{array}$$

Also, define C as shown above, just like A and B except in row i which is a multiple of row k, so $\det C=0$. Expanding $\det B$ across row i we get

$$\det B=\sum_{j=1}^{n}(-1)^{i+j}b_{ij}\det B_{ij}.$$

Since A,B, and C differ only in row i, for all j we have $A_{ij}=B_{ij}=C_{ij}$ and $b_{ij}=a_{ij}+ra_{kj}$, so

$$\begin{aligned}\det B &= \sum_{j=1}^{n}(-1)^{i+j}\left(a_{ij}+r_{ij}a_{kj}\right)\det A_{ij}\\ &= \sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}+\sum_{j=1}^{n}(-1)^{i+j}ra_{kj}\det A_{ij}\\ &= \det A+\det C\\ &= \det A+0\\ &= \det A\end{aligned}$$

Therefore, the third elementary row operation has no effect on the determinant.

A convenient way of describing the effects of all three elementary row operations on the determinant is as follows: If B is obtained from A by an elementary row operation, then $\det B=\alpha\det A$ where $\alpha\neq 0$ and $\alpha=r,-1,1$ depending on the type of elementary row operation.

Generally speaking, keep in mind that use of these properties is helpful in calculating the determinant only when the matrices are 4×4 or larger. We show two different ways in which these operations can be helpful in calculating a determinant. In Example 3.5, we use elementary row operations to reduce a matrix into an upper triangular form keeping track of the effect on the determinant as we go.

Example 3.5

$$\left|\begin{array}{cccc} 2 & 5 & 2 & -1 \\ 1 & 1 & 0 & -1 \\ -1 & 2 & 0 & 8 \\ 1 & 4 & 1 & 2 \end{array}\right| = -\left|\begin{array}{cccc} 1 & 1 & 0 & -1 \\ 2 & 5 & 2 & -1 \\ -1 & 2 & 0 & 8 \\ 1 & 4 & 1 & 2 \end{array}\right| = -\left|\begin{array}{cccc} 1 & 1 & 0 & -1 \\ 0 & 3 & 2 & 1 \\ 0 & 3 & 0 & 7 \\ 0 & 3 & 1 & 3 \end{array}\right| = -\left|\begin{array}{cccc} 1 & 1 & 0 & -1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & -2 & 6 \\ 0 & 0 & -1 & 2 \end{array}\right|$$

$$= -2\begin{vmatrix} 1 & 1 & 0 & -1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & -1 & 2 \end{vmatrix} = -2\begin{vmatrix} 1 & 1 & 0 & -1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & -1 \end{vmatrix} = (-2)(1)(3)(-1)(-1) = -6$$

Example 3.6 shows how to take advantage of naturally occurring zeros no matter where they occur to make the job of calculating the determinant easier.

Example 3.6

In this calculation, we take advantage of the 0's originally found in the third column and expand down that column after one elementary row operation.

$$\begin{vmatrix} 2 & 5 & 2 & -1 \\ 1 & 1 & 0 & -1 \\ -1 & 2 & 0 & 8 \\ 1 & 4 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 0 & -3 & 0 & -5 \\ 1 & 1 & 0 & -1 \\ -1 & 2 & 0 & 8 \\ 1 & 4 & 1 & 2 \end{vmatrix} = -(1)\begin{vmatrix} 0 & -3 & -5 \\ 1 & 1 & -1 \\ -1 & 2 & 8 \end{vmatrix} =$$

$$-(1)\begin{vmatrix} 0 & -3 & -5 \\ 1 & 1 & -1 \\ 0 & 3 & 7 \end{vmatrix} = (-1)(-1)\begin{vmatrix} -3 & -5 \\ 3 & 7 \end{vmatrix} = -21 + 15 = -6$$

Problem Set 3.2

1. Evaluate the determinants of the following matrices by inspection.

(a) $\begin{bmatrix} 3 & -2 & 2 \\ 3 & -2 & 2 \\ 1 & -1 & 3 \end{bmatrix}$ **(b)** $\begin{bmatrix} 3 & -2 & -6 \\ 2 & 4 & -4 \\ -1 & 3 & 2 \end{bmatrix}$

2. Evaluate the determinant of each of the following matrices by reducing each matrix to an upper-triangular form through elementary row operations.

(a) $\begin{bmatrix} 1 & -2 \\ 5 & 3 \end{bmatrix}$ **(b)** $\begin{bmatrix} 3 & 4 \\ 2 & 7 \end{bmatrix}$ **(c)** $\begin{bmatrix} 1 & 1 & 3 \\ 3 & 4 & 7 \\ 2 & 4 & 7 \end{bmatrix}$

(d) $\begin{bmatrix} 4 & 10 & -3 \\ 2 & 11 & 4 \\ 3 & 7 & 0 \end{bmatrix}$ **(e)** $\begin{bmatrix} 0 & 1 & 1 & 1 \\ 2/3 & 1/3 & 1 & 0 \\ 1/2 & 1/2 & 0 & 1 \\ 1/5 & 0 & 1/5 & 1 \end{bmatrix}$ **(f)** $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$

3. Given $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 7$, find

(a) $\begin{vmatrix} d & e & f \\ g & h & i \\ a & b & c \end{vmatrix}$ **(b)** $\begin{vmatrix} 3a & 3b & 3c \\ -d & -e & -f \\ 5g & 5h & 5i \end{vmatrix}$ **(c)** $\begin{vmatrix} a & b & c \\ a+2d & b+2e & c+2f \\ g-3a & h-3b & i-3c \end{vmatrix}$

4. Use elementary row operations to show
$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$

5. Prove that if A is an $n \times n$ matrix and k is a scalar, then $\det(kA) = k^n \det A$

3.3 Elementary Matrices and the Determinant

Now that we have learned the definition of determinant and how to calculate it efficiently, we develop two important properties of the determinant.

1. A square matrix is invertible if and only if its determinant is not zero.
2. The determinant of a product of matrices equals the product of the determinants.

We begin this process by examining the determinants of elementary matrices.

The $n \times n$ identity matrix, being triangular with its main diagonal all 1's, has a determinant of $\det I_n = 1$. Elementary matrices are obtained from the identity I_n by performing a single elementary row operation and in Section 3.2 we learned the effect of elementary row operations on the determinant. From this we see that if E is an elementary matrix, then $\det E = \alpha$ where

$$\alpha = \begin{cases} r \neq 0 & \text{if the row operation multiplies a row by } r \neq 0 \\ -1 & \text{if the row operation is a row swap} \\ 1 & \text{if the row operation is a replacement row operation} \end{cases}.$$

Lemma 3.7. Let B be an $n \times n$ matrix and E an $n \times n$ elementary matrix. Then

$$\det(EB) = (\det E)(\det B).$$

Proof Let C be the $n \times n$ matrix obtained from B by performing the same elementary row operation that transformed I_n to E. Let $\alpha = \det E$. From Section 3.2, we have

$\det C = \alpha \det B$. Recall from Section 1.6 that multiplication on the left by E has the same effect on B as performing the elementary row operation directly on B. It follows that $EB = C$ and $\det EB = \det C = \alpha \det B$. But $\alpha = \det E$ so that $\det EB = (\det E)(\det B)$. ■

Theorem 3.8. Let A be an $n \times n$ matrix. A is invertible if and only if $\det A \neq 0$.

Proof Let R be the reduced row-echelon form of A. There is a sequence of elementary row operations that reduces A to R: $A \longrightarrow \cdots \longrightarrow R$. Since performing an elementary row operation on a matrix has the same effect on the matrix as multiplication on the left by the corresponding elementary matrix, there is a sequence of elementary matrices $E_1, \cdots, E_k$ such that

$$R = E_1 \cdots E_k A.$$

Substituting for R and applying Lemma 3.7 k times, we peel the elementary matrices away one at a time.

$$\begin{aligned} \det R &= \det(E_1 \cdots E_k A) \\ &= (\det E_1)(\det E_2 \cdots E_k A) \\ &\vdots \\ &= (\det E_1)\cdots(\det E_k)(\det A) \end{aligned}$$

As mentioned above for each $i = 1, \cdots, k$, $\det E_i = 1, -1$, or r (all nonzero) so that $\det A = 0$ if and only if $\det R = 0$. If A is invertible, then $R = I_n$, so $\det R = 1 \neq 0$ making $\det A \neq 0$. If, on the other hand, A is singular, then R has a row of 0's and $\det R = 0$ making $\det A = 0$. Therefore, A is invertible if and only if $\det A \neq 0$. ■

We now add this result to Theorem 1.18 from Section 1.6 of statements equivalent to A is invertible.

Theorem 3.9. Let A be an $n \times n$ matrix. The following are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row-echelon form of A is the identity matrix I_n.

(d) A is a product of elementary matrices.

(e) A has n pivot columns.

(f) A has a left inverse.

(g) A has a right inverse.

(h) For all $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(i) Every n-vector $\mathbf{b}$ is a linear combination of the columns of A.

(j) A^T is invertible.

(k) $rank A = n$.

(l) $nullity A = 0$.

(m) $\det A \neq 0$.

Lemma 3.10. If A is an invertible $n \times n$ matrix and B is any $n \times n$ matrix, then

$$\det(AB) = (\det A)(\det B).$$

Proof Since A is invertible, there is a sequence of elementary matrices $E_1, \cdots, E_k$ such that $A = E_1 \cdots E_k$. Substituting for A and applying Lemma 3.7 k times we can again peel away the elemenatary matrices one at a time

$$\begin{aligned} \det(AB) &= \det(E_1 \cdots E_k B) \\ &= (\det E_1)(\det E_2 \cdots E_k B) \\ &\vdots \\ &= (\det E_1) \cdots (\det E_k)(\det B) \end{aligned}$$

Now we use Lemma 3.7 in the reverse direction $k-1$ times to put all the elementary matrices back together but leaving B out.

$$\begin{aligned} \det(AB) &= (\det E_1) \cdots (\det E_k)(\det B) \\ &= (\det E_1) \cdots (\det E_{k-2})(\det E_{k-1} E_k)(\det B) \\ &\vdots \\ &= (\det E_1 \cdots E_k)(\det B) \\ &= (\det A)(\det B) \end{aligned}$$

■

Theorem 3.11. If A and B are $n \times n$ matrices, then $\det(AB) = (\det A)(\det B)$.

Proof If A is invertible, then by Lemma 3.10 $\det(AB) = (\det A)(\det B)$. If A is singular, then by Corollary 1.17 from Section 1.6 the product AB is singular also so that both $\det A = 0$ and $\det(AB) = 0$. Therefore $\det AB = (\det A)(\det B)$. ■

Example 3.7

Let $A = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -3 \\ -2 & 0 \end{bmatrix}$. Then $AB = \begin{bmatrix} 0 & -6 \\ -7 & -9 \end{bmatrix}$ and $A + B = \begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix}$. Clearly $\det A = 7$, $\det B = -6$, $\det AB = -42$, and $\det(A+B) = 17$. Since $(7)(-6) = -42$, this illustrates the general theorem that the determinant of a product of matrices equals the product of their determinants. Is the determinant of a sum equal to the sum of the determinants? Since $7 + (-6) \neq 17$, these matrices also serve as a counterexample to disprove the *false* theorem $\det(A+B) = \det A + \det B$.

Though in general it is not true that $\det(A+B) = \det A + \det B$, there is a similar *true* theorem concerning determinants and sums. A special case of this theorem was used and proved in Section 3.2 to show that the replacement elementary row operation does not change the determinant. The proof of the next theorem is left to the exercises, but the key to the proof is found in Section 3.2.

Theorem 3.12. Let A, B, and C be $n \times n$ matrices that differ only in a single row. Suppose also that in the row where they differ, the entry in C equals the sum of the corresponding entries in A and B. Then,

$$\det C = \det A + \det B.$$

The same result holds for columns.

Example 3.8

For 3×3 matrices that are identical except in row 2, this theorem guarantees that

$$\begin{vmatrix} a & b & c \\ d+e & f+g & h+i \\ j & k & l \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & f & h \\ j & k & l \end{vmatrix} + \begin{vmatrix} a & b & c \\ e & g & i \\ j & k & l \end{vmatrix}.$$

Problem Set 3.3

1. Use determinants to determine whether the following matrices are invertible.

(a) $\begin{bmatrix} 4 & 3 \\ 5 & -2 \end{bmatrix}$ **(b)** $\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \\ 2 & 5 & 11 \end{bmatrix}$ **(c)** $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$

2. Let A and B be 3×3 matrices with $\det A = 5$ and $\det B = 4$. Find the following determinants.

(a) $\det AB$ **(b)** $\det 3A$ **(c)** $\det A^3$ **(d)** $\det AA^T$ **(e)** $\det B^{-1}$

3. Use determinants to find the values of x for which A is singular.

(a) $A = \begin{bmatrix} 3-x & 2 \\ 1 & 2-x \end{bmatrix}$ **(b)** $A = \begin{bmatrix} 1 & 3 & x \\ 4 & 1 & 2 \\ 5 & 2 & 1 \end{bmatrix}$

4. Express the following determinants as sums of determinants of matrices with entries that contain no sums.

(a) $\begin{vmatrix} a & b & c+d \\ e & f & g+h \\ i & j & k+l \end{vmatrix}$ **(b)** $\begin{vmatrix} a+b & c+d \\ e+f & g+h \end{vmatrix}$ **(c)** $\begin{vmatrix} a+b & c+d & e+f \\ g+h & i+j & k+l \\ m+n & p+q & r+s \end{vmatrix}$

5. Prove that for each positive integer n, $\det A^n = (\det A)^n$.

6. Prove that if A is invertible, then $\det A^{-1} = 1/\det A$.

7. Prove that for any $n \times n$ matrix A and for any invertible $n \times n$ matrix P, $\det(P^{-1}AP) = \det A$.

8. For $n \times n$ matrices A and B, you know that AB and BA need not be equal. Is the same true for $\det AB$ and $\det BA$? Explain.

9. Prove the following theorem: Let A, B, and C be $n \times n$ matrices that differ only in a single row. Suppose also that in that row where they differ, the entries in C equal the sum of the two corresponding entries from A and B. Then

$$\det C = \det A + \det B.$$

The same result holds for columns. *(Hint: Use LaPlace expansion by expanding* $\det C$ *across the row where the matrices differ.)*

3.4 Applications of the Determinant

Cramer's Rule

You have learned how to solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ by using Gauss-Jordan elimination and Gaussian elimination with back substitution. **Cramer's Rule** provides us with another method for solving square systems that uses determinants. Though helpful in some situations, Cramer's Rule has some distinct disadvantages of which you should be aware.

Disadvantages of Cramer's Rule

1. Cramer's Rule applies only to square systems in which the coefficient matrix is invertible.
2. The computational complexity of Cramer's Rule is greater than either of the above mentioned methods for solving a system. So it tends to be more work unless the system is a 2×2 system.
3. Cramer's Rule tends to have more problems with computational stability of its solutions than the other methods mentioned. That means that particularly with larger systems, if calculations involve roundoff error, the approximations you obtain through the other methods are likely to be better than those obtained via Cramer's Rule.

So why use it? Well, Cramer's Rule does have some advantages. The main advantage is that whereas the other methods mentioned are algorithms that bring us to the solution to a system, Cramer's Rule provides us with a *formula* for the solution. This is particularly helpful for theoretical results. Many statistics formulas, for example, are derived using Cramer's Rule.

It makes no practical sense to use Cramer's Rule for simply solving a system of size 3×3 or larger, but it works really well for 2×2 invertible systems. In fact, no matter how messy the fractions are in the solution, using Cramer's Rule yields quite simple calculations.

Theorem 3.13 (Cramer's Rule)**.** Given an $n \times n$ invertible system $A\mathbf{x} = \mathbf{b}$, for $j = 1, \cdots, n$, let A_j be the $n \times n$ matrix obtained from A by replacing column j of A with the column vector $\mathbf{b}$. For each $j = 1, \cdots, n$, in the solution $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ to the system,

$$x_j = \frac{\det A_j}{\det A}.$$

Before proving Cramer's Rule, we present Example 3.9.

Example 3.9

Solve

$$\begin{array}{rcrcr} 3x & + & 4y & = & 10 \\ 2x & + & 5y & = & 7 \end{array}$$

using Cramer's Rule. In this case $A = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix}$, $A_1 = \begin{bmatrix} 10 & 4 \\ 7 & 5 \end{bmatrix}$, and $A_2 = \begin{bmatrix} 3 & 10 \\ 2 & 7 \end{bmatrix}$. It follows that

$$x_1 = \frac{\begin{vmatrix} 10 & 4 \\ 7 & 5 \end{vmatrix}}{\begin{vmatrix} 3 & 4 \\ 2 & 5 \end{vmatrix}} = \frac{50 - 28}{15 - 8} = \frac{22}{7}, \quad \text{and} \quad x_2 = \frac{\begin{vmatrix} 3 & 10 \\ 2 & 7 \end{vmatrix}}{\begin{vmatrix} 3 & 4 \\ 2 & 5 \end{vmatrix}} = \frac{21 - 20}{15 - 8} = \frac{1}{7}.$$

The solution is then $\begin{bmatrix} 22/7 \\ 1/7 \end{bmatrix}$.

Proof Let $\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$ represent the solution to $A\mathbf{x} = \mathbf{b}$. Since A is square and invertible, the solution $\mathbf{s}$ is unique. Let S_j be the $n \times n$ matrix obtained from I_n by replacing column j of I_n with $\mathbf{s}$ for $j = 1, \cdots, n$. Note that $\det S_j = s_j$ for $j = 1, \cdots, n$ by expansion across row j. To solve $A\mathbf{x} = \mathbf{b}$ using Gauss-Jordan elimination we would apply elementary row operations to the augmented matrix $[A|\mathbf{b}]$ to place it into reduced row-echelon form. Since A is invertible, the reduced row-echelon form of $[A|\mathbf{b}]$ is $[I_n|\mathbf{s}]$.

$$[A|\mathbf{b}] \longrightarrow \cdots \longrightarrow [I_n|\mathbf{s}]$$

Clearly, applying the same sequence of elementary row operations to A and A_j yields

$$A \longrightarrow \cdots \longrightarrow I_n \quad \text{and} \quad A_j \longrightarrow \cdots \longrightarrow S_j.$$

Undoing the elementary row operations with other elementary row operations yields $[I_n|\mathbf{s}] \longrightarrow \cdots \longrightarrow [A|\mathbf{b}]$, $I_n \longrightarrow \cdots \longrightarrow A$, and $S_j \longrightarrow \cdots \longrightarrow A_j$. Applying an elementary row operation to a matrix has the same effect on the matrix as multiplying on the left by an elementary matrix, so there is a sequence of elementary matrices $E_1, \cdots, E_k$ such that $[A|\mathbf{b}] = E_1 \cdots E_k [I_n|\mathbf{s}]$, $A = E_1 \cdots E_k I_n$ and $A_j = E_1 \cdots E_k S_j$, so

$$\frac{\det A_j}{\det A} = \frac{\det(E_1 \cdots E_k S_j)}{\det(E_1 \cdots E_k I_n)} = \frac{(\det E_1 \cdots E_k)(\det S_j)}{(\det E_1 \cdots E_k)(\det I_n)} = \frac{s_j}{1} = s_j.$$

■

Interpretation of the Determinant as Area and Volume

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. In the xy-plane, we graph the vectors $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} c \\ d \end{bmatrix}$ and we wish to find the area of the parallelogram determined by these two vectors.

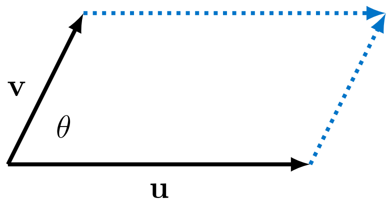


Figure 3.2 *A parallelogram determined by* $\mathbf{u}$ *and* $\mathbf{v}$. *It's area is given by* $\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta$.

In chapter 2, we then wrote this as the norm of a cross product. We run into a little problem here because the cross product is only for vectors in $\mathbb{R}^3$ and these are vectors in $\mathbb{R}^2$. We get around that problem by letting $\mathbf{u}' = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$ and $\mathbf{v}' = \begin{bmatrix} c \\ d \\ 0 \end{bmatrix}$. The area of the parallelogram determined by these two vectors is the same, so

$$\begin{aligned} Area &= \|\mathbf{u}\|\|\mathbf{v}\|\sin\theta \\ &= \|\mathbf{u}'\|\|\mathbf{v}'\|\sin\theta \\ &= \|\mathbf{u}' \times \mathbf{v}'\| \end{aligned}$$

But $\mathbf{u}' \times \mathbf{v}' = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a & b & 0 \\ c & d & 0 \end{vmatrix} = 0\mathbf{e}_1 - 0\mathbf{e}_2 + (ad - bc)\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ ad - bc \end{bmatrix}$. So,

$$\begin{aligned} Area &= \|\mathbf{u}' \times \mathbf{v}'\| \\ &= \sqrt{(ad - bc)^2} \\ &= |ad - bc| \\ &= |\det A|. \end{aligned}$$

This could just as well have been done with the column vectors of A instead of the row vectors of A since $\det A = \det A^T$. This gives us a nice geometric interpretation of the determinant of 2×2 matrices. First, 2×2 matrices are singular if their determinant is

0. Also, the area of a parallelogram being 0 means the parallelogram is flat (i.e. $\mathbf{u}$ and $\mathbf{v}$ are parallel vectors).

Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} d \\ e \\ f \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} g \\ h \\ i \end{bmatrix}$. We start by finding the volume of the parallelopiped determined by $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. In Section 2.3 we saw that

$$Volume = |(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}|.$$

Calculating gives

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a & b & c \\ d & e & f \end{vmatrix} \\ &= (bf - ce)\mathbf{e}_1 - (af - cd)\mathbf{e}_2 + (ae - bd)\mathbf{e}_3 \\ &= \begin{bmatrix} bf - ce \\ cd - af \\ ae - bd \end{bmatrix} \end{aligned}$$

so $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) = g(bf - ce) - h(af - cd) + i(ae - bd)$. But expansion of $\det A$ across the third row yields the same result. Thus,

$$\det A = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$$

and

$$Volume = |\det A|.$$

Here again we could have done this with column vectors of A as well as row vectors.

Singular matrices are the matrices that have a determinant of 0. So, for 3×3 matrices the parallelopiped determined by their column vectors are flat. That is, the three column vectors lie on the same plane through the origin.

Problem Set 3.4

1. Use Cramer's rule to solve the following systems.

(a) $\begin{aligned} 2x + 5y &= 4 \\ 3x - 4y &= 7 \end{aligned}$

(b) $\begin{aligned} 3x + 2y + 5z &= 7 \\ 2x + 5y + 3z &= 4 \\ 5x + 3y + 2z &= 6 \end{aligned}$

2. Use Cramer's rule to solve for x_2 without solving for x_1, x_3, and x_4 in the following system.

$$\begin{aligned} 3x_1 + x_2 + 2x_4 &= 1 \\ x_1 + x_2 + 4x_3 &= 2 \\ 2x_1 + x_3 + 3x_4 &= 1 \\ x_2 + 2x_3 + x_4 &= 0 \end{aligned}$$

3. Assuming $ad - bc \neq 0$, use Cramer's rule to find formulas for the values of x and y in terms of a, b, c, d, e, and f that satisfy the 2×2 system below.

$$\begin{array}{ccccc} ax & + & by & = & e \\ cx & + & dy & = & f \end{array}$$

4. Use determinants to find the areas of the parallelograms determined by the following pairs of vectors in $\mathbb{R}^2$.

(a) $\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}$ **(b)** $\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

5. Use the determinant to find the volume of the parallelepiped determined by the following three vectors in $\mathbb{R}^3$.

$$\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

6. Use the determinant to find the area of the triangle in the plane with vertices at the points $(2, 1)$, $(4, 4)$, and $(7, 3)$.

7. Use the determinant to find the volume of the tetrahedron with vertices at the points (1,2,1), (4,3,3), (2,5,1), and (3,2,4).

8. Suppose $A\mathbf{x} = \mathbf{b}$ is an $n \times n$ system of linear equations in which all the entries of A and $\mathbf{b}$ are integers and $\det A = \pm 1$. Prove that the solution to the system has all integer entries.

9. The law of cosines can be proved using Cramer's rule. Follow the outline below to prove the law of cosines.

(a) Given a triangle $\triangle ABC$ with angle measures α, β, γ at the vertices A, B, C and opposite sides of length a, b, c respectively. Using simple trigonometry, explain why $c \cos \beta + b \cos \gamma = a$.

(b) Using analogous reasoning, find similar expressions for b and c as found for a above.

(c) Letting $x = \cos \alpha$, $y = \cos \beta$, and $z = \cos \gamma$, write the three equations above as a 3×3 system of equations in x, y, and z.

(d) Use Cramer's rule to solve for z.

(e) In this last equation, substitute $\cos \gamma$ back for z.

(f) Solve for c. This should be the law of cosines.

4 Vector Spaces and Subspaces

4.1 Vector Spaces

Back in chapter 2, we defined for each $n \geq 1$,

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{R} \text{ for } i = 1, \cdots, n \right\}.$$

The set $\mathbb{R}^n$ (read "R-n") is the set of all column vectors with n entries. We have geometric interpretations for these sets for $n = 1, 2$, and 3. They are the real line, the coordinate plane, and three space respectively. We discussed the geometric interpretation of $\mathbb{R}^2$ and $\mathbb{R}^3$ in detail in chapter 2. For $n > 3$ we do not have enough spatial dimensions to visualize these vectors in the same way but space-time provides us with a physical interpretation of $\mathbb{R}^4$. Though visualizing these vectors geometrically may be impossible for $n \geq 4$, algebraically (vector addition, multiplication by scalars, linear combinations, systems of linear equations, elementary row operations, etc.) they are handled in pretty much the same way. Theorem 4.1 gives a list of several important properties of $\mathbb{R}^n$.

Theorem 4.1. Suppose $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ and $c, d \in \mathbb{R}$.

1. $\mathbf{u} + \mathbf{v} \in \mathbb{R}^n$

2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

4. There is a zero vector $\mathbf{0} \in \mathbb{R}^n$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$

5. There is a negative of $\mathbf{u}$ in $\mathbb{R}^n$, denoted $-\mathbf{u}$, such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

6. $c\mathbf{u} \in \mathbb{R}^n$

7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

9. $c(d\mathbf{u}) = (cd)\mathbf{u}$

10. $1\mathbf{u} = \mathbf{u}$

Though this list does not contain every single algebraic property of $\mathbb{R}^n$, it turns out that this list is sufficient to prove practically every other algebraic property of $\mathbb{R}^n$. That is, in order to prove most all other algebraic properties on $\mathbb{R}^n$, there is no need to go back to the definition of $\mathbb{R}^n$. One could simply go back to the properties on the list found in Theorem 4.1.

As an example, property 4 states that $\mathbb{R}^n$ has a zero vector (an additive identity). Property 4 does not state that $\mathbb{R}^n$ has only one additive identity. We prove that $\mathbb{R}^n$ has only one additive identity in Example 4.1 below, but notice that the proof makes no mention of the definition of $\mathbb{R}^n$. It uses only properties from the list of 10 given in Theorem 4.1. After this example, we discuss why that is important.

Example 4.1

Suppose $\mathbf{0}$ and $\mathbf{0}'$ behave as additive identities in $\mathbb{R}^n$. We show that $\mathbf{0} = \mathbf{0}'$. But

$$\begin{aligned} \mathbf{0} &= \mathbf{0} + \mathbf{0}' \longleftarrow \text{ since } \mathbf{0}' \text{ is an additive identity} \\ &= \mathbf{0}' + \mathbf{0} \longleftarrow \text{ by property 2} \\ &= \mathbf{0}' \longleftarrow \text{ since } \mathbf{0} \text{ is an additive identity} \end{aligned}$$

Why is it important to provide a proof that depends only on these 10 properties and not on the definition of $\mathbb{R}^n$? Because there are many other sets of mathematical objects that also satisfy those 10 properties. For example, the set of all polynomials, $\mathbb{P}$, satisfy those 10 properties also (the sum of two polynomials is another polynomial, polynomial addition is commutative and associative, etc. down the list). So the proof not only shows that $\mathbb{R}^n$ has only one additive identity, it also shows that there is only one polynomial additive identity (namely the constant zero polynomial $p(x) = 0$).

This example was chosen to be simple and easy to understand, and it may seem trivial and insignificant, but it illustrates something that is very important and profound about mathematics. The power of mathematics is in its abstraction. When abstracting we find the underlying similarities between very different settings. That is why we can use the same mathematics to answer questions in physics, economics, biology, and engineering.

In linear algebra, we abstract the idea of a vector. In various settings a vector may represent the wind, gravity, a continuous function, or a list of raw materials for factory production. Much of the mathematics is the same in each.

We define a vector space as a set of objects called vectors that satisfy these 10 properties. There are many different vector spaces. They can differ in many regards, but they are similar in many others. They are similar in that they all satisfy the 10 properties and all the properties (like uniqueness of additive identity) that follow from those 10.

We have a geometric notion of dimension. From that we think of $\mathbb{R}^1$ (the real line) as one dimensional, $\mathbb{R}^2$ as two dimensional, and $\mathbb{R}^3$ as three dimensional. In this chapter we formalize the notion of dimension for all vector spaces. We see, as expected, that $\mathbb{R}^n$ is n dimensional but also that some vector spaces, like the set of all polynomials, are infinite dimensional.

Abstract mathematics can be very powerful, but its abstractions can be difficult. To keep things easier, we focus most of our attention on $\mathbb{R}^n$. One needs to be aware of the existence of these other vector spaces, but we will not study them in as much detail here as we study $\mathbb{R}^n$. Theorems that apply to all vector spaces are presented and proved that way, but virtually all the examples we consider are from $\mathbb{R}^n$.

Definition 4.1 (Vector Space)**.** A (real) **vector space** is a nonempty set V of objects called vectors, on which are defined two operations called addition and multiplication by scalars (real numbers) subject to the 10 axioms listed below. The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in V$ and for all scalars c, d.

1. $\mathbf{u} + \mathbf{v} \in V$
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
4. There is a zero vector $\mathbf{0} \in V$ such that for all $\mathbf{u} \in V$, $\mathbf{u} + \mathbf{0} = \mathbf{u}$
5. For all $\mathbf{u} \in V$, there exists a vector $-\mathbf{u} \in V$, such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
6. $c\mathbf{u} \in V$
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$
10. $1\mathbf{u} = \mathbf{u}$

Example 4.2

The following is a list of several vector spaces. While other vector spaces exist, we study in detail only $\mathbb{R}^n$ for various n and their subspaces. This list is certainly not complete, but one for which the reader should be aware.

1. $\mathbb{R}^n$ for $n = 1, 2, 3, \cdots$

2. $\mathbb{P} =$ the set of all polynomials

3. $\mathbb{P}_n =$ the set of all polynomials of degree n or less for $n = 1, 2, 3, \cdots$

4. $C =$ the set of continuous functions over $\mathbb{R}$

5. $D =$ the set of all differentiable functions on $\mathbb{R}$

6. $C[a, b] =$ the set of all continuous functions on the interval $[a, b]$

7. $D[a, b] =$ the set of all functions continous on $[a, b]$ and differentiable on (a, b)

8. Any plane through the origin in $\mathbb{R}^3$

9. Any line through the origin in $\mathbb{R}^3$

10. Any line through the origin in $\mathbb{R}^2$

11. For each $m, n \geq 1$, $M_{mn} =$ the set of all $m \times n$ matrices.

Theorem 4.2 lists six additional properties that all vector spaces share. They are not included in the definition of vector space, but they follow logically from the 10 properties in the definition.

Theorem 4.2. Let V be a vector space, $\mathbf{u} \in V$, and $c \in \mathbb{R}$.

1. The zero vector in V is unique.

2. For each $\mathbf{u} \in V$, its negative is unique.

3. $0\mathbf{u} = \mathbf{0}$.

4. $c\mathbf{0} = \mathbf{0}$.

5. $(-1)\mathbf{u} = -\mathbf{u}$.

6. If $c\mathbf{u} = \mathbf{0}$, then $c = 0$ or $\mathbf{u} = \mathbf{0}$.

Proof

1. See Example 4.1.

2. Suppose $\mathbf{v}$ and $\mathbf{w}$ are negatives of $\mathbf{u}$. We show $\mathbf{v} = \mathbf{w}$. But $\mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{v} + (\mathbf{u} + \mathbf{w}) = (\mathbf{v} + \mathbf{u}) + \mathbf{w} = (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{0} + \mathbf{w} = \mathbf{w} + \mathbf{0} = \mathbf{w}$.

3. $0\mathbf{u} = 0\mathbf{u} + \mathbf{0} = 0\mathbf{u} + [\mathbf{u} + (-\mathbf{u})] = (0\mathbf{u} + \mathbf{u}) + (-\mathbf{u}) = (0\mathbf{u} + 1\mathbf{u}) + (-\mathbf{u}) = (0 + 1)\mathbf{u} + (-\mathbf{u}) = 1\mathbf{u} + (-\mathbf{u}) = \mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

4. $c\mathbf{0} = c\mathbf{0} + \mathbf{0} = c\mathbf{0} + [c\mathbf{0} + (-c\mathbf{0})] = (c\mathbf{0} + c\mathbf{0}) + (-c\mathbf{0}) = c(\mathbf{0} + \mathbf{0}) + (-c\mathbf{0}) = c\mathbf{0} + (-c\mathbf{0}) = \mathbf{0}$.

5. $(-1)\mathbf{u} = (-1)\mathbf{u} + \mathbf{0} = (-1)\mathbf{u} + [\mathbf{u} + (-\mathbf{u})] = [(-1)\mathbf{u} + \mathbf{u}] + (-\mathbf{u}) = [(-1)\mathbf{u} + 1\mathbf{u}] + (-\mathbf{u}) = (-1 + 1)\mathbf{u} + (-\mathbf{u}) = 0\mathbf{u} + (-\mathbf{u}) = \mathbf{0} + (-\mathbf{u}) = -\mathbf{u}$.

6. Suppose $c\mathbf{u} = \mathbf{0}$ and $c \neq 0$. We show $\mathbf{u} = \mathbf{0}$. But $\mathbf{u} = 1\mathbf{u} = \left[\left(\frac{1}{c}\right) c\right] \mathbf{u} = \frac{1}{c}(c\mathbf{u}) = \frac{1}{c}\mathbf{0} = \mathbf{0}$.

$\blacksquare$

These proofs are tricky and a bit tedious. They do not represent the types of proofs students would be expected to develop in this course. Instead, go through these proofs and try to justify each equal sign with an appropriate vector space axiom, theorem, or property of $\mathbb{R}$. This will help you become familiar with the definition of vector space.

Definition 4.2. Let V be a vector space. A **subspace** of V is a subset of V that is itself a vector space.

Example 4.3

Many of the 11 examples listed in Example 4.2 are subspaces of other vector spaces. In fact a line through the origin and a plane through the origin in $\mathbb{R}^3$ are examples of subspaces of $\mathbb{R}^3$. If that line happened to fall on the plane then the line would be a subspace of the plane which is itself a subspace of $\mathbb{R}^3$.

Example 4.4

Since $\mathbb{P}_1 \subseteq \mathbb{P}_2 \subseteq \mathbb{P}_3 \subseteq \cdots \subseteq \mathbb{P} \subseteq D \subseteq C$ and all are vector spaces, each vector space in this chain is a subspace of vector spaces to its right.

To prove that a subset of a vector space is a subspace from the definition, we would have to prove that the subset satisfies all 10 properties in the definition of a vector space. That would be a lot of work. Fortunately, the fact that we are dealing with a subset of a known vector space allows us to reduce the list of properties we must check from 10 down to 3.

The reason why we can get by without checking many of the properties is because all subsets of a vector space satisfy some of the properties. For example, suppose V is a vector space and W is any old subset of V. Property 2 in the definition of a vector space states that vector addition is commutative. Since V is a vector space, we know that for all $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$. Now suppose $\mathbf{w}_1, \mathbf{w}_2 \in W$. To show vector addition is commutative in W we must show that $\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{w}_2 + \mathbf{w}_1$. But the fact that $W \subseteq V$ tells us that $\mathbf{w}_1, \mathbf{w}_2 \in V$. So $\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{w}_2 + \mathbf{w}_1$.

Properties like the commutative property that are passed on to all of the subsets of V are said to be **inherited** properties. In the definition of a vector space, properties 2,3,7,8,9, and 10 are inherited by all subsets of a vector space. That leaves only properties 1,4,5, and 6 to check.

Property 1 in the defintion of a vector space is called **closure of vector addition**. Suppose V is a vector space, $W \subseteq V$, and $\mathbf{w}_1, \mathbf{w}_2 \in W$. The fact that V is a vector space and $W \subseteq V$ tells us that $\mathbf{w}_1 + \mathbf{w}_2 \in V$, but to show that W is a subspace we need to show that W is closed under vector addition. So we need to show $\mathbf{w}_1 + \mathbf{w}_2 \in W$. That is why property 1 is not inherited by every subset of V. Properties 4,5, and 6 are similarly not inherited properties.

Theorem 4.3 (The Subspace Test)**.** Let V be a vector space and W a subset of V. A subset W of V is a subspace of V if and only if W satisfies these three properties.

(a) W is nonempty.

(b) W is closed under vector addition.

(c) W is closed under scalar multiplication.

Proofs involving the subspace test are indeed the types of proofs linear algebra students are expected to know how to do. You will see the subspace test used many times in the text and in class. You need to know it and how to use it. What follows is about how to show each of the three properties (a), (b), and (c).

(a) *W is nonempty.* You must show that W contains at least one element. How you do that depends on what information is known to you about W. It is often easy to show $\mathbf{0} \in W$.

(b) *W is closed under vector addition.* The standard way to show W is closed under vector addition is as follows: Suppose $\mathbf{w}_1, \mathbf{w}_2 \in W$ and show $\mathbf{w}_1 + \mathbf{w}_2 \in W$.

(c) *W is closed under scalar multiplication.* The standard way to show this type of closure is to suppose that $\mathbf{w} \in W$ and $c \in \mathbb{R}$. Then, show that $c\mathbf{w} \in W$.

Next, we prove the subspace test theorem. Because it is an if and only if theorem, we have two directions to prove.

Proof (The Subspace Test)
First, suppose V is a vector space and W is a subset of V that satisfies properties (a), (b), and (c). We show that W is a subspace of V. Since $W \subseteq V$, W inherits properties 2, 3, 7, 8, 9, and 10 in the definition of a vector space, so we need only show that properties 1, 4, 5, and 6 are satisfied. Property 1 and property (b) are the same, so W satisfies property 1. Property 6 and property (c) are the same, so W satisfies property 6. For property 4 note that by property (a), $W \neq \varnothing$ so there exists $\mathbf{w} \in W$. Then, by property (c), $0\mathbf{w} \in W$. But $0\mathbf{w} = \mathbf{0}$ by part 3 of Theorem 4.2 so $\mathbf{0} \in W$ and since $\mathbf{0}$ serves as the identity for all of V, it serves as the identity for all of W. So W has an additive identity and hence satisfies property 4. For property 5 assume $\mathbf{w} \in W$. We show that $-\mathbf{w} \in W$. By property (c), $(-1)\mathbf{w} \in W$. But $(-1)\mathbf{w} = -\mathbf{w}$ by part 5 of Theorem 4.2, so $-\mathbf{w} \in W$. So the negative of every element of W is in W and hence W satisfies property 5. Therefore, if W satisfies properties (a), (b), and (c), then W is a subspace of V.

For the other direction, suppose that W is a subspace of a vector space V. We show that W is a subset of V that satisfies properties (a), (b), and (c). Now W a subspace of V implies that W satisfies properties 1-10 of the definition of a vector space. Clearly, we also see that property 4 implies property (a), property 1 implies property (b), and property 6 implies property (c). Therefore W is a subspace of V if and only if W satisfies (a), (b), and (c). ∎

Example 4.5

To see the subspace test at work in a concrete example, let W be the xy-plane in $\mathbb{R}^3$. The equation describing this plane is $z = 0$ and

$$W = \left\{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} : x, y \in \mathbb{R} \right\}.$$

We show W is a subspace of $\mathbb{R}^3$.

(a) Since $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in W$, $W \neq \varnothing$.

(b) Suppose $\mathbf{w}_1, \mathbf{w}_2 \in W$. We show that $\mathbf{w}_1 + \mathbf{w}_2 \in W$. But since $\mathbf{w}_1, \mathbf{w}_2 \in W$, there exist $x_1, y_1, x_2, y_2 \in \mathbb{R}$ such that $\mathbf{w}_1 = \begin{bmatrix} x_1 \\ y_1 \\ 0 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} x_2 \\ y_2 \\ 0 \end{bmatrix}$. So $\mathbf{w}_1 + \mathbf{w}_2 = \begin{bmatrix} x_1 \\ y_1 \\ 0 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ 0 \end{bmatrix} \in W$ since its z coordinate is 0.

(c) Suppose $\mathbf{w} \in W$ and $c \in \mathbb{R}$. We show $c\mathbf{w} \in W$. Here $\mathbf{w} \in W$ tells us that there exist $x, y \in \mathbb{R}$ such that $\mathbf{w} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$. So $c\mathbf{w} = c\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} cx \\ cy \\ 0 \end{bmatrix} \in W$ since its z coordinate is 0.

By the subspace test, W is a subspace of $\mathbb{R}^3$.

Problem Set 4.1

1. Consider the set that contains a single element which is the symbol $\triangle$. Is this set a vector space under the operations of vector addition defined by $\triangle + \triangle = \triangle$ and scalar multiplication defined by $k\triangle = \triangle$? Provide justification if your answer is no.

2. Consider the set that contains only two elements which are the symbols $\triangle$ and $\square$. Is the set a vector space under the operations of vector addition defined by $\triangle + \triangle = \triangle$, $\square + \square = \triangle$, and $\triangle + \square = \square + \triangle = \square$ and scalar multiplication defined by $k\triangle = \triangle$ and $k\square = \square$ for all real numbers k. Provide justification if your answer is no.

3. Provide the justification (i.e. axiom from Definition 4.1 or property of the real numbers, or given in the hypothesis of the theorem) for each equal sign in the proof of Theorem 4.2.

4. For parts (a) - (d) below, let W be the given subset of $\mathbb{R}^3$. Use Theorem 4.3 to determine whether each subset W is a subspace of $\mathbb{R}^3$. If W is a subspace of $\mathbb{R}^3$, use Theorem 4.3 to prove W is a subspace of $\mathbb{R}^3$. If W is not a subspace of $\mathbb{R}^3$, find an example to demonstrate that W does not satisfy one of the three properties in Theorem 4.3.

(a) $W = \left\{ \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} : a \in \mathbb{R} \right\}$

(b) $W = \left\{ \begin{bmatrix} a \\ b \\ 1 \end{bmatrix} : a, b \in \mathbb{R} \right\}$

(c) $W = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a + b + c = 0 \right\}$

(d) $W = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : c = a + b \right\}$

5. Use Theorem 4.3 to determine which of the following are subspaces of $M_{2,2}$.

(a) The set of all upper triangular 2×2 matrices .

(b) The set of all 2×2 matrices such that the sum of the entries equal 0.

(c) The set of all 2×2 matrices with integer entries.

(d) The set of all 2×2 matrices A such that $\det A = 0$.

6. Use Theorem 4.3 to determine which of the following are subspaces of $\mathbb{P}$, the vector space of all polynomials.

(a) The set of all polynomials of degree 3.

(b) The set of all polynomials of the form $p(x) = ax^3$ where a is any real number.

(c) The set of all polynomials $p(x) = a_n x^n + \cdots + a_1 x + a_0$ such that $a_3 = a_0$.

(d) The set of all polynomials, $p(x)$, with derivatives at $x = 3$ equal to 0 (i.e. $p'(3) = 0$).

7. Use Theorem 4.3 to determine which of the following are subspaces of D, the vector space of all differentiable functions on $(-\infty, \infty)$.

(a) The set of all differentiable functions, f, such that $f(0) = f(1)$.

(b) The set of all differentiable functions, f, such that $f'(x) = 3f(x)$.

(c) The set of all differentiable functions, f, such that $f'(3) = 0$.

(d) The set of all differentiable functions, f, such that $f'(0) = 1$.

8. Use the subspace test to prove that each of the following are subspaces.

(a) Let $W_1 = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 : x + 2y + 3z = 0 \right\}$. Prove W_1 is a subspace of $\mathbb{R}^3$.

(b) Let $D_2 = \{A \in M_{2,2} : A \text{ is a diagonal matrix}\}$. For the definition of diagonal matrix, see exercise 16 of problem set 1.5. Prove D_2 is a subspace of $M_{2,2}$.

(c) Let $Z_3 = \{p(x) \in \mathbb{P} : p(3) = 0\}$. Prove Z_3 is a subspace of $\mathbb{P}$.

(d) Let A be a fixed $n \times n$ matrix and let $E_4 = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 4\mathbf{x}\}$. Prove E_4 is a subspace of $\mathbb{R}^n$.

9. Suppose V is a vector space and W_1, and W_2 are subspaces of V. Determine whether the following three subsets of V must be subspaces of V. Use Theorem 4.3 to prove the subset is a subspace or give a counterexample to demonstrate the subset need not be a subspace.

(a) $W_1 \cap W_2$

(b) $W_1 \cup W_2$

(c) $W_1 + W_2 = \{\mathbf{w_1} + \mathbf{w_2} : \mathbf{w_1} \in W_1, \mathbf{w_2} \in W_2\}$

4.2 Subspaces

Theorem 4.4. Let V be a vector space. The sets V and $\{\mathbf{0}\}$ are subspaces of V.

Proof Since every set is a subset of itself, V is a subset of a vector space (itself) that is itself a vector space. So by the definition of subspace, V is a subspace of itself. To prove that $\{\mathbf{0}\}$ is a subspace of V, we use the subspace test.

(a) Since $\mathbf{0} \in \{\mathbf{0}\}$, $\{\mathbf{0}\} \neq \varnothing$.

(b) Since $\mathbf{0}$ is the only element of $\{\mathbf{0}\}$, the only possible vector sum in $\{\mathbf{0}\}$ is $\mathbf{0}+\mathbf{0}$. But $\mathbf{0}+\mathbf{0} = \mathbf{0} \in \{\mathbf{0}\}$, so $\{\mathbf{0}\}$ is closed under vector addition.

(c) Let $c \in \mathbb{R}$. Since $\{\mathbf{0}\}$ has $\mathbf{0}$ as its only element, any multiplication by scalars has the form $c\mathbf{0}$. But $c\mathbf{0} = \mathbf{0} \in \{\mathbf{0}\}$, so $\{\mathbf{0}\}$ is closed under multiplication by scalars. Therefore $\{\mathbf{0}\}$ is a subspace of V.

■

Definition 4.3. Let V be a vector space. The subspaces V and $\{\mathbf{0}\}$ are called the **trivial subspaces** of V.

What follows is an important way of generating a subspace of a vector space given any finite set of vectors.

Definition 4.4. Let $\mathcal{S} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a finite but nonempty set of vectors from a vector space V. Let the **span of** $\mathcal{S}$, denoted span $\mathcal{S}$, be the set of all possible linear combinations of the vectors in $\mathcal{S}$. That is,

$$span\ \mathcal{S} = \{c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n | c_1, \cdots, c_n \in \mathbb{R}\}.$$

We define span $\varnothing = \{\mathbf{0}\}$.

Theorem 4.5. If $\mathcal{S}$ is a finite set of vectors from a vector space V, then span $\mathcal{S}$ is a subspace of V.

Proof We use the subspace test. Suppose $\mathcal{S} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$.

(a) Since $\mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n \in span\ \mathcal{S}$, span $\mathcal{S} \neq \varnothing$.

(b) Let $\mathbf{w}_1, \mathbf{w}_2 \in span\ \mathcal{S}$. We show $\mathbf{w}_1 + \mathbf{w}_2 \in span\ \mathcal{S}$. Since $\mathbf{w}_1, \mathbf{w}_2 \in span\ \mathcal{S}$, there exist scalars $c_1, \cdots, c_n, d_1, \cdots, d_n$ such that $\mathbf{w}_1 = c_1\mathbf{v}_1 \cdots + c_n\mathbf{v}_n$ and $\mathbf{w}_2 = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$, so

$$\begin{aligned}\mathbf{w}_1 + \mathbf{w}_2 &= (c_1\mathbf{v}_1 \cdots + c_n\mathbf{v}_n) + (d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n) \\ &= (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_n + d_n)\mathbf{v}_n \in span\ \mathcal{S}.\end{aligned}$$

So $span\ \mathcal{S}$ is closed under vector addition.

(c) Let $\mathbf{w} \in span\ \mathcal{S}$ and $d \in \mathbb{R}$. We show that $d\mathbf{w} \in span\ \mathcal{S}$. Since $\mathbf{w} \in span\ \mathcal{S}$, there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. So

$$\begin{aligned}d\mathbf{w} &= d(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= (dc_1)\mathbf{v}_1 + \cdots + (dc_n)\mathbf{v}_n \in span\ \mathcal{S}.\end{aligned}$$

So $span\ \mathcal{S}$ is closed under multiplication by scalars. By the subspace test, $span\ \mathcal{S}$ is a subspace of V.

If $\mathcal{S} = \varnothing$, then by definition $span\ \mathcal{S} = \{\mathbf{0}\}$, the trivial subspace of V. In either case, $span\ \mathcal{S}$ is a subspace of V. ■

Example 4.6

Let $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} \right\}$ in $\mathbb{R}^3$. Because the two vectors in $\mathcal{S}$ are not parallel, we learned back in chapter 2 that the set of all linear combinations of those two vectors, i.e. $span\ \mathcal{S}$, is the plane through the origin and the two points $(1, 2, 1)$ and $(2, 3, 3)$. That plane can be described in a variety of ways. The set $\mathcal{S}$ itself provides us with a good description in that $span\ \mathcal{S}$ consists of all linear combinations of the form

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = s \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}.$$

With this we can crank out all kinds of vectors in $span\ \mathcal{S}$ by varying s and t. We call this an **explicit description** of $span\ \mathcal{S}$. This type of description is not so good, however, if you want to check to see whether a particular vector like $\begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ is in $span\ \mathcal{S}$ because that would involve solving a system of linear equations in each case. What would be more helpful in that case would be the equation of the plane, because checking to see whether they satisfy the equation would be a simple matter of plugging in and checking. The equation of the plane would be called an **implicit description** of $span\ \mathcal{S}$.

We present two methods for finding the equation of this plane. The first one is more what you learned in chapter 2 to focus on the geometrical aspects of this problem. The second method is more algebraic. This method may be more helpful when dealing with vectors in $\mathbb{R}^n$ where $n > 3$.

Solution 1 To find the equation of the plane $span\ \mathcal{S}$, we need a point on the plane and a vector normal to the plane. Since $\mathbf{0} \in span\ \mathcal{S}$, we use $\mathbf{0}$ as our point on the plane. To find a normal vector $\mathbf{n}$, we let

$$\mathbf{n} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 1 & 2 & 1 \\ 2 & 3 & 3 \end{vmatrix} = (6-3)\mathbf{e}_1 - (3-2)\mathbf{e}_2 + (3-4)\mathbf{e}_3 = \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}$$

so

$$\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) = 0$$

is a vector form of the equation. In standard form, we get $3x - y - z = 0$.

Solution 2 A vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is in *span* $\mathcal{S}$ if and only if there are scalars s and t such that

$$s \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

That is, $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in$ *span* $\mathcal{S}$ if and only if the system

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

is consistent. Forming the augmented matrix and reducing we see

$$\left[\begin{array}{cc|c} 1 & 2 & x \\ 2 & 3 & y \\ 1 & 3 & z \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 2 & x \\ 0 & -1 & y-2x \\ 0 & 0 & -3x+y+z \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 2 & x \\ 0 & 1 & 2x-y \\ 0 & 0 & -3x+y+z \end{array}\right].$$

At this point we see that the system is consistent if and only if $-3x + y + z = 0$. Thus the vectors $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ that are in *span* $\mathcal{S}$ are precisely those that satisfy $3x - y - z = 0$.

When talking about the span of $\mathcal{S}$, we are using the word "span" as a noun. It can also be used as a verb as the next definition shows.

Definition 4.5. Let V be a vector space, W a subspace of V, and $\mathcal{S}$ a finite set of vectors from W. We say $\mathcal{S}$ **spans** W or the vectors in $\mathcal{S}$ **span** W or W is **spanned by** $\mathcal{S}$ or W is **spanned by** the vectors in $\mathcal{S}$ if *span* $\mathcal{S} = W$.

Example 4.7 builds on Example 4.6 and illustrates an important point that is proved in Theorem 4.7 that follows.

Example 4.7

Is $\mathbb{R}^3$ spanned by $\left\{\begin{bmatrix}1\\2\\1\end{bmatrix},\begin{bmatrix}2\\3\\3\end{bmatrix},\begin{bmatrix}2\\5\\1\end{bmatrix}\right\}$?

Solution Notice the first two vectors come from the set $\mathcal{S}$ of Example 4.6. Now, if the answer to the question is yes, then the system

$$\begin{bmatrix}1&2&2\\2&3&5\\1&3&1\end{bmatrix}\begin{bmatrix}r\\s\\t\end{bmatrix}=\begin{bmatrix}x\\y\\z\end{bmatrix}$$

would be consistent for all x, y and z. Solving as before,

$$\left[\begin{array}{ccc|c}1&2&2&x\\2&3&5&y\\1&3&1&z\end{array}\right]\longrightarrow\left[\begin{array}{ccc|c}1&2&2&x\\0&-1&1&-2x+y\\0&1&-1&-x+z\end{array}\right]\longrightarrow\left[\begin{array}{ccc|c}1&2&2&x\\0&1&-1&2x-y\\0&0&0&-3x+y+z\end{array}\right].$$

Again, we see that this set of vectors spans the plane $3x-y-z=0$, so the answer is no, the set does not span $\mathbb{R}^3$. It might seem reasonable to expect the introduction of the third vector would expand the span, but that is not the case this time. Why not? We get a hint at why not by observing that the third vector $\begin{bmatrix}2\\5\\1\end{bmatrix}$ satisfies the equation $3x-y-z=0$, so it is in the span of the other two vectors. Since all three of these vectors lie on the same plane through the origin, all linear combinations of those three vectors lie on that plane also.

Lemma 4.6. If $\mathcal{S}$ is a finite set of vectors from a vector space V and $\mathcal{T}$ is a finite set of vectors from *span* $\mathcal{S}$, then *span* $\mathcal{T} \subseteq$ *span* $\mathcal{S}$.

Proof Exercise. ■

Theorem 4.7. Let $\mathcal{S}$ be a finite set of vectors from a vector space V. If a vector from $\mathcal{S}$ is a linear combination of the other vectors in $\mathcal{S}$, then removing that vector from $\mathcal{S}$ does not shrink *span* $\mathcal{S}$.

Proof Let $\mathcal{S}=\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$ be a finite set of vectors from a vector space V, and suppose a vector in $\mathcal{S}$ is a linear combination of the other vectors in $\mathcal{S}$. Without loss of generality, let $\mathbf{v}_n$ be this linear combination of the other vectors in $\mathcal{S}$. That means there exist scalars $c_1,\cdots,c_{n-1}$ such that $\mathbf{v}_n=c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}$.

To show that *span* $\mathcal{S}$ does not shrink by removing $\mathbf{v}_n$ from $\mathcal{S}$, let $\mathcal{T}=\{\mathbf{v}_1,\cdots,\mathbf{v}_{n-1}\}$. We show that *span* $\mathcal{S}$ = *span* $\mathcal{T}$. To accomplish this, we let $\mathbf{w}\in$ *span* $\mathcal{S}$ and we show that

$\mathbf{w} \in span\ \mathcal{T}$. Since $\mathbf{w} \in span\ \mathcal{S}$, there exist scalars $d_1, \cdots, d_n$ such that $\mathbf{w} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$. Substituting $c_1\mathbf{v}_1 + \cdots + c_{n-1}\mathbf{v}_{n-1}$ for $\mathbf{v}_n$ gives us

$$\begin{aligned} \mathbf{w} &= d_1\mathbf{v}_1 + \cdots + d_{n-1}\mathbf{v}_{n-1} + d_n(c_1\mathbf{v}_1 + \cdots + c_{n-1}\mathbf{v}_{n-1}) \\ &= (d_1 + d_n c_1)\mathbf{v}_1 + \cdots + (d_{n-1} + d_n c_{n-1})\mathbf{v}_{n-1} \in span\ \mathcal{T}. \end{aligned}$$

Thus $span\ \mathcal{S} \subseteq span\ \mathcal{T}$. Of course, it is clear that since $\mathcal{T} \subseteq \mathcal{S} \subseteq span\ \mathcal{S}$, we also have $span\ \mathcal{T} \subseteq span\ \mathcal{S}$ by Lemma 4.6. It follows that $span\ \mathcal{S} = span\ \mathcal{T}$. ■

For each matrix A we associate three different subspaces.

Definition 4.6. Let A be an $m \times n$ matrix.

1. The **column space** of A, denoted *col* A, is the span of the column vectors in A.

2. The **row space** of A, denoted *row* A, is the span of the row vectors of A.

3. The **null space** of A, denoted *null* A, is the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Theorem 4.8. Let A be an $m \times n$ matrix. The column space of A is a subspace of $\mathbb{R}^m$, and the row space and null space of A are subspaces of $\mathbb{R}^n$.

Proof Since A is an $m \times n$ matrix and because *col* A and *row* A are defined as spans of finite sets of vectors, *col* A is a subspace of $\mathbb{R}^m$ and *row* A is a subspace of $\mathbb{R}^n$. We show that *null* A is a subspace of $\mathbb{R}^n$ by using the subspace test.

(a) Since $A\mathbf{0} = \mathbf{0}$, $\mathbf{0} \in null\ A$ so $null\ A \neq \varnothing$.

(b) Suppose $\mathbf{u}, \mathbf{v} \in null\ A$. We show that $\mathbf{u} + \mathbf{v} \in null\ A$. But,

$$\begin{aligned} A(\mathbf{u} + \mathbf{v}) &= A\mathbf{u} + A\mathbf{v} \\ &= \mathbf{0} + \mathbf{0} \text{ since } \mathbf{u}, \mathbf{v} \in null A \\ &= \mathbf{0} \end{aligned}$$

so that $\mathbf{u} + \mathbf{v} \in null\ A$.

(c) Suppose $\mathbf{u} \in null\ A$ and c is a scalar. We show that $c\mathbf{u} \in null\ A$. But

$$\begin{aligned} A(c\mathbf{u}) &= c(A\mathbf{u}) \\ &= c\mathbf{0} \text{ since } \mathbf{u} \in null\ A \\ &= \mathbf{0} \end{aligned}$$

so $c\mathbf{u} \in null\ A$.

Therefore *null* A is a subspace of $\mathbb{R}^n$. ■

Example 4.8

Let

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

We would like to describe the column, row, and null spaces of A. To begin, we note that the matrix A itself provides us with explicit descriptions of *col* A and *row* A since we can generate all the elements of these two subspaces of $\mathbb{R}^3$ we want by taking different linear combinations of the columns and rows of A.

We might wonder whether either of these subspaces are $\mathbb{R}^3$ or if not whether we can find a way to describe them implicitly. We start with *col* A. Since *col* A is a span of a finite set, we use a method we learned earlier.

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & x \\ 1 & 1 & 2 & y \\ 0 & 1 & -1 & z \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & x \\ 0 & -1 & 1 & -x+y \\ 0 & 1 & -1 & z \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & x \\ 0 & 1 & -1 & x-y \\ 0 & 0 & 0 & -x+y+z \end{array}\right]$$

We see that *col* A is the plane defined by $x-y-z=0$ and we have our implicit description.

By ignoring the right-hand side of the matrices above, we see too that the third column of A is a linear combination of the first two. It follows that we can simplify our explicit description of *col* A from $span\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}\right\}$ to $span\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}\right\}$.

We will discuss the row space again later, but for now we note that the columns of A^T are the rows of A. So we can answer the same questions about *row* A by looking at *col* A^T.

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 2 & 1 & 1 & y \\ 1 & 2 & -1 & z \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & -1 & 1 & -2x+y \\ 0 & 1 & -1 & -x+z \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & 1 & -1 & 2x-y \\ 0 & 0 & 0 & -3x+y+z \end{array}\right]$$

So *row* A is the plane $3x-y-z=0$ in $\mathbb{R}^3$ and again the third vector can be thrown out to yield *row* $A = span\left\{\begin{bmatrix} 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}\right\}$.

For the null space we note that the matrix A really provides us with an implicit description because it is a simple matter to check whether a vector is in *null* A by checking to see whether it satisfies $A\mathbf{x} = \mathbf{0}$ by plugging it in for $\mathbf{x}$. To find an explicit description of *null* A we simply solve the system $A\mathbf{x} = \mathbf{0}$ as in chapter 1, but that is particularly easy because we did most of the work already when working on *col* A. We just need to replace x, y, and z with zeros.

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

If we let $z = t$, then $x = -3t, y = t$, and $z = t$ parametrically describes the solution to $A\mathbf{x} = \mathbf{0}$. We recognize this solution set as the parametric equations of a line through the origin. In vector form we have $\mathbf{x}(t) = t\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$. We get a simpler, implicit description of *null* A (simpler, at least, than the matrix equation $A\mathbf{x} = \mathbf{0}$) by writing symmetric equations for this line: $\frac{x}{-3} = y = z$.

Next, we describe the subspaces of $\mathbb{R}^3$ geometrically. We do this because they provide us with opportunities to practice the subspace test and to provide some geometric intuition of what a subspace is. The subspaces of $\mathbb{R}^2$ are developed in the exercises.

Subspaces of $\mathbb{R}^3$

The trivial subspaces of $\{\mathbf{0}\}$ and $\mathbb{R}^3$ are, of course, subspaces of $\mathbb{R}^3$.

Theorem 4.9. Planes through the origin of $\mathbb{R}^3$ are subspaces of $\mathbb{R}^3$.

Proof All planes in $\mathbb{R}^3$ can be described in point-normal form $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x}_0) = 0$ for some $\mathbf{n} \neq \mathbf{0}$. For planes through the origin we take $\mathbf{x}_0 = \mathbf{0}$. Let π be a plane through the origin in $\mathbb{R}^3$. Then, there exists a vector $\mathbf{n} \neq \mathbf{0}$ such that $\mathbf{n} \cdot \mathbf{x} = 0$ is satisfied by precisely the vectors in π. We show π is a subspace of $\mathbb{R}^3$ by using the subspace test.

(a) Since $\mathbf{n} \cdot \mathbf{0} = 0$, $\mathbf{0} \in \pi$ so $\pi \neq \varnothing$.

(b) Suppose $\mathbf{u}, \mathbf{v} \in \pi$. We show $\mathbf{u} + \mathbf{v} \in \pi$. But

$$\begin{aligned} \mathbf{n} \cdot (\mathbf{u} + \mathbf{v}) &= \mathbf{n} \cdot \mathbf{u} + \mathbf{n} \cdot \mathbf{v} \\ &= 0 + 0 \text{ since } \mathbf{u}, \mathbf{v} \in \pi \\ &= 0 \end{aligned}$$

So $\mathbf{u} + \mathbf{v} \in \pi$.

(c) Suppose $\mathbf{u} \in \pi$ and c is a scalar. We show that $c\mathbf{u} \in \pi$. But,

$$\begin{aligned} \mathbf{n} \cdot (c\mathbf{u}) &= c(\mathbf{n} \cdot \mathbf{u}) \\ &= c(0) \text{ since } \mathbf{u} \in \pi \\ &= 0 \end{aligned}$$

So $c\mathbf{u} \in \pi$.

Therefore, π is a subspace of $\mathbb{R}^3$. ∎

This brings up the natural question: What about planes that do not pass through the origin?

The answer is no. This is easily understood algebraically because all vector spaces (hence all subspaces) must contain a zero vector. In $\mathbb{R}^3$ the zero vector is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. If the plane does not go through the origin, it does not contain $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Theorem 4.10. Lines through the origin of $\mathbb{R}^3$ are subspaces of $\mathbb{R}^3$.

Proof All lines in $\mathbb{R}^3$ can be written in point-parallel form $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$, where $\mathbf{v} \neq \mathbf{0}$. For lines through the origin we take $\mathbf{x}_0 = \mathbf{0}$. Let L be a line through the origin in $\mathbb{R}^3$. Then, there exists a vector $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{x}(t) = t\mathbf{v}$ describes precisely the vectors in L for various t. We show L is a subspace of $\mathbb{R}^3$ by using the subspace test.

(a) Since $\mathbf{x}(0) = 0\mathbf{v} = \mathbf{0}$, $\mathbf{0} \in L$ so $L \neq \varnothing$.

(b) Suppose $\mathbf{u}, \mathbf{w} \in L$. We show $\mathbf{u} + \mathbf{w} \in L$. Now $\mathbf{u}, \mathbf{w} \in L$ implies that there exist t_1 and t_2 such that $\mathbf{u} = t_1\mathbf{v}$ and $\mathbf{w} = t_2\mathbf{v}$. But then $\mathbf{u} + \mathbf{w} = t_1\mathbf{v} + t_2\mathbf{v} = (t_1 + t_2)\mathbf{v}$. So $\mathbf{u} + \mathbf{w} \in L$.

(c) Suppose $\mathbf{u} \in L$ and c is a scalar. We show $c\mathbf{u} \in L$. Now $\mathbf{u} \in L$ implies that there exists t_1 such that $\mathbf{u} = t_1\mathbf{v}$. But then $c\mathbf{u} = c(t_1\mathbf{v}) = (ct_1)\mathbf{v}$. So $c\mathbf{u} \in L$.

Therefore L is a subspace of $\mathbb{R}^3$. ■

Of course, lines that do not pass through the origin are not subspaces either. Table 4.1 summarizes all subspaces of $\mathbb{R}^3$.

$\{\mathbf{0}\}$	0-dimensional subspace
lines through $\mathbf{0}$	1-dimensional subspaces
planes through $\mathbf{0}$	2-dimensional subspaces
$\mathbb{R}^3$	3-dimensional subspace

Table 4.1 *The subspaces of* $\mathbb{R}^3$.

Next, we show two ways to generate new subspaces out of old.

Definition 4.7. Suppose U and W are subspaces of a vector space V. Define the **sum** of U and W to be

$$U + W = \{\mathbf{u} + \mathbf{w} | \mathbf{u} \in U \text{ and } \mathbf{w} \in W\}.$$

Theorem 4.11. Suppose U and W are subspaces of a vector space V.

1. $U \cap W$ is a subspace of V.
2. $U + W$ is a subspace of V.

Proof

1. Exercise.
2. We use the subspace test.
   (a) Since $\mathbf{0} \in U$ and $\mathbf{0} \in W$, $\mathbf{0} + \mathbf{0} \in U + W$, so $U + W \neq \varnothing$.
   (b) Suppose $\mathbf{v}_1, \mathbf{v}_2 \in U + W$. We show $\mathbf{v}_1 + \mathbf{v}_2 \in U + W$. Now since $\mathbf{v}_1, \mathbf{v}_2 \in U + W$, there exist $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$ such that $\mathbf{v}_1 = \mathbf{u}_1 + \mathbf{w}_1$ and $\mathbf{v}_2 = \mathbf{u}_2 + \mathbf{w}_2$. So
   $$\mathbf{v}_1 + \mathbf{v}_2 = (\mathbf{u}_1 + \mathbf{w}_1) + (\mathbf{u}_2 + \mathbf{w}_2) = (\mathbf{u}_1 + \mathbf{u}_2) + (\mathbf{w}_1 + \mathbf{w}_2).$$
   But $\mathbf{u}_1 + \mathbf{u}_2 \in U$ and $\mathbf{w}_1 + \mathbf{w}_2 \in W$ since U and W are subspaces. So $\mathbf{v}_1 + \mathbf{v}_2 \in U + W$.
   (c) Suppose $\mathbf{v} \in U + W$ and c is a scalar. We show that $c\mathbf{v} \in U + W$. Now, since $\mathbf{v} \in U + W$, there exist $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. So $c\mathbf{v} = c(\mathbf{u} + \mathbf{w}) = c\mathbf{u} + c\mathbf{w}$. But $c\mathbf{u} \in U$ and $c\mathbf{w} \in W$ since U and W are subspaces. So $c\mathbf{v} \in U + W$.

   Therefore, $U + W$ is a subspace of V by the subspace test.

■

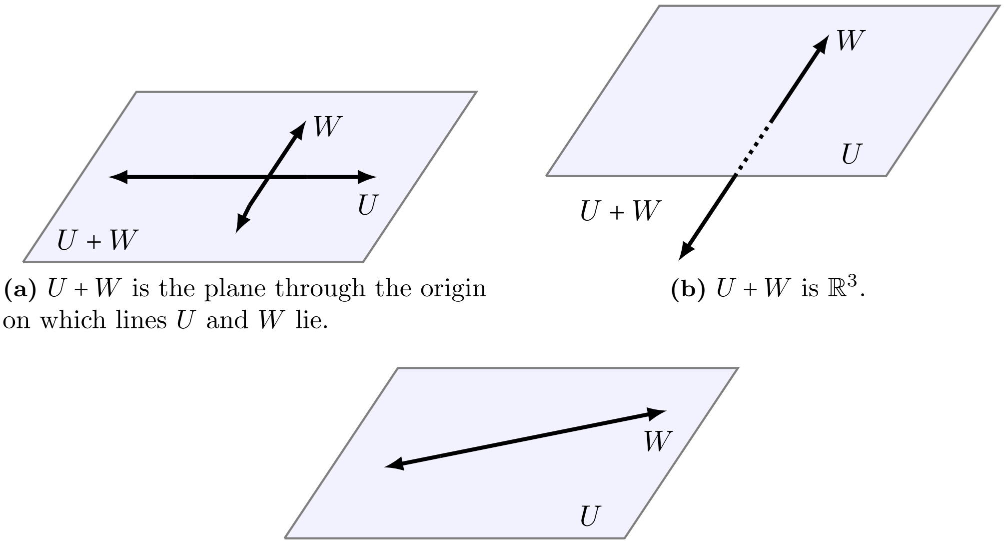
We now draw a few pictures to offer some geometric intuition about the intersection and sum of subspaces in $\mathbb{R}^3$.

If U and W are two distinct lines through the origin, then $U \cap W = \{\mathbf{0}\}$ and $U + W$ is the plane through the origin on which U and W lie (see Figure 4.1a).

If U is a plane through the origin and W is a line through the origin that is not on the plane, then $U \cap W = \{\mathbf{0}\}$ and $U + W = \mathbb{R}^3$ (see Figure 4.1b).

If U is a plane through the origin and W is a line through the origin that lies on U, then $U \cap W = W$ and $U + W = U$ (see Figure 4.1c).

We end this section by relating rank and nullity to the column space.



(a) $U+W$ is the plane through the origin on which lines U and W lie.

(b) $U+W$ is $\mathbb{R}^3$.

(c) $U \cap W = W$ and $U + W = U$.

Figure 4.1

Theorem 4.12. Let A be an $m \times n$ matrix. The following are equivalent.

(a) The set of column vectors of A span $\mathbb{R}^m$ ($col\ A = \mathbb{R}^m$).

(b) The system $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^m$.

(c) Every row of the reduced row-echelon form of A contains a leading 1.

(d) There are no zero rows in any row-echelon form of A.

(e) rank $A = m$

(f) nullity $A = n - m$

Problem Set 4.2

1. Find an equation of the plane that equals $\operatorname{span}\left\{\begin{bmatrix}1\\3\\1\end{bmatrix},\begin{bmatrix}2\\1\\1\end{bmatrix}\right\}$. In this problem you are given an explicit description of a subspace and are asked to find an implicit description of the same subspace.

2. The plane through the origin that satisfies the equation $3x+2y-z=0$ is a subspace of $\mathbb{R}^3$. Describe this subspace explicitly as the span of a two-vector set. In this problem you are given an implicit description of a subspace and asked for an explicit description of the same subspace.

3. Find symmetric equations for the line that equals $\operatorname{span}\left\{\begin{bmatrix}4\\3\\2\end{bmatrix}\right\}$. In this problem you are given an explicit description of a subspace and are asked to find an implicit description of the same subspace.

4. The line through the origin that equals the intersection of the two planes $x+3y+z=0$ and $x+4y+3z=0$ is a subspace of $\mathbb{R}^3$. Find an explicit description of this subspace.

5. Find an explicit description of the line $3x+2y=0$ in $\mathbb{R}^2$ through the origin .

6. Find an implicit description of the subspace $\operatorname{span}\left\{\begin{bmatrix}5\\3\end{bmatrix}\right\}$ in $\mathbb{R}^2$.

7. For each of the following matrices find the (i) column space, (ii) row space, and (iii) null space. For subspaces other than the trivial subspaces find both explicit and implicit descriptions. Use a minimum number of vectors in the spanning sets and a minimum number of equations in the implicit descriptions.

(a) $\begin{bmatrix}1&2&5\\1&3&8\\1&0&-1\end{bmatrix}$ **(b)** $\begin{bmatrix}1&2&3\\2&4&6\\4&8&12\end{bmatrix}$ **(c)** $\begin{bmatrix}1&2&6\\1&3&8\\2&3&9\end{bmatrix}$ **(d)** $\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix}$

(e) $\begin{bmatrix}1&2&3\\1&3&5\end{bmatrix}$ **(f)** $\begin{bmatrix}1&2\\2&4\\4&8\end{bmatrix}$ **(g)** $\begin{bmatrix}1&1&1&1\\1&2&3&2\\-1&0&2&2\\1&3&6&5\end{bmatrix}$

8. Suppose $\mathcal{S}$ is a finite set of vectors from a vector space V and $\mathcal{T}$ is a finite set of vectors from span $\mathcal{S}$. Prove span $\mathcal{T}\subseteq$ span $\mathcal{S}$.

9. Recall the dot product for vectors in $\mathbb{R}^n$ (definition and algebraic properties in section 2.2). For a fixed vector $\mathbf{u}\in\mathbb{R}^n$, let $W_{\mathbf{u}}=\{\mathbf{v}\in\mathbb{R}^n:\mathbf{u}\cdot\mathbf{v}=0\}$. Use the subspace test and algebraic properties of the dot product to prove that $W_{\mathbf{u}}$ is a subspace of $\mathbb{R}^n$.

10. Let C be the space of continuous functions on $(-\infty,\infty)$ and D the subset of C of all differentiable functions on $(-\infty,\infty)$. Use the subspace test along with well-known properties of the derivative from calculus to prove that D is a subspace of C.

11. Suppose U and W are subspaces of a vector space V. Use the subspace test to prove that $U \cap W$, the intersection of U and W, is a subspace of V.

4.3 Linear Dependence and Independence

Definition 4.8. A finite nonempty set of vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ in a vector space V is **linearly independent** if the homogeneous vector equation

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$$

has only the trivial solution $x_1 = \cdots = x_n = 0$. We define the empty set as **linearly independent** also. If a finite set of vectors is not linearly independent, we say the set is **linearly dependent**.

It is important to realize that every homogeneous vector equation has a trivial solution. The key to determining whether a set is linear independent is in determining whether the trivial solution is the *only* solution.

Theorem 4.13. Let A be an $m \times n$ matrix. The following are equivalent.

(a) The set of column vectors of A is linearly independent.

(b) The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) Every column of the reduced row-echelon form of A contains a leading 1.

(d) Every column of A is a pivot column.

(e) $rank\ A = n$

(f) $nullity\ A = 0$

Proof Let $\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ be the set of column vectors of A (i.e. $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$). We know that the vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$ and the matrix equation $A\mathbf{x} = \mathbf{0}$ have the same solutions, so the set of column vectors of A is linearly independent if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. But $A\mathbf{x} = \mathbf{0}$ has only the trivial solution precisely when each column of the reduced row-echelon form of A has a leading 1 which is what we mean when we say every column of A is a pivot column. It follows that (a), (b), (c), and (d) are equivalent. By definition, the rank of A equals the number of pivot columns of A, so $rank\ A = n$ is equivalent to saying every column of A is a pivot column. By definition, $nullity\ A = n - rank\ A$, so it is clear that $rank\ A = n$ if and only if *nullity* $A = 0$. Thus, (a), (b), (c), (d), (e), and (f) are all equivalent. ■

Example 4.9

Let

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -3 \\ 3 \end{bmatrix} \right\} \text{ and } \mathcal{T} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -2 \\ 3 \end{bmatrix} \right\}.$$

Determine whether $\mathcal{S}$ and $\mathcal{T}$ are linearly independent or dependent.

Solution For $\mathcal{S}$: $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 1 & 1 & -3 \\ -1 & -1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$

Since the third column is not a pivot column, $\mathcal{S}$ is linearly dependent.

For $\mathcal{T}$: $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 1 & 1 & -2 \\ -1 & -1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & -1 & -1 \\ 0 & 1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$

All three columns are pivot columns. $\mathcal{T}$ is linearly independent.

We leave the proof of the next two Theorems as exercises.

Theorem 4.14. Any finite set of vectors that contains the zero vector in a vector space V is linearly dependent.

Theorem 4.15. In a vector space V, any set consisting of only a single nonzero vector is linearly independent.

Many important theorems can be worded in terms of linear independence or in terms of linear dependence. Because it is often important to be aware of both formulations, we write them both ways. Since they are equivalent, we prove them only one way.

Theorem 4.16. A finite set of two or more vectors from a vector space V is linearly dependent if and only if at least one of the vectors is a linear combination of the others.

Theorem 4.16 (restated)**.** A finite set of two or more vectors from a vector space V is linearly independent if and only if none of the vectors are linear combinations of the others.

Proof Let $\mathcal{S} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a finite set of two or more vectors from a vector space V. Suppose at least one of the vectors in $\mathcal{S}$ is a linear combination of the others. Without loss of generality, suppose $\mathbf{v}_n$ is a linear combination of the others. Then, there exists scalars $c_1, \cdots, c_{n-1}$ such that $c_1\mathbf{v}_1 + \cdots + c_{n-1}\mathbf{v}_{n-1} = \mathbf{v}_n$. Bringing the vector $\mathbf{v}_n$ to the other side of the equation we see that

$$c_1\mathbf{v}_1 + \cdots + c_{n-1}\mathbf{v}_{n-1} + (-1)\mathbf{v}_n = \mathbf{0}.$$

So the vector equation $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$ has a nontrivial solution. Therefore $\mathcal{S}$ is linearly dependent.

To prove the other direction, suppose $\mathcal{S}$ is linearly dependent. This implies there exist scalars $c_1, \cdots, c_n$ not all zero such that $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$. Without loss of generality, suppose $c_n \neq 0$. Bringing everything else to the other side of the equation we get $c_n\mathbf{v}_n = -c_1\mathbf{v}_1 - \cdots - c_{n-1}\mathbf{v}_{n-1}$. Since $c_n \neq 0$ we can multiply both side by $\frac{1}{c_n}$ and obtain

$$\mathbf{v}_n = -\frac{c_1}{c_n}\mathbf{v}_1 - \cdots - \frac{c_{n-1}}{c_n}\mathbf{v}_{n-1}.$$

So, at least one vector is a linear combination of the others. ■

Theorem 4.16 provides us with a nice intuitive way of thinking about linear dependence and independence. For the most part, linear dependence means a vector is a linear combination of the others, and linear independence means none of the vectors are linear combinations of the others. Special consideration must be made for the empty set and sets with a single vector. A little more information can be gleaned from the last two theorems. Though not as intuitive, Corollary 4.17 gets at the same idea but with a wording that is helpful in Section 4.4.

Corollary 4.17. Let V be a vector space and $\mathcal{S} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a finite set of vectors from V. The set $\mathcal{S}$ is linearly dependent if and only if $\mathbf{v}_1 = \mathbf{0}$ or at least one of the vectors in $\mathcal{S}$ is a linear combination of its predecessors in the list $\mathbf{v}_1, \cdots, \mathbf{v}_n$.

Proof In light of Theorem 4.16, one direction of this proof is simple. We do that direction first.

Suppose $\mathbf{v}_1 = \mathbf{0}$ or one of the vectors in $\mathcal{S}$ is a linear combination of its predecessors. Under these assumptions, Theorems 4.14 and 4.16 imply that $\mathcal{S}$ is linearly dependent. Next, suppose $\mathcal{S}$ is linearly dependent. We show $\mathbf{v}_1 = \mathbf{0}$ or at least one of the vectors is a linear combination of its predecessors.

Let $\mathcal{S}_1 = \{\mathbf{v}_1\}$, $\mathcal{S}_2 = \{\mathbf{v}_1, \mathbf{v}_2\}$, $\cdots$, $\mathcal{S}_n = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = \mathcal{S}$. Since $\mathcal{S}_n$ is linearly dependent, there must be a first set, $\mathcal{S}_k$, in this list that is linearly dependent. If $k = 1$, then

$\mathbf{v}_1 = \mathbf{0}$ by Theorems 4.14 and 4.15. If $1 < k \le n$, then there exist scalars $c_1, \cdots, c_k$ not all zero such that $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$. If $c_k = 0$, then the vector equation simplifies to $c_1\mathbf{v}_1 + \cdots + c_{k-1}\mathbf{v}_{k-1} = \mathbf{0}$ and $c_1, \cdots, c_{k-1}$ are not all zero. That would imply $\mathcal{S}_{k-1}$ is linearly dependent contrary to the fact that $\mathcal{S}_k$ is the first. So $c_k \ne 0$. But then we can solve the equation for $\mathbf{v}_k$,

$$\mathbf{v}_k = -\frac{c_1}{c_k}\mathbf{v}_1 - \cdots - \frac{c_{k-1}}{c_k}\mathbf{v}_{k-1},$$

making $\mathbf{v}_k$ a linear combination of its predecessor. ■

Referring back to Example 4.9, since $\mathcal{S}$ is linearly dependent, Theorem 4.16 implies that at least one of the vectors in $\mathcal{S}$ is a linear combination of the others. By completing the row reduction to reduced row-echelon form, we see

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 1 & 1 & -3 \\ -1 & -1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

showing

$$\begin{bmatrix} -1 \\ 2 \\ -3 \\ 3 \end{bmatrix} = -5 \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \end{bmatrix}.$$

So the third vector is a linear combination of its predecessors. Corollary 4.17 tells us that regardless of which order these three column vectors are entered into the matrix one of the vectors will be a linear combination of its predecessors.

Next, we have two theorems. Each theorem has two wordings – one in terms of linear independence, the other linear dependence. The proof of the first pair is left as an exercise.

Theorem 4.18. Let $\mathcal{S}$ be a finite linearly independent set in a vector space V. Any subset of $\mathcal{S}$ is linearly independent.

Theorem 4.18 (restated)**.** Let $\mathcal{S}$ be a finite linearly dependent set in a vector space V that is not in $\mathcal{S}$. Any finite set of vectors in V that contains $\mathcal{S}$ is linearly dependent.

Theorem 4.19. Let $\mathcal{S}$ be a finite linearly independent set of vectors from a vector space V, and let $\mathbf{v} \in V$ that is not in $\mathcal{S}$. The set $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly dependent if and only if $\mathbf{v} \in \mathit{span}\ \mathcal{S}$.

Theorem 4.19 (restated)**.** Let $\mathcal{S}$ be a finite linearly independent set of vectors from a vector space V, and let $\mathbf{v} \in V$ that is not in $\mathcal{S}$. The set $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \notin \mathit{span}\ \mathcal{S}$.

Proof Suppose $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly dependent. List the vectors in $\mathcal{S} \cup \{\mathbf{v}\}$ with $\mathbf{v}$ last. By Corollary 4.17, at least one of the vectors in $\mathcal{S} \cup \{\mathbf{v}\}$ is a linear combination of its predecessors. Since $\mathcal{S}$ is linearly independent and the predecessors of the vectors in $\mathcal{S}$ are other vectors in $\mathcal{S}$, none of the vectors in $\mathcal{S}$ are linear combinations of their predecessors. Therefore $\mathbf{v} \in \mathit{span}\ \mathcal{S}$.

On the other hand, if $\mathbf{v} \in \mathit{span}\ \mathcal{S}$, then $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly dependent by Theorem 4.16. ■

Let's look at an example in $\mathbb{R}^3$ and discuss what is going on geometrically.

Example 4.10

Determine whether the following sets are linearly dependent or independent. Describe the spans of these sets.

$$\mathcal{S}_0 = \varnothing,\ \mathcal{S}_1 = \left\{\begin{bmatrix}1\\0\\2\end{bmatrix}\right\},\ \mathcal{S}_2 = \left\{\begin{bmatrix}1\\0\\2\end{bmatrix}, \begin{bmatrix}2\\-4\\1\end{bmatrix}\right\},\ \mathcal{S}_3 = \left\{\begin{bmatrix}1\\0\\2\end{bmatrix}, \begin{bmatrix}2\\-4\\1\end{bmatrix}, \begin{bmatrix}0\\4\\5\end{bmatrix}\right\},\ \mathcal{T}_0 = \left\{\begin{bmatrix}0\\0\\0\end{bmatrix}\right\},$$
$$\mathcal{T}_1 = \left\{\begin{bmatrix}1\\0\\2\end{bmatrix}, \begin{bmatrix}3\\0\\6\end{bmatrix}\right\},\ \mathcal{T}_2 = \left\{\begin{bmatrix}1\\0\\2\end{bmatrix}, \begin{bmatrix}2\\-4\\1\end{bmatrix}, \begin{bmatrix}0\\4\\3\end{bmatrix}\right\},\ \mathcal{T}_3 = \left\{\begin{bmatrix}1\\0\\2\end{bmatrix}, \begin{bmatrix}2\\-4\\1\end{bmatrix}, \begin{bmatrix}0\\4\\5\end{bmatrix}, \begin{bmatrix}3\\0\\-2\end{bmatrix}\right\}$$

Solution $\mathcal{S}_0$ is linearly independent by definition. Its span is the trivial subspace $\{\mathbf{0}\}$. $\mathcal{T}_0$ is linearly dependent by Theorem 4.14. Its span is the trivial subspace $\{\mathbf{0}\}$. $\mathcal{S}_1$ is linearly independent by Theorem 4.15. Its span is the line through the origin of all multiples of $\begin{bmatrix}1\\0\\2\end{bmatrix}$.

$\mathcal{T}_1$ contains two vectors and the second one is a multiple (linear combination) of the first. $\mathcal{T}_1$ is linearly dependent. Its span is the same as the span of $\mathcal{S}_1$, a line through the origin.

$\mathcal{S}_2$ contains two nonparallel nonzero vectors and is linearly independent. Its span is the plane through the origin on which the two vectors in $\mathcal{S}_2$ lie.

For $\mathcal{T}_2$, row reduction on the matrix having these column vectors is useful.

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

Since the third column is not a pivot column, the third vector is a linear combination of its predecessors, so it falls on the plane spanned by the first two making $\mathcal{T}_2$ linearly dependent. *Span* $\mathcal{T}_2$ is the same as *span* $\mathcal{S}_2$.

For $\mathcal{S}_3$, we again row reduce a matrix having these column vectors.

$$\begin{bmatrix}1&2&0\\0&-4&4\\2&1&5\end{bmatrix} \longrightarrow \begin{bmatrix}1&2&0\\0&1&-1\\0&-3&5\end{bmatrix} \longrightarrow \begin{bmatrix}1&2&0\\0&1&-1\\0&0&2\end{bmatrix}$$

Because every column is a pivot column, $\mathcal{S}_3$ is linearly independent. Another way of seeing it is that $\mathcal{S}_2$ is linearly independent and the third vector is not in the plane spanned by the first two. Because a row echelon form of this matrix has no zero rows, it is clear that $\mathcal{S}_3$ spans all of $\mathbb{R}^3$.

$\mathcal{T}_3$ contains four vectors from $\mathbb{R}^3$. Let A be the 3×4 matrix with vectors in $\mathcal{T}_3$ as columns. It is clear that A cannot have four pivot columns since it has only three rows, so $\mathcal{T}_3$ is linearly dependent. In fact, since $\mathcal{S}_3$ spans $\mathbb{R}^3$, the fourth vector in $\mathcal{T}_3$ must be in the span of the other vectors in $\mathcal{T}_3$.

In this example, we see geometrically how the spatial dimension of the span grows from a point to a line to a plane to all of three space as we include vectors that are not in the span of the previous set in moving from $\mathcal{S}_0$ to $\mathcal{S}_1$ to $\mathcal{S}_2$ to $\mathcal{S}_3$. By including vectors from outside the span of the previously included vectors, these sets stayed linearly independent. On the other hand, when including vectors that are in the span of the previous sets (from $\mathcal{S}_0$ to $\mathcal{T}_0$, from $\mathcal{S}_1$ to $\mathcal{T}_1$, from $\mathcal{S}_2$ to $\mathcal{T}_2$, and from $\mathcal{S}_3$ to $\mathcal{T}_3$) the spans did not grow and linear independence was lost.

In Example 4.10 we see a strong connection between the linear algebraic notions of linear independence and span with the spatial dimensions of the subspaces. In section 4.4 we use linear independence and span to generalize the notion of dimension.

We have learned in this section that in order to determine whether a finite set of vectors in $\mathbb{R}^n$ is linearly independent we place the vectors into a matrix as columns and reduce. If we are working in a vector space different from $\mathbb{R}^n$ like some function space, for example, or just an abstract vector space, then the vectors aren't columns of any matrix. It's not so clear how we would determine linear dependence or independence in these cases. We offer two examples to illustrate.

In Example 4.11 we see how even when we work in an abstract vector space, determining linear independence or dependence can still boil down to solving homogeneous matrix equations.

Example 4.11 **(1)** Suppose $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of vectors in a vector space V. Determine whether $\{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_3 + \mathbf{v}_1\}$ is linearly independent or dependent.

To do this, we must determine whether

$$x_1(\mathbf{v}_1 + \mathbf{v}_2) + x_2(\mathbf{v}_2 + \mathbf{v}_3) + x_3(\mathbf{v}_3 + \mathbf{v}_1) = \mathbf{0}$$

has nontrivial solutions. Rearranging the terms gives

$$(x_1 + x_3)\mathbf{v}_1 + (x_1 + x_2)\mathbf{v}_2 + (x_2 + x_3)\mathbf{v}_3 = \mathbf{0}.$$

Since this is a homogeneous vector equation involving the vectors of the linearly independent set $\mathcal{S}$ we know

$$\begin{array}{ccccccc} x_1 & & & + & x_3 & = & 0 \\ x_1 & + & x_2 & & & = & 0 \\ & & x_2 & + & x_3 & = & 0 \end{array}$$

Using row reduction to solve this system gives $x_1 = 0$, $x_2 = 0$, and $x_3 = 0$. Thus $\mathcal{S}$ is linearly independent.

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

(2) Similar to part (1), suppose $\mathcal{T} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent. Determine whether $\{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_3 + \mathbf{v}_4, \mathbf{v}_4 + \mathbf{v}_1\}$ is linearly independent or dependent.

As before, set

$$x_1(\mathbf{v}_1 + \mathbf{v}_2) + x_2(\mathbf{v}_2 + \mathbf{v}_3) + x_3(\mathbf{v}_3 + \mathbf{v}_4) + x_4(\mathbf{v}_4 + \mathbf{v}_1) = \mathbf{0}.$$

Rearrange terms to obtain

$$(x_1 + x_4)\mathbf{v}_1 + (x_1 + x_2)\mathbf{v}_2 + (x_2 + x_3)\mathbf{v}_3 + (x_3 + x_4)\mathbf{v}_4 = \mathbf{0}.$$

Since $\mathcal{T}$ is linearly independent, we get

$$\begin{array}{ccccccccc} x_1 & & & & & + & x_4 & = & 0 \\ x_1 & + & x_2 & & & & & = & 0 \\ & & x_2 & + & x_3 & & & = & 0 \\ & & & & x_3 & + & x_4 & = & 0 \end{array}$$

Row reducing this system yields

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

At this point we see that there are nontrivial solutions, so $\mathcal{T}$ is linearly dependent.

Example 4.12 deals with the vector space of continuous functions.

Example 4.12

Here we determine whether two given sets of functions are linearly dependent or linearly independent.

1. Let $\mathbf{f}(t) = \sin t$, $\mathbf{g}(t) = \cos t$, $\mathbf{h}(t) = 1$ (a constant function). Determine whether $\{\mathbf{f}, \mathbf{g}, \mathbf{h}\}$ is linearly dependent or independent.

   To accomplish this, we determine whether $x_1\mathbf{f} + x_2\mathbf{g} + x_3\mathbf{h} = \mathbf{0}$ has nontrivial solutions. Note that in this vector space, $\mathbf{0}$ is the constant function $\mathbf{z}(t) = 0$. To this end, we need to determine whether there are nontrivial values of x_1, x_2, x_3 such that

$$x_1 \sin t + x_2 \cos t + x_3(1) \quad = \quad 0 \tag{4.1}$$

for all t. Since this equation must be satisfied for all t, it must be satisfied in particular for $t = 0$, $t = \frac{\pi}{2}$, and $t = \pi$ giving us

$$\begin{array}{rcl} x_1 \sin 0 + x_2 \cos 0 + x_3(1) & = & 0 \\ x_1 \sin \dfrac{\pi}{2} + x_2 \cos \dfrac{\pi}{2} + x_3(1) & = & 0 \\ x_1 \sin \pi + x_2 \cos \pi + x_3(1) & = & 0 \end{array}$$

or

$$\begin{array}{ccccccc} & & x_2 & + & x_3 & = & 0 \\ x_1 & & & + & x_3 & = & 0 \\ & - & x_2 & + & x_3 & = & 0 \end{array} .$$

Solving this system yields only the trivial solution $x_1 = x_2 = x_3 = 0$.

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

Since no scalars except $x_1 = x_2 = x_3 = 0$ satisfy equation 4.1 for those three values of t simultaneously, there can be no other values of x_1, x_2, x_3 that satisfy the equation for all values of t. Therefore, $\{\mathbf{f}, \mathbf{g}, \mathbf{h}\}$ is linearly independent.

2. Let $\mathbf{f}(t) = \sin^2 t$, $\mathbf{g}(t) = \cos^2 t$, and $\mathbf{h}(t) = 1$. Determine whether $\{\mathbf{f}, \mathbf{g}, \mathbf{h}\}$ is linearly dependent or independent.

   Because of the identity $\cos^2 t + \sin^2 t = 1$, we see that the third vector in the set is a linear combination of the other two ($\mathbf{h} = 1\mathbf{f} + 1\mathbf{g}$). Thus this set is linearly dependent.

It is important to note that the technique used in part (a) worked for showing linear independence, but it would not work to show linear dependence. If you happened to find nontrivial values for the variables x_1, x_2, and x_3 that work for a select few values of t, perhaps you just haven't checked enough values of t to rule out the nontrivial solutions. You need to know some other information (like this identify from trigonometry) to show linear dependence.

We turn now to intersections and sums of subspaces to see how they relate to linear independence.

Theorem 4.20. Suppose U and W are subspaces of a vector space V with finite linearly independent sets $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_q\}$ in U and W respectively. If $U \cap W = \{\mathbf{0}\}$, then $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is linearly independent also.

Proof Suppose $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p + d_1\mathbf{w}_1 + \cdots + d_q\mathbf{w}_q = 0$. We show that $c_1 = \cdots = c_p = d_1 = \cdots = d_q = 0$. Bringing the $\mathbf{w}$'s to the other side of the equation gives

$$c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = -d_1\mathbf{w}_1 - \cdots - d_q\mathbf{w}_q.$$

But $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p \in U$ and $-d_1\mathbf{w}_1 - \cdots - d_q\mathbf{w}_q \in W$. Since these are equal, they fall in $U \cap W = \{\mathbf{0}\}$, so $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = \mathbf{0}$ and $-d_1\mathbf{w}_1 - \cdots - d_q\mathbf{w}_q = \mathbf{0}$. But since $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_q\}$ are linearly independent, $c_1 = \cdots = c_p = 0$ and $-d_1 = \cdots = -d_q = 0$. Therefore $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is linearly independent. ■

When $U \cap W = \{\mathbf{0}\}$, $U + W$ takes on some special qualities.

Definition 4.9. If U and W are subspaces of a vector space V and $U \cap W = \{\mathbf{0}\}$, then the sum of U and W is called the **direct sum** of U and W, and is denoted $U \oplus W$. A direct sum is a special case of a sum.

Theorem 4.21. If U and W are subspaces of a vector space V and $U \cap W = \{\mathbf{0}\}$, then each vector $\mathbf{v} \in U \oplus W$ has exactly one pair of vectors $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$.

Proof Since a direct sum is a sum, $\mathbf{v} \in U \oplus W$ implies that there exist $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. To show that this $\mathbf{u}$ and $\mathbf{w}$ are unique, we suppose $\mathbf{v} = \mathbf{u}' + \mathbf{w}'$ where $\mathbf{u}' \in U$ and $\mathbf{w}' \in W$ and we show $\mathbf{u}' = \mathbf{u}$ and $\mathbf{w}' = \mathbf{w}$. Now since $\mathbf{v} = \mathbf{u} + \mathbf{w}$ and $\mathbf{v} = \mathbf{u}' + \mathbf{w}'$, we have $\mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$. Taking the $\mathbf{u}$'s to one side of the equation and the $\mathbf{v}$'s to the other, we have $\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w} \in U \cap W = \{\mathbf{0}\}$. So $\mathbf{u} - \mathbf{u}' = \mathbf{0}$ and $\mathbf{w}' - \mathbf{w} = \mathbf{0}$ which shows that $\mathbf{u}' = \mathbf{u}$ and $\mathbf{w}' = \mathbf{w}$. ■

Problem Set 4.3

1. Determine which of the following sets of vectors from $\mathbb{R}^3$ are linearly dependent and which are linearly independent. Whenever possible write one of the vectors in the set as a linear combination of predecessors from the set.

(a) $\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\7\\2\end{bmatrix}, \begin{bmatrix}2\\6\\0\end{bmatrix}\right\}$ **(b)** $\left\{\begin{bmatrix}1\\1\\-1\end{bmatrix}, \begin{bmatrix}2\\3\\1\end{bmatrix}, \begin{bmatrix}1\\4\\3\end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix}2\\3\\1\end{bmatrix}, \begin{bmatrix}3\\0\\1\end{bmatrix}\right\}$ **(d)** $\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}3\\3\\6\end{bmatrix}, \begin{bmatrix}2\\3\\6\end{bmatrix}\right\}$

2. Determine which of the following sets of vectors from $\mathbb{R}^4$ are linearly dependent and which are linearly independent. Whenever possible write one of the vectors in the set as a linear combination of predecessors from the set.

(a) $\left\{\begin{bmatrix}1\\1\\-1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\0\\0\end{bmatrix}, \begin{bmatrix}2\\3\\0\\3\end{bmatrix}, \begin{bmatrix}2\\3\\4\\6\end{bmatrix}\right\}$ **(b)** $\left\{\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\2\\1\\3\end{bmatrix}, \begin{bmatrix}1\\3\\-1\\-2\end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix} 2 \\ 3 \\ 1 \\ 5 \end{bmatrix}\right\}$ **(d)** $\left\{\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \\ 0 \end{bmatrix}\right\}$

3. Suppose V is a vector space and $\mathcal{S} = \{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is a linearly independent set of vectors from V. Which of the following sets of vectors are linearly independent and which are linearly dependent in V. Use the definition of linear independence (Definition 4.8) and follow example 4.11 as a model to prove your claim.

(a) $\{\mathbf{v_1}, \mathbf{v_1} + \mathbf{v_2}, \mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3}\}$ **(b)** $\{\mathbf{v_1} - \mathbf{v_2}, \mathbf{v_2} - \mathbf{v_3}, \mathbf{v_3} - \mathbf{v_1}\}$

(c) $\{\mathbf{v_1} + 2\mathbf{v_2} + \mathbf{v_3}, 3\mathbf{v_1} + 5\mathbf{v_2} + 2\mathbf{v_3}, 2\mathbf{v_1} + 2\mathbf{v_2} + \mathbf{v_3}\}$

In general, it is difficult to determine linear dependence or independence of a finite set $\mathcal{S} = \{\mathbf{f_1}, \mathbf{f_2}, \ldots, \mathbf{f_n}\}$ of functions from the vector space $C^{n-1}(-\infty, \infty)$ of functions with $n-1$ continuous derivatives. It turns out that the determinant can be of assistance. This gives us another application of the determinant.

Let $\mathbf{f_i} = f_i(t)$ for $i = 1, \ldots, n$. The **Wronskian** of $\mathcal{S}$ is the determinant of the $n \times n$ matrix

$$\begin{bmatrix} f_1(t) & f_2(t) & \cdots & f_n(t) \\ f_1'(t) & f_2'(t) & \cdots & f_n'(t) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(t) & f_2^{(n-1)}(t) & \cdots & f_n^{(n-1)}(t) \end{bmatrix}.$$

The Wronskian is, of course, a function of t. We show that if the Wronskian is not uniformly equal to 0 for all values of t, then $\mathcal{S}$ is linearly independent.

Suppose $\mathcal{S}$ is linearly dependent. Then there exists scalars $c_1, c_2, \ldots, c_n$, not all 0, such that

$$c_1\mathbf{f_1} + c_2\mathbf{f_2} + \cdots + c_n\mathbf{f_n} = \mathbf{0}$$

That is, there are constants $c_1, \ldots, c_n$, not all 0, such that

$$c_1 f_1(t) + c_2 f_2(t) + \cdots + c_n f_n(t) = 0$$

for all values of t.

Differentiating this equation with respect to t yields

$$c_1 f_1'(t) + c_2 f_2'(t) + \cdots + c_n f_n'(t) = 0.$$

Differentiating again and again a total of $n-1$ times shows that $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ is a nontrivial solution to the homogeneous system

$$\begin{bmatrix} f_1(t) & f_2(t) & \cdots & f_n(t) \\ f_1'(t) & f_2'(t) & \cdots & f_n'(t) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(t) & f_2^{(n-1)}(t) & \cdots & f_n^{(n-1)}(t) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

for all values of t. But that implies the coefficient matrix is singular for all values of t, which means its determinant, the Wronskian, equals 0 for all values of t. Therefore, if the Wronskian is not uniformly 0 for all values of t, $\mathcal{S}$ must be linearly independent.

The converse of this theorem is false. That is, it is possible for $\mathcal{S}$ to be linearly independent even though its Wronskian is uniformly 0. The set $\mathcal{S} = \left\{t^2, t|t|\right\}$ serves as a counterexample of this converse. There are other theorems that add extra conditions beyond the Wronskian being uniformly 0 that do imply the set is linearly dependent.

4. Determine whether the following sets of functions are linearly independent or dependent.

(a) $\left\{1, e^t, e^{2t}\right\}$

(b) $\left\{6t^2 + 4t + 3, t^2 + t + 2, 2t + 9\right\}$

(c) $\{\sin 2t, \sin t \cos t\}$

(d) $\left\{t, e^t, \sin t\right\}$

5. Prove that any finite set of vectors from a vector space V that contains the zero vector is linearly dependent.

6. Prove that in a vector space V, any set that consists of a single nonzero vector is linearly independent.

7. Let $\mathcal{S}$ be a finite linearly independent set of vectors from a vector space V. Prove that any subset of $\mathcal{S}$ is linearly independent.

4.4 Basis and Dimension

In this section we define what we mean by the dimension of a vector space. This definition is consistent with our geometric understanding of dimension in $\mathbb{R}^2$ and $\mathbb{R}^3$, and it gives meaning to the term dimension for other vector spaces. We start by defining basis, which combines both span and linear independence.

Definition 4.10. Let V be a vector space and $\mathcal{B}$ a finite set of vectors from V. The set $\mathcal{B}$ is a **basis** for V if $\mathcal{B}$ spans V and is linearly independent.

Example 4.13

In the vector space $\mathbb{R}^n$, let $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ where the vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$ defined earlier are the standard basis vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$. The vector $\mathbf{e}_j$ is the j^{th} column of the $n \times n$ identity matrix I_n. It is clear that $\mathcal{S}_n$ spans $\mathbb{R}^n$ since $\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$, and $\mathcal{S}_n$ is linearly independent because every column of $[\mathbf{e}_1 \cdots \mathbf{e}_n] = I_n$ is a pivot column. The set $\mathcal{S}_n$ is called the **standard basis** for $\mathbb{R}^n$.

Since subspaces of vector spaces are themselves vector spaces, we can talk about bases of subspaces too.

Example 4.14

In a vector space V, the empty set, $\varnothing$, by definition is linearly independent and *span* $\varnothing = \{\mathbf{0}\}$, so the empty set is a basis for the trivial subspace $\{\mathbf{0}\}$.

Example 4.15

If $\mathbf{v}$ is a nonzero vector in $\mathbb{R}^2$ or $\mathbb{R}^3$, then the line through the origin $\mathbf{x}(t) = t\mathbf{v}$ is a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$ and $\mathcal{B} = \{\mathbf{v}\}$ is a basis for that subspace.

Example 4.16

The plane $x + 2y - z = 0$ passes through the origin. Therefore it is a subspace of $\mathbb{R}^3$. Find a basis for that subspace.

Solution By solving for x we can find an explicit description for the plane. Since $x = -2y + z$, if we let $y = s$ and $z = t$, then

$$\begin{array}{rcrcr} x & = & -2s & + & t \\ y & = & s & & \\ z & = & & & t \end{array}$$

or

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

Clearly $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ spans the plane and is linearly independent. Hence $\mathcal{B}$ is a basis for that plane. You can check to see that $\left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \right\}$ is also a basis for the same plane. In fact, any two nonparallel nonzero vectors that lie on that plane through the origin form a basis for that subspace of $\mathbb{R}^3$. This points out an important fact about bases. Every vector space and every subspace except the trivial subspace $\{\mathbf{0}\}$ have infinitely many different bases.

Though we focus on the vector spaces $\mathbb{R}^n$ for various n and their subspaces, we present one example of a basis for a different vector space so you are aware of their existence.

Example 4.17

Let $\mathbb{P}_3$ be the set of all polynomials with real coefficients of degree 3 or less. We show that the set $\mathcal{B} = \{1, t, t^2, t^3\}$ is a basis for $\mathbb{P}_3$.

It is clear that $\mathcal{B}$ spans $\mathbb{P}_3$ because an arbitrary polynomial of degree 3 or less has the form $a_3t^3 + a_2t^2 + a_1t + a_0$ and

$$a_3t^3 + a_2t^2 + a_1t + a_0 = a_0(1) + a_1(t) + a_2(t^2) + a_3(t^3).$$

To see that $\mathcal{B}$ is linearly independent we solve the homogeneous vector equation

$$x_1(1) + x_2(t) + x_3(t^2) + x_4(t^3) = \mathbf{0}.$$

We must realize that $\mathbf{0}$ is the constant function $z(t) = 0$, and we seek the values for x_1, x_2, x_3, x_4 that make this homogeneous vector equation true for all t. So, in particular, the homogeneous equation has to be true for $t = 0, 1, 2, 3$. Each of those values of t gives us a different linear equation in x_1, x_2, x_3, and x_4.

$$\begin{array}{lrcrcrcrcl}
t=0: & x_1 & & & & & & & = & 0 \\
t=1: & x_1 & + & x_2 & + & x_3 & + & x_4 & = & 0 \\
t=2: & x_1 & + & 2x_2 & + & 4x_3 & + & 8x_4 & = & 0 \\
t=3: & x_1 & + & 3x_2 & + & 9x_3 & + & 27x_4 & = & 0
\end{array}$$

Solving simultaneously we get

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 4 & 8 & 0 \\ 1 & 3 & 9 & 27 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 4 & 0 \\ 0 & 1 & 3 & 9 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

so that $x_1 = x_2 = x_3 = x_4 = 0$. Thus $\mathcal{B}$ spans $\mathbb{P}_3$ and is linearly independent (linear independence can also be shown using the Wronskian). Therefore $\mathcal{B}$ is a basis for $\mathbb{P}_3$.

Using a process similar to that in Example 4.17, it can be shown that two polynomials are equal as vectors, that is, equal for all values of t, if and only if they have the same degree and the coefficients of the same powers of t are equal. We use this fact in later examples involving polynomials.

Theorem 4.22. Suppose $\mathcal{S}$ is a finite set of vectors from a vector space V and $W = span\ \mathcal{S}$. There is a subset of $\mathcal{S}$ that is a basis for W.

Proof The process of removing one vector at a time from the finite set $\mathcal{S}$ eventually leads to a linearly independent set because if all the vectors were removed, the empty set, $\varnothing$, would be left which is linearly independent. If $\mathcal{S}$ is linearly independent, then by definition, $\mathcal{S}$ is a basis for W. If $\mathcal{S}$ is linearly dependent, then $\mathcal{S}$ contains a vector $\mathbf{v}$ that is in the span of $\mathcal{S}-\mathbf{v}$, and *span* $(\mathcal{S}-\mathbf{v}) =$ *span* $\mathcal{S}$. This process can continue without shrinking the span as long as the set of remaining vectors is linearly dependent. At that point the set of remaining vectors is linearly independent and spans W. Hence, it is a basis for W. ∎

Example 4.18

Let

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ -1 \\ 1 \end{bmatrix} \right\}$$

and $W =$ *span* $\mathcal{S}$. W is a subspace of $\mathbb{R}^4$. Find a basis for W.

Solution Let A be the matrix with the vectors of $\mathcal{S}$ as its columns. Reduce A.

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

At this point we see that the pivot columns of A are 1, 2, and 4. We claim that columns 1, 2, and 4 of A form a basis for W. To see what is happening, focus first on the first three columns. The same elementary row operations on the first three columns yield

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

Since the third column is not a pivot column, it is a linear combination of its predecessors. Thus it can be removed without shrinking the span.

Now focus on the remaining columns 1,2, 4, and 5.

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

This shows that the last column is not a pivot column. It is a linear combination of columns 1, 2, and 4 so it can be removed without shrinking the span. Finally, focus on columns 1, 2, and 4.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Since all three are pivot columns, they form a linearly independent set. They still span W so $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for W.

Theorem 4.22 and Example 4.18 show how to shrink a spanning set down to a basis. The next theorem and example show how to expand a linearly independent set to a basis.

Theorem 4.23. Suppose V is a vector space with a finite basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. If $\mathcal{S} = \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is a finite linearly independent set of vectors from V, then $\mathcal{S}$ can be expanded to a basis for V.

Proof Let $\mathcal{T} = \{\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{v}_1, \cdots, \mathbf{v}_n\}$. It is clear that $\mathcal{T}$ spans V since it has a subset $\mathcal{B}$ that does. If $\mathcal{T}$ is linearly independent, then $\mathcal{T}$ is a basis for V (this happens if $\mathcal{S} = \varnothing$). If $\mathcal{T}$ is linearly dependent, then at least one of the vectors in $\mathcal{T}$ is a linear combination of its predecessors. Removing it from $\mathcal{T}$ does not shrink the span. Since the predecessors in $\mathcal{T}$ of the vectors in $\mathcal{S}$ are all from $\mathcal{S}$ and since $\mathcal{S}$ is linearly independent, any vector that is a linear combination of its predecessors in $\mathcal{T}$ must be from $\mathcal{B}$. This process can be repeated without shrinking the span so long as the remaining subset of $\mathcal{T}$ is linearly dependent. Once it is no longer linearly dependent, it is linearly independent and spans V, hence a basis for V. Since no vectors in $\mathcal{S}$ are removed in this process, this basis contains $\mathcal{S}$. ■

Example 4.19

Let $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \\ 0 \end{bmatrix} \right\}$. This set, $\mathcal{S}$, is linearly independent as is demonstrated below. Expand $\mathcal{S}$ to a basis for $\mathbb{R}^4$.

Solution We form a matrix A with the vectors in $\mathcal{S}$ as the first columns augmented by I_4 because its columns form the standard basis for $\mathbb{R}^4$. The columns of A span $\mathbb{R}^4$

because the last four columns of A span $\mathbb{R}^4$.

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

At this point we see that the columns 1, 2, 3, and 5 are the pivot columns of A. The fact that columns 1, 2, and 3 are pivot columns verifies that $\mathcal{S}$ is indeed linearly independent as claimed. Since columns 4, 6, and 7 are not pivot columns, they are linear combinations of their predecessors that are pivot columns. Thus columns 1, 2, 3, and 5 form a basis for $\mathbb{R}^4$. So

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

expands $\mathcal{S}$ to a basis for $\mathbb{R}^4$.

Let A be an $m \times n$ matrix. We have discussed the column space of A that is a subspace of $\mathbb{R}^m$ and the row and null spaces of A that are subspaces of $\mathbb{R}^n$. Next we show how to find bases for these subspaces by reducing A.

Theorem 4.24. Let A be an $m \times n$ matrix. The pivot columns of A form a basis for the column space of A.

Proof By definition, the columns of A span the column space of A. We know that the non-pivot columns of A are linear combinations of their predecessor pivot columns, so they can be removed without shrinking the span making the pivot columns of A alone span the column space of A. In addition, the pivot columns of A form a linearly independent set, so they form a basis for *col* A. ■

Corollary 4.25. Let A be an $m \times n$ matrix. The pivot columns of A^T form a basis for the row space of A.

Proof Since the columns of A^T are the rows of A, the span of the columns of A^T equals the span of the rows of A. By Theorem 4.24, the pivot columns of A^T form a basis for the row space of A. ■

Though theoretically clear, the practical problem with Corollary 4.25 is that it requires reducing A^T rather than A. Finding bases for *col* A and *null* A can be done by reducing A. We could save work if we could find a basis for *row* A by reducing A. We take a short diversion to accomplish this.

Lemma 4.26. Let A be an $m \times n$ matrix, and P a matrix compatible with A so that the matrix product $B = PA$ is defined. Then $rowB \subseteq rowA$.

Proof From the end of section 1.5, we know that each row of B is a linear combination of the rows of A. By Lemma 4.6, $rowB \subseteq rowA$. ■

Theorem 4.27. If A is an $m \times n$ matrix and R a row-echelon form of A, then the nonzero rows of R form a basis for the row space of A.

Proof Since R is a row-echelon form of A, there is a sequence of elementary matrices $E_1, \cdots, E_k$ such that $R = E_k \cdots E_1 A$. Let $P = E_k \cdots E_1$. Lemma 4.26 shows $row\ R \subseteq row\ A$. But since elementary matrices are invertible, so is P. Thus $A = P^{-1}R$ so $row\ A \subseteq row\ R$ also and we have $row\ A = row\ R$. So the rows of R span $row\ A$. The zero rows of R add nothing to span A. It is also clear that the echelon form of R implies that the nonzero rows of R are linearly independent too. So the nonzero rows of R form a basis for $row\ A$. ■

Theorem 4.28. Let A be an $m \times n$ matrix. The general solution to $A\mathbf{x} = \mathbf{0}$ was studied in chapter 1 and written in vector form as $\mathbf{x} = t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k$ for some nonzero vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$. The set $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a basis for *null* A.

Proof Since $\mathbf{x} = t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k$ is the general solution for $A\mathbf{x} = \mathbf{0}$, the set $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ spans *null* A. Each parameter t_j results from a column of A that is not a pivot column and results in a 1 in an entry of $\mathbf{v}_i$ such that all other vectors in $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ have 0 in that entry. Each of those 1's with corresponding zeros imply that the vector equation $t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k = \mathbf{0}$ has only the trivial solution. Therefore $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a basis for *null* A. ■

Example 4.20

Let $A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 2 & 0 & 2 & 2 & -2 \\ 0 & 1 & -1 & 1 & -1 \\ 1 & 2 & -1 & 1 & 1 \end{bmatrix}$. Find bases for

(a) *col* A,

(b) *row* A, and

(c) *null* A.

Solution Start by putting A in reduced row-echelon form.

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

(a) Since columns 1, 2, and 4 are the pivot columns of A, they form a basis for $col A$ so $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for *col* A.

(b) Since the first three rows of the reduced row-echelon form of A are nonzero,

$$\left\{ \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & -1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 & -2 \end{bmatrix} \right\}$$

is a basis for *row* A.

(c) Solving $A\mathbf{x} = \mathbf{0}$ we let $x_3 = s$ and $x_5 = t$ to arrive at

$$\begin{array}{rcrcr} x_1 & = & -s & - & t \\ x_2 & = & s & - & t \\ x_3 & = & s & & \\ x_4 & = & & & 2t \\ x_5 & = & & & t \end{array}$$

which can be written in vector form as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

so

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}$$

forms a basis for *null* A. Note that because $x_3 = s$ and $x_5 = t$, the two vectors that generate the solutions to $A\mathbf{x} = \mathbf{0}$ have 1 and 0 in positions 3 and 5. This implies the two vectors are linearly independent.

Finding bases for *col* A, *row* A, and *null* A are all done differently, but they can all be done by reducing A to its reduced row-echelon form.

Here is another important property of bases.

Theorem 4.29. Let V be a vector space and $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ a basis for V. Every vector in V has a unique representation as a linear combination of vectors from $\mathcal{B}$ up to the order of the basis vectors.

Proof Since bases are spanning sets, *span* $\mathcal{B} = V$ so every vector in V has *at least one* representation as a linear combination of the vectors in $\mathcal{B}$. We show that because $\mathcal{B}$ is linearly independent, that every vector in V has *at most* one such representation. To that end, suppose $\mathbf{w} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ and $\mathbf{w} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$. We show $a_1 = b_1, a_2 = b_2, \cdots, a_n = b_n$. Now, since $\mathbf{w} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ and $\mathbf{w} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$, we have

$$a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n.$$

Combining like terms, we get

$$(a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n = \mathbf{0}.$$

Since $\mathcal{B}$ is linearly independent, $a_1 - b_1 = 0$, $a_2 - b_2 = 0$, $\cdots$, $a_n - b_n = 0$. Therefore, $a_1 = b_1$, $a_2 = b_2$, $\cdots$, $a_n = b_n$. ■

The next theorem is very important and the foundation of the notion of dimension of a vector space.

Theorem 4.30. Let V be a vector space with a finite basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. All bases of V contain the same number of elements.

Proof Suppose $\mathcal{C} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ is also a basis for V. We show $m = n$. Without loss of generality, suppose that $m \le n$. We present a process whereby we generate a finite sequence of sets $\mathcal{C}_0, \mathcal{C}_1, \cdots, \mathcal{C}_m$ in which $\mathcal{C}_0 = \mathcal{C}$ and as we move from $\mathcal{C}_i$ to $\mathcal{C}_{i+1}$ we add the next element from $\mathcal{B}$ and remove one of the remaining elements from $\mathcal{C}$. We do this in such a way that each $\mathcal{C}_i$ spans V. Once we reach $\mathcal{C}_m$ all the vectors from $\mathcal{C}$ have been removed.

Let $\mathcal{D}_1 = \{\mathbf{v}_1, \mathbf{u}_1, \cdots, \mathbf{u}_m\}$. This set is linearly dependent because $\mathbf{v}_1$ is a linear combination of the other vectors in $\mathcal{D}_1$. Since $\mathbf{v}_1 \ne \mathbf{0}$, one of the vectors in $\mathcal{D}_1$ is a linear combination of the predecessors in $\mathcal{D}_1$. That must be one of the vectors $\mathbf{u}_1, \cdots, \mathbf{u}_m$. Without loss of generality suppose it is $\mathbf{u}_1$ (renumber $\mathbf{u}_1, \cdots, \mathbf{u}_m$ if necessary). Since $\mathbf{u}_1$ is a linear combination of other vectors in $\mathcal{D}_1$, it can be removed without shrinking the span. Let $\mathcal{C}_1 = \{\mathbf{v}_1, \mathbf{u}_2, \mathbf{u}_3, \cdots, \mathbf{u}_m\}$. Then *span* $\mathcal{C}_1 = V$.

Suppose we have done this process i times $(1 \le i < m)$ with $\mathcal{C}_i = \{\mathbf{v}_1, \cdots, \mathbf{v}_i, \mathbf{u}_{i+1}, \cdots, \mathbf{u}_m\}$ and *span* $\mathcal{C}_i = V$. Let $\mathcal{D}_{i+1} = \{\mathbf{v}_1, \cdots, \mathbf{v}_i, \mathbf{v}_{i+1}, \mathbf{u}_{i+1}, \cdots, \mathbf{u}_m\}$. This set is linearly dependent since $\mathbf{v}_{i+1}$ is a linear combination of the others. Since we know $\mathbf{v}_1 \neq \mathbf{0}$, at least one of these vectors is a linear combination of its predecessors. If it is one of the vectors $\mathbf{v}_k$, that would imply $\mathcal{B}$ is linearly dependent since $\mathbf{v}_k$ and all of its predecessors come from $\mathcal{B}$. But that is impossible since $\mathcal{B}$ is a basis, so it must be $\mathbf{u}_j$ for some j where $i+1 \le j \le m$. Renumber if necessary so it is $\mathbf{u}_{i+1}$. Remove it to form $\mathcal{C}_{i+1} = \{\mathbf{v}_1, \cdots, \mathbf{v}_{i+1}, \mathbf{u}_{i+2}, \cdots, \mathbf{u}_m\}$ and span $\mathcal{C}_{i+1} = V$. This process stops with $\mathcal{C}_m = \{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ where *span* $\mathcal{C}_m = V$. If $m < n$, then $\mathbf{v}_n \in$ *span* $\mathcal{C}_m$ implying $\mathcal{B}$ is linearly dependent. Again, that is impossible, so $m = n$. ■

Definition 4.11. If V is a vector space and $\mathcal{B}$ is a basis for V containing n vectors, we say that V is an n-**dimensional** vector space or that the **dimension** of V is n.

Example 4.21

For each positive integer n, $\mathbb{R}^n$ is n dimensional because its standard basis $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ contains n vectors.

Example 4.22

In a vector space V, the trivial subspace, $\{\mathbf{0}\}$, is zero dimensional because its basis $\varnothing$ contains zero elements.

Example 4.23

Any line through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$ has a point-parallel form of $\mathbf{x}(t) = t\mathbf{v}$ where $\mathbf{v} \neq \mathbf{0}$. The set $\{\mathbf{v}\}$ is a basis for the subspace so lines through the origin are one-dimensional subspaces.

Example 4.24

Any plane through the origin in $\mathbb{R}^3$ can be written in the form $\mathbf{x}(s,t) = s\mathbf{u} + t\mathbf{v}$ where $\mathbf{u}, \mathbf{v}$ are nonzero and nonparallel. The set $\{\mathbf{u}, \mathbf{v}\}$ is a basis for this subspace of $\mathbb{R}^3$. All planes through the origin are two-dimensional subspaces of $\mathbb{R}^3$.

Example 4.25

The space of polynomials of degree n or less, $\mathbb{P}_n$, has $\{1, t, t^2, \cdots, t^n\}$ for a basis. The dimension of $\mathbb{P}_n$ is $n+1$.

Theorem 4.31. Let A be an $m \times n$ matrix.

(a) The dimension of the column space of A equals *rank* A.

(b) The dimension of the row space of A equals *rank* A.

(c) The dimension of the null space of A equals *nullity* $A = n - \textit{rank } A$.

(d) *rank* $A^T = \textit{rank } A$

Proof

(a) Because the pivot columns of A form a basis for *col* A, and the rank of A equals the number of pivot columns of A, the dimension of *col* A equals the rank of A.

(b) Because the nonzero rows of a row-echelon form of A form a basis for *row* A, and the number of nonzero rows in a row-echelon form of A equals the rank of A, the dimension of *row*A equals the rank of A.

(c) The homogeneous system $A\mathbf{x} = \mathbf{0}$ has a k-parameter family of solutions where k equals the number of columns of A that are not pivot columns. That is, $k = n - \textit{rank } A = \textit{nullity } A$. But if $A\mathbf{x} = \mathbf{0}$ has a k parameter family of solutions, we know that *null* A has a dimension of k. Therefore the dimension of *null* A equals *nullity* $A = n - \textit{rank } A$.

(d) Exercise.

■

Definition 4.12. Let V be a vector space. If there is a nonnegative integer n such that V is n-dimensional, then we say V is **finite dimensional**. If V is not finite dimensional, then we say V is **infinite dimensional**.

Though it is important to be aware of the existence of infinite dimensional vector spaces, they are not studied in detail in an introductory linear algebra class. We present one example.

Example 4.26

Let $\mathbb{P}$ be the set of all polynomials with real coefficients. $\mathbb{P}$ is an infinite dimensional vector space. We saw that the vector space $\mathbb{P}_n$ of polynomials of degree n or less has dimension $n+1$, since $\{1, t, t^2, \cdots, t^n\}$ is a basis for $\mathbb{P}_n$. But the vector space of all polynomials has no finite set that forms a basis, so $\mathbb{P}$ is infinite dimensional. Though we haven't defined what an infinite basis would be, the infinite set $\{1, t, t^2, \cdots\}$ is a basis for $\mathbb{P}$.

A basis for a finite dimensional vector space V is a finite set that is linearly independent and spans V. To show that a set is a basis by the definition, we show both linear independence and spanning. Knowing the dimension of the vector space can make that job easier.

Theorem 4.32. Let V be an n-dimensional vector space. Let $\mathcal{B}$ be a set of vectors from V. Any two of the three criteria listed below imply the third and that $\mathcal{B}$ is a basis for V.

(a) $\mathcal{B}$ is linearly independent.

(b) $\mathcal{B}$ spans V.

(c) $\mathcal{B}$ contains exactly n vectors.

Proof Suppose (a) and (b). By the definition of basis, $\mathcal{B}$ is a basis for V. Since V is n-dimensional, all bases of V contain exactly n vectors, so $\mathcal{B}$ contains exactly n vectors.

Suppose (a) and (c). Since $\mathcal{B}$ is linearly independent, it can be expanded to a basis for V. Since V is n-dimensional, that expansion must contain n vectors. Since $\mathcal{B}$ already contains n vectors, no extra vectors are necessary, so $\mathcal{B}$ is a basis and $\mathcal{B}$ spans V.

Suppose (b) and (c). Since $\mathcal{B}$ spans V, $\mathcal{B}$ contains a subset that is a basis for V. That basis must contain n vectors. Since $\mathcal{B}$ is the only subset of itself that contains n vectors, $\mathcal{B}$ is a basis, hence linearly independent. ■

Example 4.27

Suppose $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for a vector space V. Which of the following are bases for V?
$\mathcal{A} = \{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_2 + \mathbf{v}_3\}$
$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\}$
$\mathcal{C} = \{\mathbf{v}_3, \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3\}$
$\mathcal{D} = \{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 - \mathbf{v}_2, \mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_2 - \mathbf{v}_3\}$

Solution Since $\mathcal{S}$ has three vectors, V is three dimensional and all bases of V must contain three vectors. That means $\mathcal{A}$ and $\mathcal{D}$ are not bases for V. Since $\mathcal{B}$ and $\mathcal{C}$ contain the correct number of vectors, we need only check one additional criterion. Linear independence is easier.

First, we check $\mathcal{B}$. Now $x_1\mathbf{v}_1 + x_2(\mathbf{v}_1 + \mathbf{v}_2) + x_3(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) = \mathbf{0}$ if and only if $(x_1 + x_2 + x_3)\mathbf{v}_1 + (x_2 + x_3)\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$. Since $\mathcal{S}$ is a basis, we know

$$\begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 0 \\ & & x_2 & + & x_3 & = & 0 \\ & & & & x_3 & = & 0 \end{array}$$

which gives $x_1 = x_2 = x_3 = 0$ using back substitution. Therefore, $\mathcal{B}$ is a basis for V.

Next we check $\mathcal{C}$. Here $x_1\mathbf{v}_3 + x_2(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) + x_3(\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3) = \mathbf{0}$ if and only if $(x_2 + x_3)\mathbf{v}_1 + (x_2 + x_3)\mathbf{v}_2 + (x_1 + x_2 - x_3)\mathbf{v}_3 = \mathbf{0}$. Since $\mathcal{S}$ is a basis, we know

$$\begin{array}{ccccccc} & & x_2 & + & x_3 & = & 0 \\ & & x_2 & + & x_3 & = & 0 \\ x_1 & + & x_2 & - & x_3 & = & 0 \end{array}$$

Since nontrivial solutions like $x_1 = 2, x_2 = -1$, and $x_3 = 1$ exist, $\mathcal{C}$ is linearly dependent and not a basis (so it must not span V either).

We return now to direct sums.

Theorem 4.33. Suppose U and W are finite dimensional subspaces of a vector space V such that $U \cap W = \{\mathbf{0}\}$.

(a) If $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_q\}$ are bases of U and W respectively, then $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is a basis for $U \oplus W$.

(b) $\mathit{dim}(U \oplus W) = (\mathit{dim}\ U) + (\mathit{dim}\ W)$

Proof

(a) Since bases are linearly independent, $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is linearly independent by Theorem 4.20. To show it spans $U \oplus V$, suppose $\mathbf{v} \in U \oplus V$ and show $\mathbf{v}$ is a linear combination of $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$. Since $\mathbf{v} \in U \oplus V$, there exist $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. Since $\mathbf{u} \in U$ and $\mathbf{w} \in W$, there exist scalars $c_1, \cdots, c_p, d_1, \cdots, d_q$ such that $\mathbf{u} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ and $\mathbf{w} = d_1\mathbf{w}_1 + \cdots + d_q\mathbf{w}_q$ so $\mathbf{v} = \mathbf{u} + \mathbf{w} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p + d_1\mathbf{w}_1 + \cdots + d_q\mathbf{w}_q$. Thus $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ spans $U \oplus W$. Since $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is linearly independent and spans $U \oplus W$, it is a basis for $U \oplus W$.

(b) From part (a), $\mathit{dim}(U \oplus W) = p + q$, but $\mathit{dim}\ U = p$ and $\mathit{dim}\ W = q$, so $\mathit{dim}(U \oplus W) = (\mathit{dim}\ U) + (\mathit{dim}\ W)$

■

Theorem 4.34. Suppose U and W are subspaces of an n-dimensional vector space V. Any two of the following implies the third and that $V = U \oplus W$.

(a) $U \cap W = \{\mathbf{0}\}$

(b) $U + W = V$

(c) $(\mathit{dim}\ U) + (\mathit{dim}\ W) = n$

Proof ((a) and (b) $\Longrightarrow$ (c)) By the definition of direct sum, $V = U \oplus W$. Theorem 4.33 implies $(dim\ U) + (dim\ W) = n$.

((a) and (c) $\Longrightarrow$ (b)) Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_q\}$ be bases of U and W respectively. Then $dim\ U = p$ and $dim\ W = q$. By Theorem 4.33, $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is a basis for $U \oplus W$. So $U \oplus W$ is a subspace of V of dimension $p + q = dim\ U + dim\ W = n$. But V is the only subspace of V of dimension n, so $V = U \oplus W = U + W$.

((b) and (c) $\Longrightarrow$ (a)) Again, let $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_q\}$ be bases of U and W respectively. Since $U + W = V$, $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is a set of $p + q = n$ vectors that spans V, so it contains a basis of V as a subset. But all bases of V contain n vectors, so it must be a basis for V. To show $U \cap W = \{\mathbf{0}\}$, we suppose $\mathbf{v} \in U \cap W$ and show $\mathbf{v} = \mathbf{0}$. Since $\mathbf{v} \in U$ and $\mathbf{v} \in W$, there exist scalars $c_1, \cdots, c_p, d_1, \cdots, d_q$ such that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ and $\mathbf{v} = d_1\mathbf{w}_1 + \cdots + d_q\mathbf{w}_q$. So,

$$c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = d_1\mathbf{w}_1 + \cdots + d_q\mathbf{w}_q$$

and

$$c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p - d_1\mathbf{w}_1 - \cdots - d_q\mathbf{w}_q = \mathbf{0}.$$

But being a basis of V, $\{\mathbf{u}_1, \cdots, \mathbf{u}_p, \mathbf{w}_1, \cdots, \mathbf{w}_q\}$ is linearly independent hence $c_1 = \cdots = c_p = d_1 = \cdots = d_q = 0$ and $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = \mathbf{0}$. Therefore $U \cap W = \{\mathbf{0}\}$ and $V = U \oplus W$. ■

In Theorem 1.18 of section 1.6 we proved that there are many statements that are equivalent to saying an $n \times n$ matrix is invertible. The list was expanded in Theorem 3.9 of section 3.3. We add several statements to that list now with more to come later.

Theorem 4.35. Let A be an $n \times n$ matrix. The following are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row-echelon form of A is the identity matrix I_n.

(d) A is a product of elementary matrices.

(e) A has n pivot columns.

(f) A has a left inverse.

(g) A has a right inverse.

(h) For all $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(i) Every n-vector $\mathbf{b}$ is a linear combination of the columns of A.

(j) A^T is invertible.

(k) $\mathit{rank}\ A = n$.

(l) $\mathit{nullity}\ A = 0$.

(m) $\det A \neq 0$.

(n) The columns (or rows) of A span $\mathbb{R}^n$.

(o) The columns (or rows) of A are linearly independent.

(p) The columns (or rows) of A form a basis for $\mathbb{R}^n$.

(q) $\mathit{col}\ A = \mathbb{R}^n$.

(r) $\mathit{row}\ A = \mathbb{R}^n$.

(s) $\mathit{null}\ A = \{\mathbf{0}\}$.

Problem Set 4.4

1. Find the dimension of the subspace of $\mathbb{R}^3$ spanned by each of the following sets of vectors.

(a) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \\ 0 \end{bmatrix} \right\}$

(b) $\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} \right\}$

(c) $\left\{ \begin{bmatrix} 5 \\ 2 \\ 7 \end{bmatrix} \right\}$

(d) $\left\{ \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix} \right\}$

(e) $\left\{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right\}$

2. Which of the following are bases for the plane $3x - 5y + 2z = 0$?

(a) $\left\{\begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 0 \end{bmatrix}\right\}$ **(b)** $\left\{\begin{bmatrix} 3 \\ -5 \\ 2 \end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ -5 \end{bmatrix}\right\}$ **(d)** $\left\{\begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}\right\}$

3. Find a basis for and the dimension of each of the following subspaces of $\mathbb{R}^3$.

(a) The solution set of the equation $2x + 3y - 4z = 0$.

(b) The span of $\left\{\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 11 \\ -2 \end{bmatrix}\right\}$.

(c) The line $\dfrac{x}{2} = \dfrac{y}{5} = \dfrac{z}{7}$.

(d) The set of all vectors that are orthogonal to $\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$.

(e) The set of all vectors that are orthogonal to both $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$.

4. Find a basis for and the dimension of (i) the column space, (ii) the row space, and (iii) the null space for each of the following matrices.

(a) $\begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix}$ **(b)** $\begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 5 \\ -1 & 0 & 1 \end{bmatrix}$ **(d)** $\begin{bmatrix} 1 & 2 & 3 & 3 \\ 2 & 4 & 7 & 8 \\ -1 & -2 & 0 & 3 \end{bmatrix}$

5. Let H be the subspace of $\mathbb{R}^4$ of vectors that satisfy the equation $3x_1 - 4x_2 - 5x_3 + 6x_4 = 0$, and let K be the subspace of $\mathbb{R}^4$ of vectors that satisfy the symmetric equations $x_1 = x_2 = x_3 = x_4$.

(a) Find a basis for H.

(b) Find a basis for K.

(c) Extend your answer to (a) to a basis for $\mathbb{R}^4$.

(d) Extend your answer to (b) to a basis for $\mathbb{R}^4$.

(e) Verify that K is a subspace of H.

(f) Extend your answer to (b) to a basis for H.

6. Suppose $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for a vector space V. Determine whether each of the following are bases for V.

(a) $\{\mathbf{v}_1 + 2\mathbf{v}_2 + 3\mathbf{v}_3, 2\mathbf{v}_1 + 5\mathbf{v}_2 + 5\mathbf{v}_3, 2\mathbf{v}_1 + 8\mathbf{v}_2 + 3\mathbf{v}_3\}$

(b) $\{\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3, \mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3\}$

(c) $\{\mathbf{v}_1 + \mathbf{v}_2 + 3\mathbf{v}_3, 3\mathbf{v}_1 + 2\mathbf{v}_2 + 7\mathbf{v}_3, 2\mathbf{v}_1 - \mathbf{v}_2\}$

(d) $\{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_3, \mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_1 - \mathbf{v}_3\}$

7. Suppose A is a 5×7 matrix and rank $A = 3$.

(a) The column space of A is a subspace of $\mathbb{R}^n$ for $n =$ _____.

(b) What is the dimension of the column space of A?

(c) The row space of A is a subspace of $\mathbb{R}^n$ for $n =$ _____.

(d) What is the dimension of the row space of A?

(e) The null space of A is a subspace of $\mathbb{R}^n$ for $n =$ _____.

(f) What is the dimension of the null space of A?

(g) The null space of A^T is a subspace of $\mathbb{R}^n$ for $n =$ _____.

(h) What is the dimension of the null space of A^T?

8. Let $U = span\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}\right\}$ and $W = span\ \mathcal{S}$ where $\mathcal{S}$ might be any of the four sets below. Which of those four choices of $\mathcal{S}$ would make $\mathbb{R}^3 = U + W$? For those choices of $\mathcal{S}$ for which $\mathbb{R}^3 = U + W$, find $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \mathbf{u} + \mathbf{w}$.

(a) $\left\{\begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix}\right\}$

(b) $\left\{\begin{bmatrix} 7 \\ 10 \\ -8 \end{bmatrix}, \begin{bmatrix} 8 \\ 13 \\ -11 \end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}\right\}$

(d) $\left\{\begin{bmatrix} 8 \\ 11 \\ -9 \end{bmatrix}\right\}$

9. Suppose that A is an $m \times n$ matrix. Prove that $rank\ A^T = rank\ A$.

5 Linear Transformations

5.1 Definition of Linear Transformations

Now that we know what a vector space is, we are ready to begin studying functions from one vector space (domain) to another (codomain). In calculus we study functions that are differentiable, continuous, and integrable. In a linear algebra course, we study a particular type of function, called a linear transformation, that "preserves" vector space operations. What exactly do we mean by "preserves" in this context? Definition 5.1 explains.

Definition 5.1 (Linear Transformation)**.** Let V and W be vector spaces and $T : V \longrightarrow W$ a function from V to W. The function T is a **linear transformation** if it has the following two properties: For all vectors $\mathbf{u}, \mathbf{v}$ in V and for all scalars c,

(1) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(2) $T(c\mathbf{u}) = cT(\mathbf{u})$

We say a function is **linear** if it is a linear transformation.

A linear transformation preserves the two vector space operations of vector addition and scalar multiplication in the following sense:

(1) With linear transformations it doesn't matter whether you add the vectors first and then apply the function or apply the function first and then add the resulting vectors. You get the same result either way.

(2) The same goes with scalar multiplication. You get the same result whether you multiply by the scalar first and then apply the function or apply the function first and then multiply by the scalar.

Example 5.1

Let A be an $m \times n$ matrix. Define $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $T_A(\mathbf{x}) = A\mathbf{x}$. We show that T_A is a linear transformation by showing that it satisfies the two properties in the definition. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(1)
$$\begin{aligned} T_A(\mathbf{u}+\mathbf{v}) &= A(\mathbf{u}+\mathbf{v}) \text{ by the definition of } T_A \\ &= A\mathbf{u} + A\mathbf{v} \text{ by the distributive property of multiplication over matrix addition} \\ &= T_A(\mathbf{u}) + T_A(\mathbf{v}) \text{ by the definition of } T_A \end{aligned}$$

(2)
$$\begin{aligned} T_A(c\mathbf{u}) &= A(c\mathbf{u}) \\ &= c(A\mathbf{u}) \\ &= cT_A(\mathbf{u}) \end{aligned}$$

Therefore T_A is a linear transformation.

Definition 5.2. For any matrix A, the linear transformation T_A defined in Example 5.1 is called a **matrix transformation**.

We begin the next example with a geometric description of a function and show that the function is linear and, in fact, is a matrix transformation.

Example 5.2

Define $T_\theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ as follows: The function T_θ takes a vector $\mathbf{x}$ in $\mathbb{R}^2$ and rotates it counterclockwise by the angle θ about the origin.

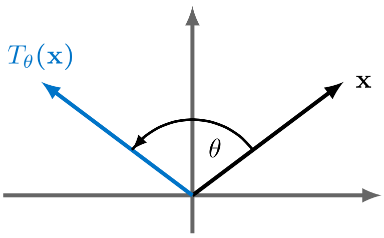
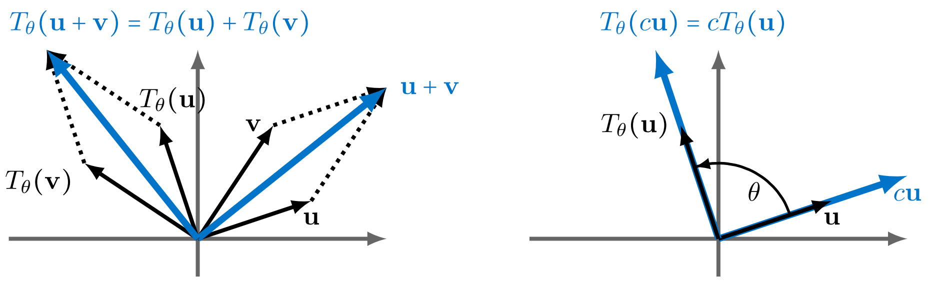


Figure 5.1 *T_θ rotates a vector counterclockwise by an angle of θ.*

By examining Figure 5.2, it is easy to convince yourself by the geometry that T_θ is linear.

We can write T_θ in terms of matrix multiplication as follows: Suppose $\mathbf{x}$ is a nonzero vector in $\mathbb{R}^2$ that makes an angle ϕ with the positive x-axis. Let $r = \|\mathbf{x}\|$. Then the unit vector in the direction of $\mathbf{x}$ is $\begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$ and $\mathbf{x} = r\begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$. Since T_θ rotates $\mathbf{x}$ by an



(a) Add first then rotate by θ. Compare to rotate first then add.

(b) Multiply first and then rotate by θ. Compare to rotate first then multiply.

Figure 5.2 *Geometric justification that T_θ is linear.*

angle θ and doesn't change its norm,

$$\begin{aligned}
T_\theta(\mathbf{x}) &= r\begin{bmatrix} \cos(\phi+\theta) \\ \sin(\phi+\theta) \end{bmatrix} \\
&= r\begin{bmatrix} \cos\phi\cos\theta - \sin\phi\sin\theta \\ \cos\phi\sin\theta + \sin\phi\cos\theta \end{bmatrix} \\
&= r\left(\cos\phi\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + \sin\phi\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}\right) \\
&= r\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\left(r\begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}\right) \\
&= R_\theta\mathbf{x}
\end{aligned}$$

where R_θ is the matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. So, we see that this rotation function T_θ is, in fact, a matrix transformation, hence T_θ is a linear transformation.

Definition 5.3. The 2×2 matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is called the **rotation matrix**.

Know Definition 5.3. To rotate a vector $\mathbf{v} \in \mathbb{R}^2$ by angle θ, simply multiply it on the left by R_θ.

Definition 5.4. A linear transformation from a vector space V to itself is called a **linear operator**. A **linear functional** is a linear transformation from a vector space to $\mathbb{R}$.

Since the domain and codomain of the rotation transformation $T_\theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ are the same, T_θ is a linear operator. A matrix transformation T_A is a linear operator if and only if A is square, and T_A is a linear functional if and only if A is a row vector.

Linear transformations in general and linear operators are found throughout mathematics. Example 5.3 shows that you worked with a linear operator back in calculus.

Example 5.3

Let $\mathbb{D}_\infty$ be the vector space of real-valued functions of a single variable that have derivatives of all orders. Define $D : \mathbb{D}_\infty \longrightarrow \mathbb{D}_\infty$ by $D(f) = f'$. That is, the function D maps a function to its derivative. The following two rules of differentiation tell us that D is a linear operator.

1. The derivative of a sum equals the sum of the derivative

$$D(f+g) = D(f) + D(g).$$

2. The derivative of a constant (scalar) times a function equals the constant times the derivative of the function

$$D(cf) = cD(f).$$

When we say a function is linear in linear algebra we almost always mean that the function is a linear transformation. This can cause some confusion because the term linear can have a different meaning in some other mathematical contexts. Example 5.4 illustrates.

Example 5.4

The set of real numbers, $\mathbb{R}$, is a one-dimensional vector space. The two functions $f(t) = 3t+1$ and $g(t) = 3t$ are two functions from $\mathbb{R}$ to $\mathbb{R}$. Are these functions linear?

The answer really depends on what you mean by linear. As you know, both f and g have graphs that are straight lines, so in that sense, both f and g are linear. But even in the sense of being a linear transformation, the function g is linear because

$$\begin{array}{rcl} g(s+t) & = & 3(s+t) \\ & = & 3s+3t \\ & = & g(s)+g(t) \end{array} \quad \text{and} \quad \begin{array}{rcl} g(ct) & = & 3(ct) \\ & = & c(3t) \\ & = & cg(t). \end{array}$$

But the function f is not linear in this sense since $f(s+t) = 3(s+t)+1$ but $f(s)+f(t) = (3s+1)+(3t+1) = 3(s+t)+2$. The function f can also be seen to be not linear in this sense since $f(2t) = 3(2t)+1 = 6t+1$ and $2f(t) = 2(3t+1) = 6t+2$. The only real-valued functions of a single variable that are actual linear transformations are those with graphs that are straight lines *through the origin.*

Example 5.5

Let $\mathbf{u}_0$ be a fixed vector in $\mathbb{R}^n$. Define $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ by $F(\mathbf{v}) = \mathbf{u}_0 \cdot \mathbf{v}$. It is easy to see that F is a linear functional since

$$\begin{array}{rcl} F(\mathbf{v}+\mathbf{w}) & = & \mathbf{u}_0 \cdot (\mathbf{v}+\mathbf{w}) \\ & = & \mathbf{u}_0 \cdot \mathbf{v} + \mathbf{u}_0 \cdot \mathbf{w} \\ & = & F(\mathbf{v}) + F(\mathbf{w}) \end{array} \quad \text{and} \quad \begin{array}{rcl} F(c\mathbf{v}) & = & \mathbf{u}_0 \cdot (c\mathbf{v}) \\ & = & c(\mathbf{u}_0 \cdot \mathbf{v}) \\ & = & cF(\mathbf{v}). \end{array}$$

Example 5.6

Let $\mathbb{I}[a,b]$ denote the space of all integrable functions on $[a,b]$. Define $F : \mathbb{I}[a,b] \longrightarrow \mathbb{R}$ by

$$F(f) = \int_a^b f(t)\, dt.$$

Properties of the definite integral imply that F is a linear functional because

$$\begin{array}{rcl} F(f+g) & = & \int_a^b (f(t)+g(t))\ dt \\ & = & \int_a^b f(t)\ dt + \int_a^b g(t)\ dt \\ & = & F(f)+F(g) \end{array} \quad \text{and} \quad \begin{array}{rcl} F(cf) & = & \int_a^b cf(t)\ dt \\ & = & c\int_a^b f(t)\ dt \\ & = & cF(f). \end{array}$$

The fact that a linear transformation preserves the operations of vector addition and scalar multiplication tells us a good deal about the function. If we know how a linear transformation acts on a basis of the domain, we know how it acts on the entire domain.

Theorem 5.1. Suppose V and W are vector spaces, $T : V \longrightarrow W$ is linear, and $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V. If $\mathbf{u} \in V$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, then $T(\mathbf{u}) = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)$.

Proof

$$\begin{array}{rcl} T(\mathbf{u}) & = & T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \text{ by the definition of } \mathbf{u} \\ & = & T(c_1\mathbf{v}_1) + \cdots + T(c_n\mathbf{v}_n) \text{ by property (1) of linear transformations} \\ & = & c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) \text{ by property (2) of linear transformations.} \end{array}$$

■

Theorem 5.1 shows us just how restricted the class of linear transformations is. There are infinitely many vectors in an n-dimensional vector space ($n \geq 1$), but once we determine how a linear transformation T acts on just n basis vectors of the domain, the rest of T is completely determined. Beyond that, however, we have complete flexibility as Theorem 5.2 shows.

Theorem 5.2. Suppose V and W are vector spaces with $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ a basis for V. Let $\mathbf{w}_1, \cdots, \mathbf{w}_n$ be a collection of any n vectors from W (repeats allowed). There exists a unique linear transformation $T : V \longrightarrow W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for $i = 1, \cdots, n$.

Proof Define $T : V \longrightarrow W$ as follows: Let $\mathbf{u} \in V$. Since $\mathcal{B}$ is a basis for V, there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Define $T(\mathbf{u}) = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$. Since $\mathbf{v}_i = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{i-1} + 1\mathbf{v}_i + 0\mathbf{v}_{i+1} + \cdots + 0\mathbf{v}_n$ we see that $T(\mathbf{v}_i) = 0\mathbf{w}_1 + \cdots + 0\mathbf{w}_{i-1} + 1\mathbf{w}_i + 0\mathbf{w}_{i+1} + \cdots + 0\mathbf{w}_n = \mathbf{w}_i$ for $i = 1, \cdots, n$. We show T is linear. Suppose $\mathbf{u}, \mathbf{v} \in V$ and k is a scalar. Since $\mathcal{B}$ is a basis, there exist unique $c_1, \cdots, c_n, d_1, \cdots, d_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ and $\mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$. Now, $\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_n + d_n)\mathbf{v}_n$, and thus

$$\begin{array}{rcl} T(\mathbf{u}+\mathbf{v}) & = & (c_1 + d_1)\mathbf{w}_1 + \cdots + (c_n + d_n)\mathbf{w}_n \\ & = & (c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n) + (d_1\mathbf{w}_1 + \cdots + d_n\mathbf{w}_n) \\ & = & T(\mathbf{u}) + T(\mathbf{v}). \end{array}$$

And,

$$\begin{aligned} T(k\mathbf{u}) &= T\left(k(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n)\right) \\ &= T\left((kc_1)\mathbf{v}_1 + \cdots + (kc_n)\mathbf{v}_n\right) \\ &= (kc_1)\mathbf{w}_1 + \cdots + (kc_n)\mathbf{w}_n \\ &= k(c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n) \\ &= kT(\mathbf{u}). \end{aligned}$$

Since T is determined on $\mathcal{B}$, T is unique by Theorem 5.1. ■

Definition 5.5. When carrying out this process of defining a linear transformation T on all of V by determining T on a basis only we say we **extend** T **linearly** to all of V.

Theorems 5.1 and 5.2 make it particularly easy to describe a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. In fact, they imply that *all* linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are, indeed, matrix transformations, and the matrix is very easy to find.

Theorem 5.3. Suppose $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear. There exists an $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ and, in fact, $A = [T(\mathbf{e}_1)\cdots T(\mathbf{e}_n)]$ where $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$.

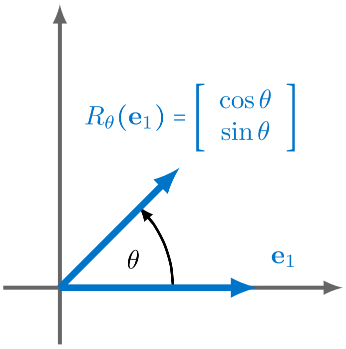
Proof Define $M : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $M(\mathbf{x}) = A\mathbf{x}$ where A is defined as above. We know that M is a matrix transformation, hence M is linear. Note that

$$M(\mathbf{e}_j) = \begin{bmatrix} & a_{1j} & \\ * & \vdots & * \\ & a_{mj} & \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} = T(\mathbf{e}_j)$$

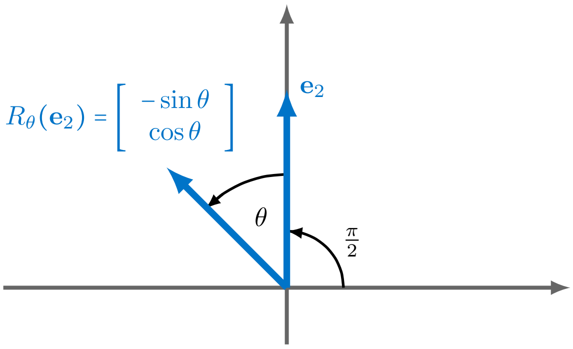
for $j = 1, \cdots, n$. Since T and M are both linear and agree on the standard basis $\mathcal{S}_n$, they must agree on all of $\mathbb{R}^n$. Therefore, $T = M$ and $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. ■

Definition 5.6. Suppose $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear. The $m \times n$ matrix $A = [T(\mathbf{e}_1)\cdots T(\mathbf{e}_n)]$ is called the **standard matrix representation of** T.

Let $\mathbf{u}$ be a nonzero vector in $\mathbb{R}^3$. Recall that the orthogonal projection of a vector $\mathbf{x}$ onto $\mathbf{u}$ is given by $proj_{\mathbf{u}}\mathbf{x} = \frac{\mathbf{u}\cdot\mathbf{x}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. The function $P_{\mathbf{u}} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ defined by $P_{\mathbf{u}}(\mathbf{x}) = proj_{\mathbf{u}}\mathbf{x}$ can be shown to be a linear operator (exercise).



(a) The image of $\mathbf{e}_1$ under R_θ.



(b) The image of $\mathbf{e}_2$ under R_θ.

Figure 5.3 *Defining the columns of the matrix R_θ.*

Example 5.7

Find the projection matrix A for $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ such that $P_{\mathbf{u}}(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^3$.

Solution $proj_{\mathbf{u}}\mathbf{e}_1 = \frac{1}{14}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $proj_{\mathbf{u}}\mathbf{e}_2 = \frac{2}{14}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, and $proj_{\mathbf{u}}\mathbf{e}_3 = \frac{3}{14}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, so

$$A = \begin{bmatrix} \frac{1}{14} & \frac{2}{14} & \frac{3}{14} \\ \frac{2}{14} & \frac{4}{14} & \frac{6}{14} \\ \frac{3}{14} & \frac{6}{14} & \frac{9}{14} \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

Example 5.8

Theorem 5.3 makes it easier to remember the 2×2 rotation matrix R_θ. The two columns of R_θ are the results after rotating $\mathbf{e}_1$ and $\mathbf{e}_2$ counterclockwise by an angle θ. But, as shown in Figure 5.3,

$$R_\theta(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \text{ and } R_\theta(\mathbf{e}_2) = \begin{bmatrix} \cos(\theta + \frac{\pi}{2}) \\ \sin(\theta + \frac{\pi}{2}) \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

So,

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

There are several ways of building new functions from old that you have learned about before. Many of them make sense in a linear algebra context.

Definition 5.7. The sum and difference of two functions: $(f \pm g)(x) = f(x) \pm g(x)$.
Scalar multiple of a function: $(kf)(x) = k(f(x))$.
Composition of two functions: $(g \circ f)(x) = g(f(x))$.

If the old functions are linear, what about the new functions?

Theorem 5.4. Suppose U, V, and W are vector spaces where $R : U \longrightarrow V$, $S : V \longrightarrow W$, and $T : V \longrightarrow W$ are linear, and k is a scalar.

(a) $(S \pm T) : V \longrightarrow W$ is linear.

(b) $(kT) : V \longrightarrow W$ is linear.

(c) $(S \circ R) : U \longrightarrow W$ is linear.

Proof Exercises. ∎

We can say more when the vector spaces are spaces of column vectors. Addition and subtraction of linear transformations correspond to matrix addition and subtraction. A scalar multiple of a linear transformation corresponds to a scalar multiple of a matrix. And, the composition of two linear transformations corresponds to matrix multiplication.

Theorem 5.5. Suppose A and B are $m \times p$ matrices, C is a $p \times n$ matrix and k is a scalar. Then $T_A : \mathbb{R}^p \longrightarrow \mathbb{R}^m$, $T_B : \mathbb{R}^p \longrightarrow \mathbb{R}^m$, and $T_C : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ are linear and

(a) $T_A \pm T_B = T_{A \pm B}$.

(b) $kT_A = T_{kA}$.

(c) $T_A \circ T_C = T_{AC}$.

Proof

(a) $(T_A \pm T_B)(\mathbf{v}) = T_A(\mathbf{v}) \pm T_B(\mathbf{v}) = A\mathbf{v} \pm B\mathbf{v} = (A \pm B)\mathbf{v} = T_{A \pm B}(\mathbf{v})$.

(b) $kT_A(\mathbf{v}) = k\left(T_A(\mathbf{v})\right) = k(A\mathbf{v}) = (kA)\mathbf{v} = T_{kA}(\mathbf{v})$.

(c) $(T_A \circ T_C)(\mathbf{v}) = T_A\left(T_C(\mathbf{v})\right) = T_A(C\mathbf{v}) = A(C\mathbf{v}) = (AC)\mathbf{v} = T_{AC}(\mathbf{v})$.

∎

Problem Set 5.1

1. Use the definition of linear transformation to determine whether the following functions are linear transformations. If linear, find the standard matrix representation of the linear transformation. If not linear, provide a counterexample to illustrates that one or the other of the two properties in the definition of linear transformation is not satisfied.

(a) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x+3y \\ x-4y \end{bmatrix}$

(b) $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} y \\ z \end{bmatrix}$

(c) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x+1 \\ x+y \\ y-1 \end{bmatrix}$

(d) $T : \mathbb{R}^2 \to \mathbb{R}$, where $T(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}$.

(e) $T : \mathbb{R}^2 \to \mathbb{R}$, where $T(\mathbf{x}) = \|\mathbf{x}\|$.

(f) $T : \mathbb{R}^3 \to \mathbb{R}^3$, where $\mathbf{u}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $T(\mathbf{x}) = \mathbf{u}_0 \times \mathbf{x}$.

2. Let $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 7 \end{bmatrix}\right\}$.

(a) Verify that $\mathcal{B}$ is a basis for $\mathbb{R}^2$.

(b) Let T be the linear transformation that maps $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ to $\begin{bmatrix} 2 \\ -5 \end{bmatrix}$ and maps $\begin{bmatrix} 3 \\ 7 \end{bmatrix}$ to $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$. Find the image of $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$ under T.

(c) Find the standard matrix representation of the linear transformation T that is defined in part (b).

3. Let $A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 3 & 1 \\ 3 & 0 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 3 & -1 \\ 0 & 2 & 1 \end{bmatrix}$. Let T_A and T_B be the matrix transformations for A and B respectively. Find the standard matrix representation for each of the following linear transformations.

(a) $T_A + T_B$ **(b)** $T_A - T_B$ **(c)** $3T_A$

(d) $T_A \circ T_B$ **(e)** $T_B \circ T_A$

4. Similar to the rotation operator $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, for each angle θ there is a reflection operator $F_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that maps each vector in $\mathbb{R}^2$ to its reflection across the line through the origin that makes an angle θ with the positive x axis. It can be shown that F_θ is a linear operator (similar to the manner in which the text shows R_θ is linear).

(a) Find the standard matrix representation of F_θ.

(b) Using matrix multiplication and trigonometric identities, show that the composition of two reflections is a rotation by finding the angle θ in terms of the angles α and β such that $F_\beta \circ F_\alpha = R_\theta$.

(c) Show that the composition of a reflection followed be a rotation is a reflection by finding the angle θ in terms of α and β such that $R_\beta \circ F_\alpha = F_\theta$.

(d) By comparing standard matrix representations, show that $R_\theta^{-1} = R_{-\theta}$ and $F_\theta^{-1} = F_\theta$.

(e) Show that the composition of a rotation followed be a reflection is a reflection by finding the angle θ in terms of α and β such that $F_\alpha \circ R_\beta = F_\theta$.

(f) Show that the composition of three reflections is a reflection by finding the angle θ in terms of α, β, and γ such that $F_\gamma \circ F_\beta \circ F_\alpha = F_\theta$.

5. From chapter 2 we see that the orthogonal projection of a vector $\mathbf{v}$ onto a nonzero vector $\mathbf{a}$ is given by $proj_{\mathbf{a}}\mathbf{v} = \dfrac{\mathbf{v}\cdot\mathbf{a}}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}$. For each nonzero vector $\mathbf{u}$ in $\mathbb{R}^3$, define the orthogonal projection operator $P_{\mathbf{u}} : \mathbb{R}^3 \to \mathbb{R}^3$ given by $P_{\mathbf{u}}(\mathbf{x}) = \dfrac{\mathbf{x}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. Prove that $P_{\mathbf{u}}$ is linear.

6. Suppose that U, V, and W are vector spaces and that $R : U \to V$, $S : U \to V$, and $T : V \to W$ are linear transformations and c is a scalar. Prove the following:

(a) $R + S$ is linear.

(b) $R - S$ is linear.

(c) cR is linear.

(d) $T \circ R$ is linear.

7. The definition of linear transformation is broken down into two parts (i) linear transformations preserve vector addition, and (ii) linear transformations preserve multiplication by scalars. Putting the two parts together tells us that linear transformations preserve linear combinations. In fact, an equivalent definition for linear transformation goes as follows:

Let V and W be vector spaces and $T : V \to W$ a function. The function T is called a linear transformation if for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars c and d we have $T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$.

Prove that the two definitions are equivalent.

5.2 The Kernel and Range of a Linear Transformation

Let V and W be vector spaces and $T : V \longrightarrow W$ a linear transformation. The vector space V is called the domain of T and W is called the codomain. There are subspaces of V and W that are associated with the linear transformation T. We begin with two of those subspaces and study more in the next two chapters.

Definition 5.8. Suppose V and W are vector spaces and $T : V \longrightarrow W$ is a linear transformation. The **kernel** of T, denoted *ker* T, is the set of all vectors in V that are mapped by T to $\mathbf{0}$ in W. That is,

$$\mathit{ker}\ T = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}.$$

Put another way, the kernel of T is the preimage of $\{\mathbf{0}\}$ under the linear transformation T.

Lemma 5.6. If V and W are vector spaces and $T : V \longrightarrow W$ is a linear transformation, then $T(\mathbf{0}_V) = \mathbf{0}_W$.

Proof $T(\mathbf{0}_V) = T(0\mathbf{0}_V) = 0T(\mathbf{0}_V) = \mathbf{0}_W$. ∎

Theorem 5.7. If V and W are vector spaces and $T : V \longrightarrow W$ is a linear transformation, then the kernel of T is a subspace of V.

Proof Use the subspace test.
▶ By Lemma 5.6, $\mathbf{0}_V \in \mathit{ker}\ T$, so $\mathit{ker}\ T \neq \varnothing$.
▶ Suppose $\mathbf{u}, \mathbf{v} \in \mathit{ker}\ T$ and show $\mathbf{u}+\mathbf{v} \in \mathit{ker}\ T$. But $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u})+T(\mathbf{v}) = \mathbf{0}_W+\mathbf{0}_W = \mathbf{0}_W$. So $\mathbf{u}+\mathbf{v} \in \mathit{ker}\ T$.
▶ Suppose $\mathbf{u} \in \mathit{ker}\ T$ and c is a scalar and show that $c\mathbf{u} \in \mathit{ker}\ T$. But $T(c\mathbf{u}) = cT(\mathbf{u}) = c\mathbf{0}_W = \mathbf{0}_W$. So $c\mathbf{u} \in \mathit{ker}\ T$.
Therefore, *ker* T is a subspace of V by the subspace test. ∎

Definition 5.9. Suppose V and W are vector spaces and $T : V \longrightarrow W$ is a linear transformation. The **range** of T, denoted *range* T, is the set of all vectors in W that are images of vectors in V under the linear transformation T. That is,

$$\mathit{range}\ T = \{\mathbf{w} \in W : T(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\}.$$

Put another way, the range of T is the image of V under the linear transformation T.

Theorem 5.8. If V and W are vector spaces and $T : V \longrightarrow W$ is a linear transformation, then the range of T is a subspace of W.

Proof Use the subspace test.
▶ By Lemma 5.6, $T(\mathbf{0}_V) = \mathbf{0}_W$, so $\mathbf{0}_W \in \mathit{range}\ T$ and $\mathit{range}\ T \neq \varnothing$.

▶ Suppose $\mathbf{w}_1, \mathbf{w}_2 \in range\ T$ and show $\mathbf{w}_1 + \mathbf{w}_2 \in range\ T$. Since $\mathbf{w}_1, \mathbf{w}_2 \in range\ T$, there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. So $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$. Thus $\mathbf{w}_1 + \mathbf{w}_2 \in range\ T$.
▶ Suppose $\mathbf{w} \in range\ T$ and c is a scalar and show $c\mathbf{w} \in range\ T$. Since $\mathbf{w} \in range\ T$, there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. So $T(c\mathbf{v}) = cT(\mathbf{v}) = c\mathbf{w}$ giving $c\mathbf{w} \in range\ T$.
Therefore the range of T is a subspace of W by the subspace test. ■

You have already encountered the range and kernel in the case when V and W are $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. In section 5.1 we showed that all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are, in fact, matrix transformations. So if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear, then there exists an $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$. So, a vector $\mathbf{b} \in \mathbb{R}^m$ is in the range of T if and only if the system $A\mathbf{x} = \mathbf{b}$ is consistent. Thus, the range of T equals the column space of A. Likewise, a vector $\mathbf{v} \in \mathbb{R}^n$ is in the kernel of T if and only if $\mathbf{v}$ is a solution to $A\mathbf{x} = \mathbf{0}$. Thus, the kernel of T equals the null space of A.

We also know the dimensions of these subspaces since the dimension of the column space of A equals the rank of A, and the dimension of the null space of A equals the nullity of A. We summarize these results in Theorem 5.9.

Theorem 5.9. Suppose A is an $m \times n$ matrix.

(a) $range\ T_A = col\ A$ and its dimension is $rank\ A$.

(b) $ker\ T_A = null\ A$ and its dimension is $nullity\ A$.

Example 5.9

Find the dimensions of and bases for the range and kernel of $T_A : \mathbb{R}^4 \longrightarrow \mathbb{R}^3$ where

$$A = \begin{bmatrix} 1 & -1 & 2 & 1 \\ 1 & 0 & 3 & 2 \\ 2 & -1 & 5 & 3 \end{bmatrix}.$$

Solution Reduce A.

$$\begin{bmatrix} 1 & -1 & 2 & 1 \\ 1 & 0 & 3 & 2 \\ 2 & -1 & 5 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

At this point we see that the first two columns are the only two pivot columns, so $rank\ A = 2$, $nullity\ A = 4 - 2 = 2$ and the first two columns of A form a basis for $range\ T_A$. By completing the reduction to reduced row-echelon form,

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

we can solve the homogeneous system $A\mathbf{x} = \mathbf{0}$ to find a basis for $ker\ T_A$. Letting $x_3 = s$ and $x_4 = t$ gives

$$\begin{array}{rcrcr} x_1 & = & -3s & - & 2t \\ x_2 & = & -s & - & t \\ x_3 & = & s & & \\ x_4 & = & & & t \end{array}$$

so that

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

is a basis for *range* T_A = *col* A and its dimension is 2. In addition,

$$\left\{ \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for *ker* T_A = *null* A and its dimension is 2.

Two important properties of functions you have studied elsewhere are one-to-one and onto. We define these properties here in case you don't recall their precise definitions.

Definition 5.10. A function $f : A \longrightarrow B$ is said to be **one-to-one** or an **injection** if $f(a_1) = f(a_2)$ implies $a_1 = a_2$. A function $f : A \longrightarrow B$ is said to be **onto** or a **surjection** if every element of B is in the range of f. If a function is both one-to-one and onto, we call it a **bijection**.

Recall that it is the bijections that have inverse functions and that the inverse functions are themselves bijections.

The kernels and ranges of linear transformations tell us which linear transformations are one-to-one and onto.

Theorem 5.10. Suppose V and W are vector spaces and $T : V \longrightarrow W$ is linear.

(a) T is one-to-one if and only if *ker* $T = \{\mathbf{0}\}$.

(b) T is onto if and only if *range* $T = W$.

Proof Suppose T is one-to-one. Since the kernel of T is the preimage of $\{\mathbf{0}\}$, T one-to-one implies *ker* T contains at most one vector. But Lemma 5.6 tells us $\mathbf{0} \in$ *ker* T, so *ker* $T = \{\mathbf{0}\}$. Suppose *ker* $T = \{\mathbf{0}\}$ and suppose $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. Show $\mathbf{v}_1 = \mathbf{v}_2$. But $T(\mathbf{v}_1) = T(\mathbf{v}_2) \implies T(\mathbf{v}_1) - T(\mathbf{v}_2) = \mathbf{0} \implies T(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0} \implies \mathbf{v}_1 - \mathbf{v}_2 \in$ *ker* T. But *ker* $T = \{\mathbf{0}\}$, so $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, which implies $\mathbf{v}_1 = \mathbf{v}_2$. This completes the proof of part (a).

Part (b) is trivial since T onto means *range* $T = W$. ∎

In the finite dimensional case, we can also determine whether a linear transformation is one-to-one or onto by looking at the image of a basis for the domain.

Theorem 5.11. Suppose V and W are vector spaces, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V, and $T : V \longrightarrow W$ is linear.

(a) T is one-to-one if and only if $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is linearly independent.

(b) T is onto if and only if $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ spans W.

Proof (a) Suppose T is one-to-one. Show $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is linearly independent. To show linear independence, suppose $c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) = \mathbf{0}$ and show $c_1 = \cdots = c_n = 0$. Now $c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) = \mathbf{0} \implies T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = \mathbf{0}$ by linearity of T. So $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in \mathit{ker}\ T$. By Theorem 5.10, T one-to-one implies $\mathit{ker}\ T = \{\mathbf{0}\}$, so $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$, and since $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V (and so linearly independent), $c_1 = \cdots = c_n = 0$. Therefore, $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is linearly independent.

Suppose $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is linearly independent. Show T is one-to-one. We use Theorem 5.10 and show $\mathit{ker}\ T = \{\mathbf{0}\}$ instead. Suppose $\mathbf{u} \in \mathit{ker}\ T$. Show $\mathbf{u} = \mathbf{0}$. Since $\mathbf{u} \in \mathit{ker}\ T \subseteq V$, there exists scalars $c_1, \cdots, c_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Since T is linear and $\mathbf{u} \in \mathit{ker}\ T$,

$$\begin{aligned} c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) &= T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= T(\mathbf{u}) \\ &= \mathbf{0}. \end{aligned}$$

But $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is linearly independent, so $c_1 = \cdots = c_n = 0$. Thus, $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = 0\mathbf{v}_1 + \cdots = 0\mathbf{v_n} = \mathbf{0}$, implying $\mathit{ker}\ T = \{\mathbf{0}\}$. So T is one-to-one.

(b) Suppose T is onto. Show $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ spans W. To do this, we suppose $\mathbf{w} \in W$ and show $\mathbf{w}$ is a linear combination of the vectors in $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$. Now, since T is onto, there exists $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{w}$, and since $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V, there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. So

$$\begin{aligned} \mathbf{w} &= T(\mathbf{u}) \\ &= T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) \end{aligned}$$

by the linearity of T. Thus $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ spans W.

Suppose $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ spans W. Show T is onto. We let $\mathbf{w} \in W$ and show there exists $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{w}$. Since $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ spans W, there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{w} = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)$. Let $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in V$. Then $T(\mathbf{u}) = T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) = \mathbf{w}$. So $\mathbf{w} \in \mathit{range}\ T$ and T is onto. ■

We have already used matrices for a variety of purposes. In this chapter we see that we can use them to describe all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. We turn now to the special case when $m = n$. In that case, the matrix A used to define the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is square. Since the domain and codomain are the same we can think of T_A as moving the vectors around in the same vector space. This perspective is particularly evident with the rotation matrix R_θ. In addition, since $\mathit{nullity}\ A = n - \mathit{rank}\ A$, we see that when A is square, T_A is onto precisely when T_A is one-to-one. This gives us several new ways to characterize invertible matrices. We continue from our last extension in section 4.4 (Theorem 4.35).

Theorem 5.12. Let A be an $n \times n$ matrix. The following are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row-echelon form of A is the identity matrix I_n.

(d) A is a product of elementary matrices.

(e) A has n pivot columns.

(f) A has a left inverse.

(g) A has a right inverse.

(h) For all $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(i) Every n-vector $\mathbf{b}$ is a linear combination of the columns of A.

(j) A^T is invertible.

(k) $\textit{rank } A = n$.

(l) $\textit{nullity } A = 0$.

(m) $\det A \neq 0$.

(n) The columns (or rows) of A span $\mathbb{R}^n$.

(o) The columns (or rows) of A are linearly independent.

(p) The columns (or rows) of A form a basis for $\mathbb{R}^n$.

(q) $\textit{col } A = \mathbb{R}^n$.

(r) $\textit{row } A = \mathbb{R}^n$.

(s) $\textit{null } A = \{\mathbf{0}\}$.

(t) $\textit{range } T_A = \mathbb{R}^n$.

(u) $\textit{ker } T_A = \{\mathbf{0}\}$.

(v) T_A is a surjection.

(w) T_A is an injection.

(x) T_A is a bijection.

Problem Set 5.2

1. For each part (a) - (c) below, let $T(\mathbf{x}) = A\mathbf{x}$. On an x,y coordinate plane, plot $\textit{ker } T$ and the preimages of $\{\mathbf{u}\}$, $\{\mathbf{v}\}$, and $\{\mathbf{w}\}$.

(a) $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(b) $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -3 \\ -6 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(c) $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

2. For each part (a) - (c) in Exercise 1 plot the range of T and determine whether T is injective and whether T is surjective. Also, for each part determine whether T has an inverse.

3. For each part (a) - (c) below, let $T(\mathbf{x}) = A\mathbf{x}$. In each case find a basis for the kernel of T and find a basis for the range of T.

(a) $A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 5 \end{bmatrix}$ **(b)** $A = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 1 & 3 & 2 & 5 \end{bmatrix}$ **(c)** $A = \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix}$

4. For each part (a) - (c) in Exercise 3 determine whether T is injective and surjective, and determine whether T has an inverse.

5. Check to see whether your answers to Exercise 3 match Theorem 5.9 regarding the dimensions of the range and kernel of T.

6. Let T be the linear operator on $\mathbb{R}^3$ defined by $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 4 & 7 \\ -2 & 2 & 0 \end{bmatrix}$.

(a) Show that *ker* T is a line through the origin in $\mathbb{R}^3$ by finding parametric equations for it.

(b) Show that *range* T is a plane through the origin in $\mathbb{R}^3$ by finding an equation for it.

7. Let A be a 5×7 matrix with *rank* $A = 4$ and T the linear transformation defined by $T(\mathbf{x}) = A\mathbf{x}$.

(a) What is the domain of T?

(b) What is the codomain of T?

(c) What is the dimension of *ker* T?

(d) What is the dimension of *range* T?

(e) Is T a one-to-one function?

(f) Is T an onto function?

(g) Does T have an inverse function?

8. Describe the kernel of the differential operator D described in Example 5.3. *Hint: Use your knowledge of calculus.*

9. Describe the kernel of the linear functional in Example 5.5 in the case where $n = 3$ and $\mathbf{u}_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

5.3 Isomorphisms and Coordinate Vectors

Functions between vector spaces that have all three properties of one-to-one, onto, and linearity are very special indeed. We give them their own name.

Definition 5.11. Suppose V and W are vector spaces. A linear transformation $T : V \longrightarrow W$ that is both onto and one-to-one is called an **isomorphism**. We say that V and W are **isomorphic** if there is an isomorphism from one to the other.

We mentioned earlier that the bijections are the functions that have inverse functions. In fact, those inverse functions are bijections too, so if $T : V \longrightarrow W$ is an isomorphism, then $T^{-1} : W \longrightarrow V$ is at least a bijection. In fact, T^{-1} is an isomorphism too.

Theorem 5.13. Suppose V and W are vector spaces and $T : V \longrightarrow W$ is an isomorphism. The isomorphism T has a unique inverse function T^{-1} and that inverse function is an isomorphism.

Proof Since $T : V \longrightarrow W$ is a bijection, we know that there is a unique inverse function $T^{-1} : W \longrightarrow V$ and that T^{-1} is a bijection. To complete the proof, we need only show T^{-1} is linear. Suppose $\mathbf{w}_1, \mathbf{w}_2 \in W$. Show $T^{-1}(\mathbf{w}_1 + \mathbf{w}_2) = T^{-1}(\mathbf{w}_1) + T^{-1}(\mathbf{w}_2)$. Since T is onto, there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Since T is a bijection, $T^{-1}(\mathbf{w}_1) = \mathbf{v}_1$ and $T^{-1}(\mathbf{w}_2) = \mathbf{v}_2$. Thus

$$\begin{aligned}
T^{-1}(\mathbf{w}_1 + \mathbf{w}_2) &= T^{-1}(T(\mathbf{v}_1) + T(\mathbf{v}_2)) \\
&= T^{-1}(T(\mathbf{v}_1 + \mathbf{v}_2)) \text{ by the linearity of } T \\
&= \mathbf{v}_1 + \mathbf{v}_2 \\
&= T^{-1}(\mathbf{w}_1) + T^{-1}(\mathbf{w}_2) \text{ by the definition of } T^{-1}.
\end{aligned}$$

Similarly, $T^{-1}(c\mathbf{w}) = cT^{-1}(\mathbf{w})$ so T^{-1} is an isomorphism. ■

Corollary 5.14. Let A be an $n \times n$ matrix. The linear transformation T_A is an isomorphism if and only if A is invertible. If A is invertible, then $T_A^{-1} = T_{A^{-1}}$.

Proof Exercise. ■

A bijection is a one-to-one correspondence between two sets, so an isomorphism is a one-to-one correspondence between to vector spaces that preserves the vector operations of vector addition and scalar multiplication. The next theorems show that if two vector spaces are isomorphic, then they are very, very similar indeed. It is as though the vector spaces are parallel universes populated by vectors in one-to-one correspondence. Two corresponding vectors might not look alike (one might be a polynomial and the other a column vector) but they play the exact same role in their respective universes.

Theorem 5.15. Suppose V and W are isomorphic vector spaces and $T : V \longrightarrow W$ is an isomorphism.

(a) $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is linearly independent if and only if $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is linearly independent.

(b) $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ spans V if and only if $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ spans W.

(c) $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V if and only if $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is a basis for W.

(d) If V and W are finite dimensional, then they have the same dimension.

Proof Exercises. ■

Because we know how to extend linearly, it is easy enough to show that two finite dimensional vector spaces are isomorphic if and only if they have the same dimension.

Theorem 5.16. Suppose V and W are finite dimensional vector spaces. These vector spaces are isomorphic if and only if they have the same dimension.

Proof Theorem 5.15 gives us one direction of this proof. If V and W are isomorphic, then they have the same dimension. For the other direction, suppose V and W are both n-dimensional vector spaces. Show V and W are isomorphic. Since V and W are both n-dimensional, they have bases $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ respectively containing n vectors each. Define $T : V \longrightarrow W$ by $T(\mathbf{v}_i) = \mathbf{w}_i$ for $i = 1, \cdots, n$ and extend linearly. So T is linear and Theorem 5.11 implies T is a bijection. Therefore, V and W are isomorphic. ■

For many purposes, the vector spaces $\mathbb{R}^n$ for various n are the nicest vector spaces with which to work. That is because those vectors lend themselves to matrix manipulation. Suppose, for example, we wished to determine whether a particular set of vectors from some abstract n-dimensional vector space V is linearly independent. One approach to answering that question would be to set up an isomorphism, T, between V and $\mathbb{R}^n$, look at the set of vectors in $\mathbb{R}^n$ that correspond to the set in question from V, and answer the question in $\mathbb{R}^n$ using matrix techniques. By Theorem 5.15 the answer must be the same for both sets. The vectors in $\mathbb{R}^n$ that represent the vectors in V are called coordinate vectors. In order to define coordinate vectors precisely we need something called an ordered basis.

Definition 5.12. Suppose V is an n-dimensional vector space. An **ordered basis** for V is a basis for V with the added property that one of its elements is designated as the first element, another is the second, etc. all the way to the n^{th}.

Definition 5.13. Let V be an n-dimensional vector space with an ordered basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. For each $\mathbf{u} \in V$, there exist unique scalars $c_1, \cdots, c_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. The **coordinate vector** for $\mathbf{u}$ relative to $\mathcal{B}$ is $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$.

Note that $\mathbf{u} \in V$ and $[\mathbf{u}]_{\mathcal{B}} \in \mathbb{R}^n$.

Theorem 5.17. Suppose V is an n-dimensional vector space with an ordered basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. The function $T_{\mathcal{B}} : V \longrightarrow \mathbb{R}^n$ given by $T_{\mathcal{B}}(\mathbf{u}) = [\mathbf{u}]_{\mathcal{B}}$ is an isomorphism.

Proof Note that $T_{\mathcal{B}}$ is the linear transformation defined by setting $T_{\mathcal{B}}(\mathbf{v}_j) = \mathbf{e}_j$ for $j = 1, \cdots, n$ where $\mathbf{e}_j$ is the j^{th} vector in the standard ordered basis $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ and then extended linearly. Since $\mathcal{B}$ and $\mathcal{S}_n$ are bases of V and $\mathbb{R}^n$ respectively, $T_{\mathcal{B}}$ is an isomorphism. ■

Example 5.10

Let $p(t) = t^2 + t + 1$, $q(t) = 2t^2 + 4t + 3$, and $r(t) = 5t^2 - t + 2$. Is the set $\{p(t), q(t), r(t)\}$ linearly independent in $\mathbb{P}_2$?

Solution The set $\mathcal{B} = \{1, t, t^2\}$ is an ordered basis for $\mathbb{P}_2$, and since $p(t) = 1(1) + 1(t) + 1(t^2)$, $q(t) = 3(1) + 4(t) + 2(t^2)$, and $r(t) = 2(1) - 1(t) + 5(t^2)$, we have

$$[p(t)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, [q(t)]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}, \text{ and } [r(t)]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}.$$

We check to see whether $\{[p(t)]_{\mathcal{B}}, [q(t)]_{\mathcal{B}}, [r(t)]_{\mathcal{B}}\}$ is linearly independent in $\mathbb{R}^3$. The answer for $\{p(t), q(t), r(t)\}$ in $\mathbb{P}_2$ must be the same. Row reduction (below) shows that there are only two pivot columns so that the set is linearly dependent.

$$\begin{bmatrix} 1 & 3 & 2 \\ 1 & 4 & -1 \\ 1 & 2 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & -3 \\ 0 & -1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

We can find coordinate vectors for any finite-dimensional vector space V and ordered basis $\mathcal{B}$ for V. Interestingly enough, we are most interested in finding coordinate vectors for vectors in $\mathbb{R}^n$ ($V = \mathbb{R}^n$). You may wonder what the point is because the vectors already lend themselves to matrix manipulation. It turns out that depending on the question we wish to answer, we can learn a great deal by looking at coordinate vectors under ordered bases $\mathcal{B}$ different from $\mathcal{S}_n$ chosen especially for the question we wish to answer. This will be made much clearer in the next chapter.

Suppose $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an ordered basis for $\mathbb{R}^n$. The function $T_{\mathcal{B}} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by $T_{\mathcal{B}}(\mathbf{u}) = [\mathbf{u}]_{\mathcal{B}}$ is an isomorphism, so there is an invertible $n \times n$ matrix A such that $[\mathbf{u}]_{\mathcal{B}} = A\mathbf{u}$. We wish to find A.

Theorem 5.18 is very important. We provide two proofs to give you two different perspectives.

Theorem 5.18. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be an ordered basis for $\mathbb{R}^n$ and let $P_{\mathcal{B}} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$. For each $\mathbf{u} \in \mathbb{R}^n$, the coordinate vector $[\mathbf{u}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{u}$ and $\mathbf{u} = P_{\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}$.

Proof (#1) Let $\mathbf{u} \in \mathbb{R}^n$. Since $\mathcal{B}$ is an ordered basis, there exist unique scalars $c_1, \cdots, c_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ and $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. That is to say $c_1, \cdots, c_n$ is the solution to the vector equation $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{u}$ or equivalently, $[\mathbf{u}]_{\mathcal{B}}$ is the solution to the matrix equation $P_{\mathcal{B}}\mathbf{x} = \mathbf{u}$. Thus $P_{\mathcal{B}}[\mathbf{u}]_{\mathcal{B}} = \mathbf{u}$.

Now $P_{\mathcal{B}}$ is invertible since its columns form a basis for $\mathbb{R}^n$. Multiplying both sides by $P_{\mathcal{B}}^{-1}$ we get $[\mathbf{u}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{u}$. ■

Proof (#2) Multiplying $P_{\mathcal{B}}$ on the left of $\mathbf{e}_j$ picks out the j^{th} column of $P_{\mathcal{B}}$, so $P_{\mathcal{B}}\mathbf{e}_j = \mathbf{v}_j$ for $j = 1, \cdots, n$. The matrix $P_{\mathcal{B}}$ is invertible since its columns are a basis for $\mathbb{R}^n$, so $P_{\mathcal{B}}^{-1}\mathbf{v}_j = \mathbf{e}_j$ for $j = 1, \cdots, n$. The coordinate vector transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by $T(\mathbf{u}) = [\mathbf{u}]_{\mathcal{B}}$ also maps $\mathbf{v}_j$ to $\mathbf{e}_j$ for $j = 1, \cdots, n$. Since these two linear transformations agree on the basis $\mathcal{B}$, they agree on all of $\mathbb{R}^n$, so $[\mathbf{u}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{u}$ for every $\mathbf{u} \in \mathbb{R}^n$. ■

Definition 5.14. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be an ordered basis for $\mathbb{R}^n$. Let $P_{\mathcal{B}}$ be the $n \times n$ matrix with columns consisting of the vectors of $\mathcal{B}$ in order, $P_{\mathcal{B}} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$. The matrix $P_{\mathcal{B}}$ is called the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{S}_n$ where $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is the standard ordered basis for $\mathbb{R}^n$. Its inverse $P_{\mathcal{B}}^{-1}$ is called the **change of basis matrix** from $\mathcal{S}_n$ to $\mathcal{B}$.

Example 5.11

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

It is easily verified that $\mathcal{B}$ is an ordered basis for $\mathbb{R}^3$. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Find the change of bases matrices $P_{\mathcal{B}}$ and $P_{\mathcal{B}}^{-1}$, and find $[\mathbf{u}]_{\mathcal{B}}$.

Solution

$$P_{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & -1 \\ 1 & 3 & 1 \end{bmatrix}.$$

So

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ -2 & -3 & -1 & 0 & 1 & 0 \\ 1 & 3 & 1 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 & -1 \end{array}\right] \longrightarrow$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -2 & -1 & 1 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 3 & 1 & -1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 3 & 1 & -1 \end{array}\right]$$

gives us

$$P_{\mathcal{B}}^{-1} = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix} \text{ and } [\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -5 \\ 2 \\ 2 \end{bmatrix}.$$

To check this, note that

$$-5\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 2 \\ -3 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Theorem 5.19. Let $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be the standard ordered basis for $\mathbb{R}^n$. For all $\mathbf{u} \in \mathbb{R}^n$, $[\mathbf{u}]_{\mathcal{S}_n} = \mathbf{u}$.

Proof Exercise. ∎

It is important to think about what is happening geometrically. You are very familiar with the Cartesian coordinate system. We start with a point on the plane called the origin. Then we draw two perpendicular lines through the origin called the x and y axes. Finally, we mark off unit distances on these axes. It is sometimes helpful to make a grid with lines parallel to the axes. In Figure 5.4, the ordered pair $(2, 3)$ identifies the point on the plane located by moving two units to the right of the origin and then three units up.

In linear algebra, we accomplish the same thing using different notation. We tend to use column vectors instead of ordered pairs. We start with a point on the plane called the origin. Then we take two unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ that are perpendicular. We say the column vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ represents the point $2\mathbf{e}_1 + 3\mathbf{e}_2$. The Cartesian coordinate system describes all points (vectors) in terms of the standard ordered basis $\mathcal{S}_2 = \{\mathbf{e}_1, \mathbf{e}_2\}$.

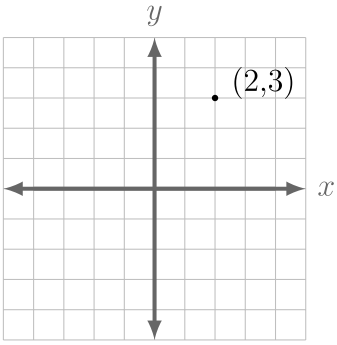


Figure 5.4 *The ordered pair* $(2,3)$ *in the Cartesian plane.*

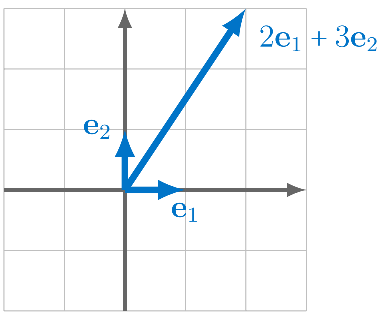


Figure 5.5 $\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2\mathbf{e}_1 + 3\mathbf{e}_2.$

But $\mathbb{R}^2$ has many bases. For $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ to be a basis $\mathbf{v}_1$ and $\mathbf{v}_2$ need only to be nonzero and nonparallel. They do not need to be perpendicular. They do not need to be of unit length. They do not even need to be the same length. A coordinate vector $[\mathbf{u}]_{\mathcal{B}}$ simply gives the coordinates of $\mathbf{u}$ relative to $\mathcal{B}$ rather than $\mathcal{S}_2$. If $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} a \\ b \end{bmatrix}$, that means $\mathbf{u} = a\mathbf{v}_1 + b\mathbf{v}_2$.

Example 5.12 Rotation Matrices and Rotation of Axes

In section 5.1 we studied the rotation matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

This is a great illustration of the fact that what goes on in linear algebra depends heavily on interpretation. Back in section 5.1 we were working with the standard basis $\mathcal{S}_2$. We thought of the matrix R_θ as moving a vector $\mathbf{u}$ to a new position. That is, $R_\theta\mathbf{u}$ is a different vector obtained by rotating $\mathbf{u}$ in a counterclockwise direction by an angle θ around the origin. It is easy enough to show (an exercise) that

$$R_\theta^{-1} = R_{-\theta} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

The basis

$$\mathcal{B} = \left\{ \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}, \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix} \right\}$$

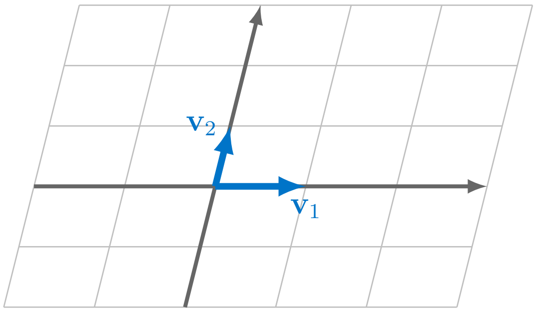


Figure 5.6 *A basis* $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ *for* $\mathbb{R}^2$ *need not have* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *perpendicular or even of the same length.*

can be obtained from $\mathcal{S}_2$ by rotating the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ by an angle $-\theta$. So, the change of basis matrices $P_{\mathcal{B}} = R_{-\theta}$ and $P_{\mathcal{B}}^{-1} = R_{-\theta}^{-1} = R_\theta$. This gives us two interpretations of $R_\theta \mathbf{u}$.

1. $R_\theta \mathbf{u}$ is the vector obtained by rotating $\mathbf{u}$ counterclockwise by an angle θ.
2. $R_\theta \mathbf{u} = [\mathbf{u}]_{\mathcal{B}}$ is the coordinate vector of $\mathbf{u}$ relative to $\mathcal{B}$.

In the second interpretation, $\mathbf{u}$ stays fixed but the reference points (the basis) are rotated **clockwise** (not counterclockwise) by the angle θ. Though the interpretation is different, the result is the same.

If A is an invertible matrix, we can think of $T_A(\mathbf{u})$ as $\mathbf{u}$ taking a trip by physically going to a new location. We can almost imagine $[\mathbf{u}]_{\mathcal{B}}$ as $\mathbf{u}$ taking a trip on drugs. The vector $\mathbf{u}$ never moves, but all points of reference change around $\mathbf{u}$.

Let V be a vector space with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$. Since $T_{\mathcal{B}}$ and $T_{\mathcal{C}}$ are isomorphisms their inverses are isomorphisms and the composition $T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is an isomorphism (see Figure 5.7). Since this composition is an isomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$, it is a matrix transformation, so there is an $n \times n$ matrix M such that $M\,[\mathbf{u}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{C}}$. We wish to find M. We consider two cases.

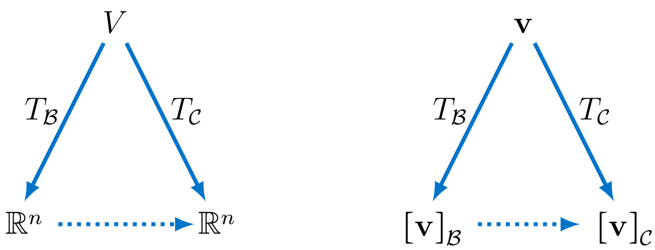


Figure 5.7 $T_{\mathcal{B}}$ *and* $T_{\mathcal{C}}$ *isomorphisms* $\Longrightarrow$ $T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}$ *is an isomorphism.*

The easy case occurs if $V = \mathbb{R}^n$ because then the isomorphisms $T_{\mathcal{B}}$ and $T_{\mathcal{C}}$ are themselves matrix transformations. Let $P_{\mathcal{B}} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ and $P_{\mathcal{C}} = [\mathbf{w}_1 \cdots \mathbf{w}_n]$, so $T_{\mathcal{B}}^{-1} = T_{P_{\mathcal{B}}}$ and $T_{\mathcal{C}} = T_{P_{\mathcal{C}}^{-1}}$, so the composition $T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1} = T_{P_{\mathcal{C}}^{-1}} T_{P_{\mathcal{B}}}$. Thus,

$$\begin{aligned} [\mathbf{u}]_{\mathcal{C}} &= T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}([\mathbf{u}]_{\mathcal{B}}) \\ &= T_{P_{\mathcal{C}}^{-1}} \circ T_{P_{\mathcal{B}}}([\mathbf{u}]_{\mathcal{B}}) \\ &= P_{\mathcal{C}}^{-1} P_{\mathcal{B}} [\mathbf{u}]_{\mathcal{B}} . \end{aligned}$$

This proves Theorem 5.20.

Theorem 5.20. Suppose $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ are ordered bases of $\mathbb{R}^n$ and $P_{\mathcal{B}} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ and $P_{\mathcal{C}} = [\mathbf{w}_1 \cdots \mathbf{w}_n]$. Then

$$[\mathbf{u}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1} P_{\mathcal{B}} [\mathbf{u}]_{\mathcal{B}}$$

for every $\mathbf{u} \in \mathbb{R}^n$.

The problem is a little more difficult if $V \neq \mathbb{R}^n$ because then $T_{\mathcal{B}}$ and $T_{\mathcal{C}}$ are not matrix transformations, but the composition $T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}$ is still a matrix transformation. For simplicity, let $S = T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}$. We know that $S(\mathbf{u}) = Q\mathbf{u}$ for all $\mathbf{u} \in \mathbb{R}^n$ where $Q = [S(\mathbf{e}_1) \cdots S(\mathbf{e}_n)]$.

For each $j = 1, \cdots, n$, $S(\mathbf{e}_j) = T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}(\mathbf{e}_j) = T_{\mathcal{C}}(\mathbf{v}_j) = [\mathbf{v}_j]_{\mathcal{C}}$. Thus $Q = [[\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}]$. This proves Theorem 5.21.

Theorem 5.21. Let V be a vector space with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$. Let the $n \times n$ matrix $Q = [[\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}]$. Then for all $\mathbf{u} \in V$,

$$[\mathbf{u}]_{\mathcal{C}} = Q [\mathbf{u}]_{\mathcal{B}} .$$

Of course, reversing the roles of $\mathcal{B}$ and $\mathcal{C}$ we obtain the inverse transformation defined by the inverse matrix

$$Q^{-1} = [[\mathbf{w}_1]_{\mathcal{B}} \cdots [\mathbf{w}_n]_{\mathcal{B}}]$$

where $[\mathbf{u}]_{\mathcal{B}} = Q^{-1} [\mathbf{u}]_{\mathcal{C}}$.

Definition 5.15. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ be ordered bases for a vector space V. The $n \times n$ matrix $Q = [[\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}]$ is called the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{C}$. Its inverse $Q^{-1} = [[\mathbf{w}_1]_{\mathcal{B}} \cdots [\mathbf{w}_n]_{\mathcal{B}}]$ is called the **change of basis matrix** from $\mathcal{C}$ to $\mathcal{B}$.

Usually one or the other of the two matrices Q and Q^{-1} is easier to calculate. Calculate that one and use matrix techniques to find its inverse. Example 5.13 illustrates.

Example 5.13

The vector space $\mathbb{P}_3$ has ordered bases $\mathcal{B} = \{1, t, t^2, t^3\}$ and $\mathcal{C} = \{1, t-1, (t-1)^2, (t-1)^3\}$. Calculate the change of bases matrices and use them to rewrite the following polynomials in terms of powers of $t-1$ rather than powers of t.

(a) $t^3 + 2t^2 + t - 1$ **(b)** $t^3 - 1$ **(c)** $t^2 + 1$

Solution By simply expanding the powers of $t-1$, we easily write the basis vectors in $\mathcal{C}$ in terms of $\mathcal{B}$, so the matrix $\left[[1]_{\mathcal{B}}\,[t-1]_{\mathcal{B}}\left[(t-1)^2\right]_{\mathcal{B}}\left[(t-1)^3\right]_{\mathcal{B}}\right]$ is easy to calculate:

$$\begin{aligned}
1 &= 1(1) + 0(t) + 0(t^2) + 0(t^3) \\
t-1 &= -1(1) + 1(t) + 0(t^2) + 0(t^3) \\
(t-1)^2 = t^2 - 2t + 1 &= 1(1) - 2(t) + 1(t^2) + 0(t^3) \\
(t-1)^3 = t^3 - 3t^2 + 3t - 1 &= -1(1) + 3(t) - 3(t^2) + 1(t^3)
\end{aligned}$$

So,

$$[1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, [t-1]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \left[(t-1)^2\right]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \left[(t-1)^3\right]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 3 \\ -3 \\ 1 \end{bmatrix}.$$

Thus

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

is the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$. Augmenting this change of basis matrix to the identity matrix and performing row reduction gives

$$\left[\begin{array}{cccc|cccc} 1 & -1 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$
$$\downarrow$$
$$\left[\begin{array}{cccc|cccc} 1 & -1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$
$$\downarrow$$
$$\left[\begin{array}{cccc|cccc} 1 & -1 & 0 & 0 & 1 & 0 & -1 & 2 \\ 0 & 1 & 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$
$$\downarrow$$
$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

so that

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

is the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$.

(a) $$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \\ 5 \\ 1 \end{bmatrix} \Longrightarrow \begin{array}{c} t^3 + 2t^2 + t - 1 = \\ (t-1)^3 + 5(t-1)^2 + 8(t-1) + 3 \end{array}.$$

(b) $$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 3 \\ 1 \end{bmatrix} \Longrightarrow t^3 - 1 = (t-1)^3 + 3(t-1)^2 + 3(t-1).$$

(c) $$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix} \Longrightarrow t^2 + 1 = (t-1)^2 + 2(t-1) + 2.$$

The method developed in Theorem 5.21 for finding change of bases matrices applies equally well to the case where $V = \mathbb{R}^n$. In fact, this case is particularly interesting and the method is different from that in Theorem 5.20. This idea is developed below and Example 5.14 illustrates both methods.

Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ be ordered bases for $\mathbb{R}^n$. We wish to find the change of basis matrix Q from $\mathcal{B}$ to $\mathcal{C}$. By Theorem 5.21, $Q = [[\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}]$. We also know that for all $\mathbf{u} \in \mathbb{R}^n$, $P_{\mathcal{C}}\,[\mathbf{u}]_{\mathcal{C}} = \mathbf{u}$ so $[\mathbf{u}]_{\mathcal{C}}$ is the solution to the system $P_{\mathcal{C}}\mathbf{x} = \mathbf{u}$. In particular then, for $j = 1, \cdots, n$, $[\mathbf{v}_j]_{\mathcal{C}}$ is the solution to the system $P_{\mathcal{C}}\mathbf{x} = \mathbf{v}_j$. This gives us n systems to solve to find each column of Q. We can do that all at once by augmenting the right-hand side of all n systems to $P_{\mathcal{C}}$ and reducing.

$$[P_{\mathcal{C}} | \mathbf{v}_1 \cdots \mathbf{v}_n] \longrightarrow \cdots \longrightarrow [I_n | \, [\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}]$$

But $[\mathbf{v}_1 \cdots \mathbf{v}_n] = P_{\mathcal{B}}$ and $[I_n | \, [\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}] = [I_n | Q]$, so the reduction above can be rewritten

$$[P_{\mathcal{C}} | P_{\mathcal{B}}] \longrightarrow \cdots \longrightarrow [I_n | Q]\,.$$

Example 5.14

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\} \text{ and } \mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

The sets $\mathcal{B}$ and $\mathcal{C}$ are ordered bases of $\mathbb{R}^3$. Find the change of basis matrix, Q, from $\mathcal{B}$ to $\mathcal{C}$.

Solution Using the method of Theorem 5.20,

$$P_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \text{ and } P_{\mathcal{C}} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

To find $P_{\mathcal{C}}^{-1}$, row reduction gives

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

so

$$P_{\mathcal{C}}^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$Q = P_{\mathcal{C}}^{-1} P_{\mathcal{B}} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Now, using the method of Theorem 5.21, we have

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right]$$

so that

$$Q = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$$

as before.

Problem Set 5.3

1. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$.

(a) Find the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}_2$.

(b) Find the change of basis matrix from $\mathcal{S}_2$ to $\mathcal{B}$.

(c) Find the coordinate vector $[\mathbf{u}]_{\mathcal{B}}$ where $\mathbf{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

2. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} \right\}$.

(a) Find the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}_3$.

(b) Find the change of basis matrix from $\mathcal{S}_3$ to $\mathcal{B}$.

(c) Find the coordinate vector $[\mathbf{u}]_{\mathcal{B}}$ where $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

3. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \right\}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. Repeat Exercise 2 for this $\mathcal{B}$ and $\mathbf{u}$ if possible. If impossible, explain why.

4. In the vector space $\mathbb{P}_3$ of all polynomials with real coefficients of degree less than or equal to 3, let $\mathcal{S} = \{1, t, t^2, t^3\}$, $\mathcal{B} = \{1, t-2, (t-2)^2, (t-2)^3\}$, and $\mathcal{C} = \{1, t, t(t-1), t(t-1)(t-2)\}$. These are three different ordered bases for $\mathbb{P}_3$.

(a) Find the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}$ and use it to write $p(t) = (t-2)^3 + 2(t-2)^2 + (t-2) + 3$ in terms of powers of t rather than powers of $t-2$. Check your answer through standard simplification.

(b) Find the change of basis matrix from $\mathcal{S}$ to $\mathcal{B}$ and use it to write $q(t) = t^3 + t + 1$ in terms of powers of $t-2$ rather than powers of t. You may recall from calculus II that this can also be done with Taylor polynomials ($q(t) = \frac{q(a)}{0!} + \frac{q'(a)}{1!}(t-a) + \frac{q''(a)}{2!}(t-a)^2 + \cdots + \frac{q^{(n)}(a)}{n!}(x-a)^n$). Check your answer using Taylor polynomials. If you have not yet taken calculus II, check your answer using standard algebraic simplification.

(c) The basis $\mathcal{C}$ is called a factorial basis. Find the change of basis matrix from $\mathcal{S}$ to $\mathcal{C}$ and use it to write $r(t) = (t+1)^3$ in terms of this factorial basis. Check your answer through standard simplification.

(d) Find the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ and use it to write $p(t) = (t-2)^3 + (t-2)^2 + (t-2) + 1$ in terms of powers of the factorial basis. Check your answer through standard simplification.

5. Let $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$.

(a) Use Theorem 5.20 to find the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ and use it to find $[\mathbf{u}]_{\mathcal{C}}$ if $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

(b) Use Theorem 5.21 to find the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$. Check your answer with part (a).

6. Repeat Exercise 5 with $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 3 \end{bmatrix} \right\}$, $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}$, and $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

7. Prove Corollary 5.14.

8. Prove Theorem 5.15.

9. Prove Theorem 5.19.

10. Given $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ from Definition 5.3 in Section 5.1, prove that $R_\theta^{-1} = R_{-\theta}$ by showing that they both equal the same 2×2 matrix.

5.4 Similarity

Suppose V and W are finite dimensional vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ respectively. In addition, suppose $T : V \longrightarrow W$ is a linear transformation. The linear transformation $T : V \longrightarrow W$ does not have a standard matrix representation if V and W are different from $\mathbb{R}^n$ and $\mathbb{R}^m$, but we see that the composition of linear transformations $T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1}$ (see Figure 5.8) maps coordinate vectors as follows: T maps $\mathbf{u}$ to $\mathbf{w}$ if and only if $T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1}$ maps $[\mathbf{u}]_\mathcal{B}$ to $[\mathbf{w}]_\mathcal{C}$.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_{\mathcal{B}}"'] & W \arrow[d, shift left, "T_{\mathcal{C}}"] \\ \mathbb{R}^n \arrow[u, shift right, "T_{\mathcal{B}}^{-1}"'] \arrow[r] & \mathbb{R}^m \arrow[u, shift left, "T_{\mathcal{C}}^{-1}"] \end{tikzcd}$$

Figure 5.8 *Visualizing the composition* $T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1}$.

Since $T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, there is an $m \times n$ matrix M such that $T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1} = T_M$. Let $S = T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1}$. Since $M = [S(\mathbf{e}_1) \cdots S(\mathbf{e}_n)]$ and through composition we see that for $j = 1, \cdots, n$,

$$\mathbf{e}_j \xrightarrow{T_\mathcal{B}^{-1}} \mathbf{v}_j \xrightarrow{T} T(\mathbf{v}_j) \xrightarrow{T_\mathcal{C}} [T(\mathbf{v}_j)]_\mathcal{C}\,.$$

So $M = [[T(\mathbf{v}_1)]_\mathcal{C} \cdots [T(\mathbf{v}_n)]_\mathcal{C}]$. This proves Theorem 5.22.

Theorem 5.22. Suppose V and W are vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ respectively. If $T : V \longrightarrow W$ is a linear transformation, then there exists an $m \times n$ matrix M such that $M\,[\mathbf{u}]_\mathcal{B} = [T(\mathbf{u})]_\mathcal{C}$ and, in fact

$$M = [[T(\mathbf{v}_1)]_\mathcal{C} \cdots [T(\mathbf{v}_n)]_\mathcal{C}]\,.$$

Though the linear transformation T may not have a matrix representation, we find, using Theorem 5.22, a matrix M acts on the coordinate vector surrogates just as T acts on the vectors in V and W.

Definition 5.16. Let $T : V \longrightarrow W$ be a linear transformation between finite dimensional vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ respectively. The $m \times n$ matrix $M = [[T(\mathbf{v}_1)]_{\mathcal{C}} \cdots [T(\mathbf{v}_n)]_{\mathcal{C}}]$ is called the **matrix representation of T relative to $\mathcal{B}$ and $\mathcal{C}$.**

Example 5.15

Because the derivative has the two properties

$$\frac{d}{dt}(f(t)+g(t)) = \frac{d}{dt}(f(t)) + \frac{d}{dt}(g(t)) \text{ and } \frac{d}{dt}(cf(t)) = c\frac{d}{dt}(f(t)),$$

the function $D : \mathbb{P}_3 \longrightarrow \mathbb{P}_2$ given by $D(f(t)) = f'(t)$ is a linear transformation. Since D acts on polynomials rather than column vectors, D is not a matrix transformation, but we can still find a matrix M to describe D through coordinate vectors.

$$\begin{tikzcd}
\mathbb{P}_3 \arrow[r, "D"] \arrow[d, shift right, "T_{\mathcal{B}}"'] & \mathbb{P}_2 \arrow[d, shift left, "T_{\mathcal{C}}"] \\
\mathbb{R}^4 \arrow[r, dotted, "M"'] \arrow[u, shift right, "T_{\mathcal{B}}^{-1}"'] & \mathbb{R}^3 \arrow[u, shift left, "T_{\mathcal{C}}^{-1}"]
\end{tikzcd}$$

Figure 5.9 *Describing D with a matrix M through the use of coordinate vectors.*

We know that $\mathcal{B} = \{1, t, t^2, t^3\}$ and $\mathcal{C} = \{1, t, t^2\}$ are ordered bases for $\mathbb{P}_3$ and $\mathbb{P}_2$ respectively. The matrix

$$\begin{aligned}
M &= \left[[D(1)]_{\mathcal{C}}\ [D(t)]_{\mathcal{C}}\ \left[D(t^2)\right]_{\mathcal{C}}\ \left[D(t^3)\right]_{\mathcal{C}}\right] \\
&= \left[[0]_{\mathcal{C}}\ [1]_{\mathcal{C}}\ [2t]_{\mathcal{C}}\ \left[3t^2\right]_{\mathcal{C}}\right] \\
&= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.
\end{aligned}$$

Differentiating polynomials is quite simple, so the use of M is not necessary, but we illustrate how M could be used to differentiate. Let $p(t) = 2t^3 - 4t^2 + 5t + 7$. Then

$$[p(t)]_{\mathcal{B}} = \begin{bmatrix} 7 \\ 5 \\ -4 \\ 2 \end{bmatrix}$$

so that

$$M[p(t)]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 7 \\ 5 \\ -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -8 \\ 6 \end{bmatrix} = \left[6t^2 - 8t + 5\right]_{\mathcal{C}}.$$

Thus, $p'(t) = 6t^2 - 8t + 5$.

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "T"] \arrow[d, shift right, "T_{\mathcal{B}}"'] & \mathbb{R}^n \arrow[d, shift left, "T_{\mathcal{B}}"] \\
\mathbb{R}^n \arrow[r, dotted] \arrow[u, shift right, "T_{\mathcal{B}}^{-1}"'] & \mathbb{R}^n \arrow[u, shift left, "T_{\mathcal{B}}^{-1}"]
\end{tikzcd}$$

(a) Taking vector spaces V and W to be $\mathbb{R}^n$ and bases $\mathcal{B}$ and $\mathcal{C}$ to be the same.

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "A"] \arrow[d, shift right, "P^{-1}"'] & \mathbb{R}^n \arrow[d, shift left, "P^{-1}"] \\
\mathbb{R}^n \arrow[r, dotted, "M"'] \arrow[u, shift right, "P"'] & \mathbb{R}^n \arrow[u, shift left, "P"]
\end{tikzcd}$$

(b) Matrices M and A represent the same linear transformation with respect to different ordered bases.

Figure 5.10

In Example 5.15 all four vector spaces in Figure 5.9 are different. That made things easy to keep straight. We are most interested, however, in applying this theory when all four vector spaces are $\mathbb{R}^n$ for some n. In addition, we want the two bases $\mathcal{B}$ and $\mathcal{C}$ to be the same (see Figure 5.10a). That is to say T and the isomorphisms $T_{\mathcal{B}}$ and $T_{\mathcal{B}}^{-1}$ are all linear operators and, in fact, they are all matrix operators.

Let A be an $n \times n$ matrix and let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be an ordered basis for $\mathbb{R}^n$. We wish to find the matrix M with the property that

$$M\,[\mathbf{u}]_{\mathcal{B}} = [\mathbf{w}]_{\mathcal{B}} \iff A\mathbf{u} = \mathbf{w} \text{ for all } \mathbf{u} \in \mathbb{R}^n.$$

Let $P = [\mathbf{v}_1 \cdots \mathbf{v}_n]$. In Figure 5.10b, we fill in the diagram found in Figure 5.10a with the matrices rather than the linear transformations. By chasing around the diagram we see that $M = P^{-1}AP$.

The matrices M and A represent the same linear transformation but with respect to different ordered bases. As set up here, M is with respect to $\mathcal{B}$ and A is with respect to the standard ordered basis $\mathcal{S}_n$ since $[\mathbf{u}]_{\mathcal{S}_n} = \mathbf{u}$.

Example 5.16

Let

$$A = \begin{bmatrix} 11 & -4 & -4 \\ -4 & 1 & 2 \\ 24 & -8 & -9 \end{bmatrix}.$$

Then $T_A : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ given by $T_A(\mathbf{x}) = A\mathbf{x}$ is a linear operator on $\mathbb{R}^3$. The set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}$$

is an ordered basis for $\mathbb{R}^3$. Find the matrix M that represents the same linear transformation T_A as A but relative to the ordered basis $\mathcal{B}$ rather than $\mathcal{S}_3$.

Solution Let

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

We identify P^{-1} through row reduction:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 2 & 4 & 1 & 0 & 0 & 1 \end{array}\right]$$
$$\downarrow$$
$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right]$$
$$\downarrow$$
$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right]$$
$$\downarrow$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 5 & -2 & -2 \\ 0 & 1 & 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right]$$

It follows that

$$\begin{aligned} M &= P^{-1}AP \\ &= \begin{bmatrix} 5 & -2 & -2 \\ -2 & 1 & 1 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 11 & -4 & -4 \\ -4 & 1 & 2 \\ 24 & -8 & -9 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \\ 2 & 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 15 & -6 & -6 \\ -2 & 1 & 1 \\ 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \\ 2 & 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \end{aligned}$$

Notice how simple M looks compared to A in Example 5.16, yet they both describe the same linear transformation but relative to different bases. It is much easier to determine how the linear transformation behaves by examining M rather than A. This did not happen by accident. The basis $\mathcal{B}$ was chosen especially to simplify the matrix A. In this section we concern ourselves with performing these changes of bases. In the next chapter we concern ourselves with choosing helpful bases.

Definition 5.17. Let A and B be $n \times n$ matrices. The matrix A is **similar** to B, denoted $A \sim B$, if there is an invertible matrix P such that $B = P^{-1}AP$.

The similarity relation between $n \times n$ matrices is an equivalence relation. That is, the similarity relation is **reflexive**, **symmetric**, and **transitive**.

Theorem 5.23. Let A, B, and C be $n \times n$ matrices.

(a) $A \sim A$. (similarity is reflexive)

(b) If $A \sim B$, then $B \sim A$. (similarity is symmetric)

(c) If $A \sim B$ and $B \sim C$, then $A \sim C$. (similarity is transitive)

Proof

(a) Since $A = I_n^{-1} A I_n$, $A \sim A$.

(b) Suppose $A \sim B$. We show $B \sim A$. Since $A \sim B$, there exists an invertible matrix P such that $B = P^{-1}AP$. But multiplying both sides of this equation on the left by P and on the right by P^{-1} yields $PBP^{-1} = A$. Let $Q = P^{-1}$. Then $Q^{-1} = P$. Substituting we get $A = Q^{-1}BQ$ so that $B \sim A$.

(c) Suppose $A \sim B$ and $B \sim C$. We show $A \sim C$. Since $A \sim B$ and $B \sim C$, there exist invertible matrices P and Q such that $B = P^{-1}AP$ and $C = Q^{-1}BQ$. Substituting $P^{-1}AP$ for B in the second equation yields $C = Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ)$. Let $R = PQ$. Then $R^{-1} = (PQ)^{-1} = Q^{-1}P^{-1}$ so that $C = R^{-1}AR$ and $A \sim C$. ∎

The fact that similarity is an equivalence relation tells us that the set of all $n \times n$ matrices breaks up into similarity classes. All matrices in the same similarity class are similar to each other and no matrix is similar to a matrix outside of its own similarity class.

If T is a linear operator on a finite dimensional vector space V, its matrix representation depends on the ordered basis chosen for V. Change the ordered basis and the square matrix that represents T must change accordingly. We have already seen how this is done if $V = \mathbb{R}^n$. In the remainder of this section we show how it is done if $V \neq \mathbb{R}^n$.

Suppose T is a linear operator on a finite-dimensional vector space V with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \cdots \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$. By Theorem 5.22, $A = [[T(\mathbf{v}_1)]_{\mathcal{B}} \cdots [T(\mathbf{v}_n)]_{\mathcal{B}}]$ and $M = [[T(\mathbf{w}_1)]_{\mathcal{C}} \cdots [T(\mathbf{w}_n)]_{\mathcal{C}}]$ are the matrix representations of T relative to $\mathcal{B}$ and relative to $\mathcal{C}$ respectively.

By Theorem 5.21, the matrix $Q = [[\mathbf{w}_1]_{\mathcal{B}} \cdots [\mathbf{w}_n]_{\mathcal{B}}]$ is the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$ and $Q^{-1} = [[\mathbf{v}_1]_{\mathcal{C}} \cdots [\mathbf{v}_n]_{\mathcal{C}}]$ is the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$.

So, for all $\mathbf{u} \in V$,

$$M\,[\mathbf{u}]_{\mathcal{C}} = [T(\mathbf{u})]_{\mathcal{C}}$$

and

$$\begin{aligned} Q^{-1}AQ\,[\mathbf{u}]_{\mathcal{C}} &= Q^{-1}A\,[\mathbf{u}]_{\mathcal{B}} \\ &= Q^{-1}\,[T(\mathbf{u})]_{\mathcal{B}} \\ &= [T(\mathbf{u})]_{\mathcal{C}}\,. \end{aligned}$$

Thus $M = Q^{-1}AQ$. This proves Theorem 5.24 and is illustrated in Example 5.17.

Theorem 5.24. Suppose T is a linear operator on a finite-dimensional vector space V with $\mathcal{B}$ and $\mathcal{C}$ ordered bases on V. Let A and M be the matrix representations of T relative to $\mathcal{B}$ and $\mathcal{C}$ respectively. Then A is similar to M and, in fact, $M = Q^{-1}AQ$ where Q is the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$.

Example 5.17

In the space of all real-valued differentiable functions of a single variable, let $V = span\{\sin t, \cos t\}$, and let $T : V \longrightarrow V$ be defined by $T(f(t)) = f(-t) - f'(t)$. Let $\mathcal{B} = \{\sin t, \cos t\}$ and $\mathcal{C} = \{\sin t + \cos t, \sin t - \cos t\}$. Find the matrix representations A and M of T relative to $\mathcal{B}$ and relative to $\mathcal{C}$ respectively plus the change of basis matrix Q from $\mathcal{C}$ to $\mathcal{B}$.

Solution $T(\sin t) = \sin(-t) - \cos t = -\sin t - \cos t$, so

$$[T(\sin t)]_{\mathcal{B}} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Similarly, $T(\cos t) = \cos(-t) + \sin t = \sin t + \cos t$, so

$$[T(\cos t)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Thus,

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

Next, we determine Q.

$$\begin{aligned} Q &= [[\sin t + \cos t]_{\mathcal{B}} \, [\sin t - \cos t]_{\mathcal{B}}] \\ &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \end{aligned}$$

so

$$Q^{-1} = -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Finally,

$$\begin{aligned} M &= Q^{-1}AQ \\ &= -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 0 & -2 \\ 0 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

We can check our work by calculating $M = [[T(\sin t + \cos t)]_{\mathcal{C}} \, [T(\sin t - \cos t)]_{\mathcal{C}}]$ directly. Since

$$\begin{aligned} T(\sin t + \cos t) &= (\sin(-t) + \cos(-t)) - (\cos t - \sin t) \\ &= -\sin t + \cos t - \cos t + \sin t \\ &= 0 \end{aligned}$$

we see that
$$[T(\sin t+\cos t)]_{\mathcal{C}}=\begin{bmatrix}0\\0\end{bmatrix}.$$

And,
$$\begin{aligned}T(\sin t-\cos t) &= (\sin(-t)-\cos(-t))-(\cos t+\sin t)\\ &= -\sin t-\cos t-\cos t-\sin t\\ &= -2(\sin t+\cos t)\end{aligned}$$

so
$$[T(\sin t-\cos t)]_{\mathcal{C}}=\begin{bmatrix}-2\\0\end{bmatrix}$$

and
$$M=\begin{bmatrix}0&-2\\0&0\end{bmatrix}.$$

Problem Set 5.4

1. Let $A=\begin{bmatrix}1&0&-1\\0&1&-1\end{bmatrix}$ and let $T:\mathbb{R}^3\to\mathbb{R}^2$ be defined by $T(\mathbf{x})=A\mathbf{x}$. The sets $\mathcal{B}=\left\{\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}-1\\1\\0\end{bmatrix},\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$ and $\mathcal{C}=\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}-1\\1\end{bmatrix}\right\}$ are ordered bases for $\mathbb{R}^3$ and $\mathbb{R}^2$ respectively. Find the matrix representation of T relative to $\mathcal{B}$ and $\mathcal{C}$.

2. Let $T:\mathbb{P}_2\to\mathbb{P}_1$ be defined by $T(f(t))=f(0)+tf(1)$. Find A, the matrix representation of T relative to the ordered bases $\mathcal{S}_2=\left\{1,t,t^2\right\}$ and $\mathcal{S}_1=\{1,t\}$, and find M, the matrix representation of T relative to the ordered bases $\mathcal{B}=\left\{t^2-t,t,1\right\}$ and $\mathcal{C}=\{t,1\}$

3. Let each $n\times n$ matrix be the standard matrix, A, for the linear operator $T(\mathbf{x})=A\mathbf{x}$. In each case, find the matrix representation of T relative to the given ordered basis $\mathcal{B}$.

(a) $A=\begin{bmatrix}0&1\\1&0\end{bmatrix}$, $\mathcal{B}=\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}-1\\1\end{bmatrix}\right\}$

(b) $A=\begin{bmatrix}3&-2\\1&0\end{bmatrix}$, $\mathcal{B}=\left\{\begin{bmatrix}2\\1\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix}\right\}$

(c) $A=\begin{bmatrix}3&-6\\1&-2\end{bmatrix}$, $\mathcal{B}=\left\{\begin{bmatrix}3\\1\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix}\right\}$

(d) $A=\begin{bmatrix}1&-2&2\\0&-2&3\\0&-2&3\end{bmatrix}$, $\mathcal{B}=\left\{\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}1\\2\\2\end{bmatrix},\begin{bmatrix}2\\3\\2\end{bmatrix}\right\}$

(e) $A = \begin{bmatrix} -1 & 4 & 2 \\ 0 & 2 & 0 \\ -1 & 1 & 2 \end{bmatrix}$, $\mathcal{B} = \left\{ \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\}$

4. Let T be the linear operator on $\mathbb{P}_2$ defined by $T(f(t)) = (t+1)f'(t)$. Let $\mathcal{B}$ and $\mathcal{C}$ be the ordered bases $\{1, t, t^2\}$ and $\{1, t+1, (t+1)^2\}$ respectively. Let Q be the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$, let A be the matrix representation of T relative to $\mathcal{B}$ and M the matrix representation of T relative to $\mathcal{C}$. Find Q, A, and M.

5. Let V be the span of $\mathcal{B} = \{e^t, e^{-t}\}$ in the space of continuous functions, and suppose $T(f(t)) = f(t+1) + f(-t)$. Let $\mathcal{C} = \{e^t - ee^{-t}, ee^t + e^{-t}\}$. Find Q, A, and M where Q is the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$, A is the matrix representation of T relative to $\mathcal{B}$, and M is the matrix representation of T relative to $\mathcal{C}$.

6. Suppose A and B are similar $n \times n$ matrices. Prove each of the following:

(a) $\det A = \det B$.

(b) A^T and B^T are similar.

(c) For each positive integer k, A^k and B^k are similar.

(d) If A is invertible, then A^{-1} and B^{-1} are similar.

7. Suppose $A = PQ$ where P and Q are square and P is invertible. Let $B = QP$. Prove A and B are similar.

8. Suppose A and B are similar $n \times n$ matrices. Prove that there exist $n \times n$ matrices P and Q in which P is invertible, $A = PQ$, and $B = QP$.

6 Eigenvalues and Eigenvectors

6.1 Eigenvalues, Eigenvectors, and Eigenspaces

In this chapter we learn how to analyze operators $\mathbf{x} \longrightarrow A\mathbf{x}$, that is, matrix transformations where A is a square matrix. This analysis has important applications all through the sciences, engineering, and mathematics.

Let A be an $n \times n$ matrix. Though an operator $\mathbf{x} \longrightarrow A\mathbf{x}$ may move vectors in a variety of directions, oftentimes it happens that the matrix operator acts on certain nonzero vectors in very simple ways. Example 6.1 illustrates.

Example 6.1

Let

$$A = \begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

Notice that

$$A\mathbf{v} = \begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -4 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = 2\mathbf{v}.$$

Definition 6.1. Let A be an $n \times n$ matrix. If there is a nonzero vector $\mathbf{v}$ and a scalar λ such that $A\mathbf{v} = \lambda\mathbf{v}$, then λ is called an **eigenvalue** of A and the nonzero vector $\mathbf{v}$ is called an **eigenvector** of A associated with the eigenvalue λ.

Example 6.1 shows that 2 is an eigenvalue of A and $\begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$ is an eigenvector of A associated with the eigenvalue 2. Are there any other eigenvectors of A associated with the eigenvalue 2 in this example?

To answer this question we must simply solve a linear system

$$\begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 2 \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

which can also be written as

$$\begin{array}{rcrcrcr} -4x & + & 3y & - & 3z & = & 2x \\ & - & y & & & = & 2y \\ 6x & - & 6y & + & 5z & = & 2z. \end{array}$$

Putting this in standard form, we see it as a homogeneous system

$$\begin{array}{rcrcrcl} (-4-2)x & + & 3y & - & 3z & = & 0 \\ & & (-1-2)y & & & = & 0 \\ 6x & - & 6y & + & (5-2)z & = & 0. \end{array}$$

Row reduction

$$\left[\begin{array}{rrr|r} -6 & 3 & -3 & 0 \\ 0 & -3 & 0 & 0 \\ 6 & -6 & 3 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{rrr|r} 2 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{rrr|r} 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

and setting $z = t$ yields the parameterized solution $x = -1/2t$, $y = 0$, and $z = t$. That is,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix}.$$

We see that, in fact, any nonzero multiple of $\begin{bmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix}$ is an eigenvector of A associated with the eigenvalue $\lambda = 2$.

Note that even though $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix}$ is a solution to the system and $A\mathbf{0} = 2\mathbf{0}$ the zero vector $\mathbf{0}$ is *not* an eigenvector because the definition of eigenvector excludes it.

Note too that the solution set

$$S = span \left\{ \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a subspace. This is no accident because it is the solution set of a homogeneous system. The system $A\mathbf{x} = \lambda\mathbf{x}$ doesn't look homogeneous at first, but once all variables are put on the left-hand side in standard form we see it is. In general,

$$\begin{gathered} A\mathbf{x} = \lambda\mathbf{x} \\ \Updownarrow \\ A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \\ \Updownarrow \\ A\mathbf{x} - \lambda I_n\mathbf{x} = \mathbf{0} \\ \Updownarrow \\ (A - \lambda I_n)\mathbf{x} = \mathbf{0}. \end{gathered}$$

Note that it is incorrect to write $(A - \lambda)\mathbf{x} = \mathbf{0}$ since A is a matrix and λ is a scalar making subtraction impossible in general. In this particular instance, the homogeneous system $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$ can be written as

$$\left(\begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix} - 2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{i.e.} \begin{bmatrix} -6 & 3 & -3 \\ 0 & -3 & 0 \\ 6 & -6 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

In determining eigenvectors, you can skip right to the homogeneous equation $(A-\lambda I_n)\mathbf{x} = \mathbf{0}$.

Definition 6.2. Let A be an $n \times n$ matrix. If λ is an eigenvalue of A, then the solution set to the homogeneous system $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$ is called the **eigenspace** of A associated with the eigenvalue λ and is denoted E_λ.

Put another way, the eigenspace of A associated with the eigenvalue λ is the null space of $A - \lambda I_n$. It is the set of all eigenvectors of A associated with the eigenvalue λ plus the zero vector.

It is easy to determine whether a vector $\mathbf{v}$ is an eigenvector of A - just check whether $A\mathbf{v}$ is a multiple of $\mathbf{v}$.

Example 6.2

Let

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

Is $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ an eigenvector of A?

Solution

$$A\mathbf{v}_1 = \begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ -1 \\ 5 \end{bmatrix}$$

No! The product $A\mathbf{v}_1$ is not a multiple of $\mathbf{v}_1$.

How about $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$?

$$A\mathbf{v}_2 = \begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix} = -1\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = -1\mathbf{v}_2$$

Yes! $\mathbf{v}_2$ is an eigenvector of A associated with $\lambda = -1$.

We can find the whole eigenspace associated with $\lambda = -1$ by solving $(A - (-1)I_3)\mathbf{x} = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} -3 & 3 & -3 & 0 \\ 0 & 0 & 0 & 0 \\ 6 & -6 & 6 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Letting $y = s$ and $z = t$ gives the parametric solution

$$\begin{array}{rcrcr} x & = & s & - & t \\ y & = & s & & \\ z & = & & & t. \end{array}$$

That is,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = s\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

So

$$E_{-1} = span\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}.$$

The eigenspace associated with $\lambda = -1$ is a 2-dimensional subspace of $\mathbb{R}^3$.

Is -5 an eigenvalue of A? We solve $(A + 5I_3)\mathbf{x} = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} 1 & 3 & -3 & 0 \\ 0 & 4 & 0 & 0 \\ 6 & -6 & 10 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 3 & -3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -24 & 28 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 3 & -3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 28 & 0 \end{array}\right]$$

Each column of $A + 5I_3$ is a pivot column making $\mathbf{x} = \mathbf{0}$ the only solution. Thus -5 is not an eigenvalue.

The moral of Example 6.2 is that it is easy to check whether a vector $\mathbf{v}$ is an eigenvector of A. Simply check to see whether $A\mathbf{v}$ is a multiple of $\mathbf{v}$. It is also easy to check whether λ is an eigenvalue of A. Simply solve $(A - \lambda I)\mathbf{x} = \mathbf{0}$. If it has nontrivial solutions, then λ is an eigenvalue. If it has only the trivial solution, then λ is not an eigenvalue of A.

So far it appears that we made some good guesses to find $\lambda = 2, -1$ in Example 6.2. But we made some bad guesses too (like -5). Are there any other eigenvalues of A? We won't know until we find a process for determining all eigenvalues.

It turns out it is easy to find eigenvalues of triangular matrices.

Theorem 6.1. The eigenvalues of triangular matrices are the entries on the main diagonal.

Proof We prove this theorem for upper triangular matrices. The same argument works for lower triangular matrices. Suppose that A is upper triangular. Then λ is an eigenvalue of A if and only if $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has nontrivial solutions which occurs if and only if $A - \lambda I$ is singular. But $A - \lambda I$ is upper triangular, so $A - \lambda I$ is singular if and only if $A - \lambda I$ has a zero on the main diagonal.

$$A - \lambda I = \begin{bmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix} - \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = \begin{bmatrix} (a_{11} - \lambda_1) & & * \\ & \ddots & \\ 0 & & (a_{nn} - \lambda_n) \end{bmatrix}$$

$A - \lambda I$ has a zero on the main diagonal if and only if $\lambda = a_{jj}$ for some j. Therefore, λ is an eigenvalue of A if and only if $\lambda = a_{jj}$ for some j. ■

Example 6.3

Let

$$A = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 0 & 2 \\ 0 & 0 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 5 & 0 \\ 1 & -1 & 7 \end{bmatrix}.$$

The eigenvalues of A are $\lambda = 1, 0, 3$. The eigenvalues of B are $\lambda = 5, 7$. In each case we could solve to find the eigenspaces as we did in Example 6.2.

In Example 6.3 the matrix A had an eigenvalue of 0. An eigenvalue of 0 is a case that deserves some special attention.

Theorem 6.2. Let A be an $n \times n$ matrix. A has an eigenvalue of 0 if and only if A is singular.

Proof

$$\begin{aligned} \text{The matrix } A \text{ has an eigenvalue of } 0 \quad &\Longleftrightarrow \quad (A - 0I)\mathbf{x} = \mathbf{0} \text{ has nontrivial solutions.} \\ &\Longleftrightarrow \quad A\mathbf{x} = \mathbf{0} \text{ has nontrivial solutions.} \\ &\Longleftrightarrow \quad A \text{ is singular.} \end{aligned}$$

■

Reworded, Theorem 6.2 says A is invertible if and only if 0 is not an eigenvalue of A. We add this to our list of equivalent descriptions of invertible matrices at the end of this section. Here is one more theorem for later use.

Theorem 6.3. If $\mathbf{v}_1, \cdots, \mathbf{v}_r$ are eigenvectors that correspond to *distinct* eigenvalues $\lambda_1, \cdots, \lambda_r$ of an $n \times n$ matrix A, then $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is linearly independent.

Proof We prove by contradiction and suppose not. That is, suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is linearly dependent. Since $\mathbf{v}_1$ is an eigenvector, $\mathbf{v}_1 \neq \mathbf{0}$, the set $\{\mathbf{v}_1\}$ is linearly independent. Let p be the smallest index such that $\mathbf{v}_{p+1} \in \mathit{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$. Then $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is linearly independent and $\{\mathbf{v}_1, \cdots, \mathbf{v}_{p+1}\}$ is linearly dependent. So, there exist $c_1, \cdots, c_p$ such that

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p. \tag{6.1}$$

Note that not all $c_1, \cdots, c_p$ are 0 since $\mathbf{v}_{p+1}$ is an eigenvector. Multiplying both sides of (6.1) by λ_{p+1} gives

$$\lambda_{p+1}\mathbf{v}_{p+1} = \lambda_{p+1}c_1\mathbf{v}_1 + \cdots + \lambda_{p+1}c_p\mathbf{v}_p. \tag{6.2}$$

Now multiply both sides of (6.1) on the left by A.

$$\begin{aligned} A\mathbf{v}_{p+1} &= Ac_1\mathbf{v}_1 + \cdots + Ac_p\mathbf{v}_p \\ &= c_1A\mathbf{v}_1 + \cdots + c_pA\mathbf{v}_p \end{aligned}$$

Since $\mathbf{v}_1, \cdots, \mathbf{v}_{p+1}$ are eigenvectors and $A\mathbf{v}_j = \lambda_j\mathbf{v}_j$ for $j = 1, \cdots, p+1$,

$$\lambda_{p+1}\mathbf{v}_{p+1} = c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p \tag{6.3}$$

Equating (6.2) and (6.3) gives

$$c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p = c_1\lambda_{p+1}\mathbf{v}_1 + \cdots + c_p\lambda_{p+1}\mathbf{v}_p.$$

Bring everything to the left-hand side and combine like terms.

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = \mathbf{0}$$

But, because the eigenvalues are all distinct, $\lambda_j - \lambda_{p+1} \neq 0$ for $j = 1, \cdots, p$ and because not all $c_1, \cdots, c_p$ are 0, we know that the scalars $c_1(\lambda_1 - \lambda_{p+1}), \cdots, c_p(\lambda_p - \lambda_{p+1})$ are not all 0. Thus $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is linearly dependent. This contradicts the fact that $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is linearly independent. So our assumption that $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is linearly dependent must be false. Therefore $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is linearly independent. ■

We add now to Theorem 5.12 which was last extended in section 5.2.

Theorem 6.4. Let A be an $n \times n$ matrix. The following are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) The reduced row-echelon form of A is the identity matrix I_n.

(d) A is a product of elementary matrices.

(e) A has n pivot columns.

(f) A has a left inverse.

(g) A has a right inverse.

(h) For all $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(i) Every n-vector $\mathbf{b}$ is a linear combination of the columns of A.

(j) A^T is invertible.

(k) $\mathit{rank}\ A = n$.

(l) $\mathit{nullity}\ A = 0$.

(m) $\det A \neq 0$.

(n) The columns (or rows) of A span $\mathbb{R}^n$.

(o) The columns (or rows) of A are linearly independent.

(p) The columns (or rows) of A form a basis for $\mathbb{R}^n$.

(q) $\mathit{col}\ A = \mathbb{R}^n$.

(r) $\mathit{row}\ A = \mathbb{R}^n$.

(s) $\mathit{null}\ A = \{\mathbf{0}\}$.

(t) $\mathit{range}\ T_A = \mathbb{R}^n$.

(u) $\mathit{ker}\ T_A = \{\mathbf{0}\}$.

(v) T_A is a surjection.

(w) T_A is an injection.

(x) T_A is a bijection.

(y) T_A is an isomorphism.

(z) 0 is not an eigenvalue of A.

Problem Set 6.1

In each part of Exercises 1-4 determine whether the given vector is an eigenvector of the given matrix A. If so, find the eigenvalue of A it is associated with and a basis for the associated eigenspace.

1. $A = \begin{bmatrix} 8 & -18 \\ 3 & -7 \end{bmatrix}$.

(a) $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ **(b)** $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ **(c)** $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

2. $A = \begin{bmatrix} -7 & 25 \\ -4 & 13 \end{bmatrix}$.

(a) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ **(b)** $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ **(c)** $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

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

(a) $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ **(b)** $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ **(c)** $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ **(d)** $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$

4. $A = \begin{bmatrix} -1 & 4 & -4 \\ 0 & 7 & -8 \\ 0 & 4 & -5 \end{bmatrix}$.

(a) $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ **(b)** $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ **(c)** $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ **(d)** $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$

In each part of Exercises 5 & 6 determine whether the given scalar is an eigenvalue of the given matrix A. If so, find a basis for the associated eigenspace.

5. $A = \begin{bmatrix} -11 & 18 \\ -6 & 10 \end{bmatrix}$.

(a) -2 **(b)** 2 **(c)** 1

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

(a) 1 **(b)** 2 **(c)** 3 **(d)** 4

7. Find the eigenvalues and bases for the associated eigenspaces of each of the following triangular matrices.

(a) $\begin{bmatrix} 4 & 1 \\ 0 & 3 \end{bmatrix}$ **(b)** $\begin{bmatrix} 1 & 0 \\ 1 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} -1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ **(d)** $\begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

8. True or false. Let A represent an $n \times n$ matrix.

(a) If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$, then λ is an eigenvalue of A.

(b) A matrix A is invertible if and only if 0 is not an eigenvalue of A.

(c) A scalar t is an eigenvalue of A if and only if the homogeneous matrix equation $(A - tI)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

(d) If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A.

(e) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent eigenvectors of A, then they are associated with distinct eigenvalues.

(f) The eigenvalues of a matrix are on its main diagonal.

(g) An eigenspace of a matrix is the null space of a related matrix.

9. Let A be an $n \times n$ matrix in which the row sums all equal the same scalar s. Show that s is an eigenvalue of A by finding an eigenvector associated with s.

10. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and let A be the 2×2 projection matrix with the property that for all $\mathbf{x} \in \mathbb{R}^2$, $A\mathbf{x} = proj_{\mathbf{u}}\mathbf{x}$ where $proj_{\mathbf{u}}\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{u}$ (see section 2.2).

(a) What are the eigenvalues of A?

(b) Describe geometrically the eigenspaces of A associated with each of the eigenvalues.

11. Let B be the 2×2 reflection matrix with the property that for all $\mathbf{x} \in \mathbb{R}^2$, $B\mathbf{x}$ is the reflection of $\mathbf{x}$ across the line $y = 2x$ (see Problem Set 5.1, Exercise 4).

(a) What are the eigenvalues of B?

(b) Describe geometrically the eigenspaces of B associated with each of the eigenvalues.

12. There are only two values of θ, for $0 \leq \theta < 2\pi$, for which there are real numbers λ that are eigenvalues for the 2×2 rotation matrix R_θ (see Section 5.1). What are those two values of θ and what are the corresponding eigenvalues for those values of θ? Give a geometric argument for why R_θ does not have real eigenvalues for other values of θ.

13. Let A be the 3×3 projection matrix with the property that for all $\mathbf{x} \in \mathbb{R}^3$, $A\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto the plane $x + 2y + 3z = 0$

(a) What are the eigenvalues of A?

(b) Describe geometrically the eigenspaces of A associated with each of the eigenvalues.

14. Let λ be an eigenvalue of a matrix A, and let $\mathbf{v}$ be an associated eigenvector. Show that for each positive integer n, λ^n is an eigenvalue of A^n and $\mathbf{v}$ is an associated eigenvector.

15. Let λ be an eigenvalue of an invertible matrix A. Show that λ^{-1} is an eigenvalue of A^{-1}.

6.2 The Characteristic Polynomial

Let A be an $n \times n$ matrix. From section 6.1 we see that if we know or can guess an eigenvalue or eigenvector, then it is easy to check and it is easy to find the entire eigenspace by solving $(A-\lambda I)\mathbf{x} = \mathbf{0}$. If a guess of the eigenvalue is wrong, then $(A-\lambda I)\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Also from section 6.1 we know that if A happens to be triangular then the eigenvalues appear on the main diagonal and solving $(A-\lambda I)\mathbf{x} = \mathbf{0}$ for various λ along the diagonal produces the eigenspaces.

For the moment then, if A is not a diagonal matrix, we are reduced to guess and check because both λ and $\mathbf{x}$ are unknown in the equation $(A-\lambda I)\mathbf{x} = \mathbf{0}$. We need to eliminate an unknown.

A correct eigenvalue of A is chosen for λ if and only if the homogeneous system $(A-\lambda I)\mathbf{x} = \mathbf{0}$ has nontrivial solutions. But $(A-\lambda I)\mathbf{x} = \mathbf{0}$ has nontrivial solutions if and only if $A-\lambda I$ is singular. In chapter 3 we learned that a matrix is singular if and only if its determinant is 0. So λ is an eigenvalue of A if and only if $\det(A-\lambda I) = 0$.

The nice thing is that the equation $\det(A-\lambda I) = 0$ has only one unknown, λ. This proves Theorem 6.5.

Theorem 6.5. Suppose A is an $n \times n$ matrix. The scalar λ is an eigenvalue of A if and only if $\det(A-\lambda I) = 0$.

Let's look at the matrix from Example 6.1 in section 6.1.

Example 6.4

Let
$$A = \begin{bmatrix} -4 & 3 & -3 \\ 0 & -1 & 0 \\ 6 & -6 & 5 \end{bmatrix}.$$
Find the eigenvalues of A.

Solution

$$A - \lambda I = \begin{bmatrix} -4-\lambda & 3 & -3 \\ 0 & -1-\lambda & 0 \\ 6 & -6 & 5-\lambda \end{bmatrix}.$$

By expanding $\det(A-\lambda I)$ across the second row,

$$\begin{aligned}
\det(A-\lambda I) &= \begin{vmatrix} -4-\lambda & 3 & -3 \\ 0 & -1-\lambda & 0 \\ 6 & -6 & 5-\lambda \end{vmatrix} \\
&= (-1-\lambda)\begin{vmatrix} -4-\lambda & -3 \\ 6 & 5-\lambda \end{vmatrix} \\
&= -(\lambda+1)\left[(-4-\lambda)(5-\lambda)-(-3)(6)\right] \\
&= -(\lambda+1)\left[-20-\lambda+\lambda^2+18\right] \\
&= -(\lambda+1)(\lambda^2-\lambda-2) \\
&= -(\lambda+1)(\lambda+1)(\lambda-2) \\
&= -(\lambda+1)^2(\lambda-2).
\end{aligned}$$

So, the eigenvalues of A are -1 and 2. This agrees with what we found in section 6.1 and it shows that A has no other eigenvalues because $\lambda=-1,2$ are the only roots of the equation $\det(A-\lambda I)=0$.

We could go on to find the eigenspaces of A by solving the systems $(A-(-1)I)\mathbf{x}=\mathbf{0}$ and $(A-2I)\mathbf{x}=\mathbf{0}$, but that was done in section 6.1. There we found

$$E_{-1} = span\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\} \text{ and } E_2 = span\left\{\begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}\right\}.$$

Definition 6.3. Let A be an $n\times n$ matrix. Then $\det(A-\lambda I)$ is called the **characteristic polynomial** of A. The equation $\det(A-\lambda I)=0$ is called the **characteristic equation** of A.

Theorem 6.5 tells us that the eigenvalues of a square matrix are the roots of its characteristic equation or the zeros of its characteristic polynomial.

It is not immediately clear that $\det(A-\lambda I)$ is a polynomial. It doesn't look much like a polynomial in that form, but by simply looking at Example 6.4 we see that indeed $\det(A-\lambda I)$ is a polynomial in the variable λ.

It is easy to see by induction that if A is an $n\times n$ matrix, then the degree of its characteristic polynomial is n. This means that finding eigenvalues of 2×2 matrices is relatively easy using the quadratic formula, but for $n\geq 3$ it involves factoring the characteristic polynomial. That can be difficult. In Example 6.4 we avoided the difficulty by expanding the determinant across the second row. Some higher degree polynomials are easily factored, but not all. Software like *Maple* can handle polynomial equations up to degree 4, but for degree 5 and higher, even software packages like *Maple* can fail to factor.

Example 6.5

Find the eigenvalues and eigenspaces of

$$A = \begin{bmatrix} -1 & 3 & -2 \\ -6 & 8 & -6 \\ -6 & 6 & -5 \end{bmatrix}.$$

Solution First, we determine the eigenvalues. The basket weave technique for computing the determinant of a 3×3 matrix comes in handy here.

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} -1-\lambda & 3 & -2 \\ -6 & 8-\lambda & -6 \\ -6 & 6 & -5-\lambda \end{vmatrix} \\
&= (-1-\lambda)(8-\lambda)(-5-\lambda) + 108 + 72 - 12(8-\lambda) + 18(-5-\lambda) + 36(-1-\lambda) \\
&= (\lambda+1)(\lambda^2 - 3\lambda - 40) + 180 - 96 + 12\lambda - 90 - 18\lambda - 36 - 36\lambda \\
&= -\lambda^3 + 3\lambda^2 + 40\lambda - \lambda^2 + 3\lambda + 40 - 42\lambda - 42 \\
&= -\lambda^3 + 2\lambda^2 + \lambda - 2 \\
&= -\lambda^2(\lambda - 2) + 1(\lambda - 2) \quad \text{(factor by grouping)} \\
&= -(\lambda - 2)(\lambda^2 - 1) \\
&= -(\lambda - 2)(\lambda + 1)(\lambda - 1)
\end{aligned}$$

so that the eigenvalues are $\lambda = 2, 1$, and -1.

Next, we determine the eigenspaces.

▶ For $\lambda = 2$, we solve $(A - 2I)\mathbf{x} = \mathbf{0}$. To that end, we row reduce $A - 2I$:

$$\begin{bmatrix} -3 & 3 & -2 \\ -6 & 6 & -6 \\ -6 & 6 & -7 \end{bmatrix} \longrightarrow \begin{bmatrix} -3 & 3 & -2 \\ 0 & 0 & -2 \\ 0 & 0 & -3 \end{bmatrix} \longrightarrow \begin{bmatrix} -3 & 3 & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} -3 & 3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Letting $y = t$ gives $x = t, y = t$, and $z = 0$. That is, we can describe the solutions by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

so that

$$E_2 = span \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

▶ For $\lambda = 1$, we solve $(A - 1I)\mathbf{x} = \mathbf{0}$. Again, row reduce $A - 1I$:

$$\begin{bmatrix} -2 & 3 & -2 \\ -6 & 7 & -6 \\ -6 & 6 & -6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 1 \\ -6 & 7 & -6 \\ -2 & 3 & -2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Letting $z = t$ gives $x = -t, y = 0$, and $z = t$. So we can describe the solutions by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

so that

$$E_1 = span\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right\}.$$

▶ For $\lambda = -1$, we solve $(A-(-1)I)\mathbf{x} = \mathbf{0}$. As before, row reduce $A-(-1)I$:

$$\begin{bmatrix} 0 & 3 & -2 \\ -6 & 9 & -6 \\ -6 & 6 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 6 & -9 & 6 \\ 0 & 3 & -2 \\ -6 & 6 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 6 & -9 & 6 \\ 0 & 3 & -2 \\ 0 & -3 & 2 \end{bmatrix} \longrightarrow$$

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

Letting $z = t$ gives $x = 0, y = \frac{2}{3}t$, and $z = t$. So we can describe the solutions by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t\begin{bmatrix} 0 \\ \frac{2}{3} \\ 1 \end{bmatrix}$$

so that

$$E_{-1} = span\left\{\begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}\right\}.$$

Factoring the characteristic polynomial by grouping made it easy to find the eigenvalues in Example 6.5, but factoring by grouping doesn't work on all polynomials. You learned several tricks for factoring polynomials in precalculus. They can all be employed for finding eigenvalues. There are other methods for finding eigenvalues. Some are numerical methods that approximate eigenvalues and eigenvectors. They are important in applications but they are beyond the scope of an introductory linear algebra course.

As the next theorem demonstrates, square matrices that are similar (similar as defined in section 5.4) have the same characteristic polynomial. This tells us that similar matrices must have exactly the same eigenvalues.

Theorem 6.6. If A and B are similar $n \times n$ matrices, then they have the same characteristic polynomial.

Proof Suppose A is similar to B. Then, there exists an invertible matrix P such that $B = P^{-1}AP$. The characteristic polynomial of B is $\det(B-\lambda I)$ so

$$\begin{aligned} \det(B-\lambda I) &= \det(P^{-1}AP - \lambda P^{-1}IP) \\ &= \det(P^{-1}AP - P^{-1}(\lambda I)P) \\ &= \det\left[P^{-1}(A-\lambda I)P\right] \\ &= (\det P^{-1})\det(A-\lambda I)(\det P) \\ &= \det(A-\lambda I) \end{aligned}$$

since $\det P^{-1} = \frac{1}{\det P}$. ■

Corollary 6.7. If A and B are similar $n \times n$ matrices, then A and B have exactly the same eigenvalues.

Proof Since A and B have the same characteristic polynomial, the zeros of their characteristic polynomials must be the same. But the zeros of the characteristic polynomials are the eigenvalues. So A and B must have the same eigenvalues. ■

Problem Set 6.2

1. Find the characteristic polynomial and eigenvalues of each of the following.

(a) $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ **(b)** $\begin{bmatrix} 3 & 3 \\ -1 & -1 \end{bmatrix}$ **(c)** $\begin{bmatrix} 0 & 4 \\ -1 & 4 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ **(e)** $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ **(f)** $\begin{bmatrix} 6 & -8 & -8 \\ 3 & -5 & -6 \\ 1 & -1 & 0 \end{bmatrix}$

(g) $\begin{bmatrix} -4 & 4 & -4 \\ 1 & -1 & 2 \\ 3 & -3 & 4 \end{bmatrix}$ **(h)** $\begin{bmatrix} 4 & -2 & -1 \\ 7 & -5 & -1 \\ -5 & 6 & 0 \end{bmatrix}$ **(i)** $\begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \\ -8 & 8 & 1 \end{bmatrix}$

2. Find bases for the eigenspaces of the matrices in Exercise 1.

3. Use the definition of characteristic polynomial and properties of transpose to show that a matrix and its transpose have the same characteristic polynomial.

4. Note that for any polynomial $p(t) = a_n t^n + \cdots + a_1 t + a_0$, the constant term of the polynomial $a_0 = p(0)$. Use this fact and the definition of characteristic polynomial to show that the constant term of the characteristic polynomial of a square matrix A equals $\det A$.

For an $n \times n$ matrix A, the **trace** of A equals the sum of its diagonal entries. That is, $tr(A) = a_{11} + \cdots + a_{nn}$.

5. Let A be a 2×2 matrix. Show that the characteristic polynomial of A is $p(\lambda) = \lambda^2 - tr(A)\lambda + \det(A)$ where $tr(A)$ is the trace of A and $\det(A)$ is the determinant of A. Use this to check your results in Exercise 1.

6. Let λ_1 and λ_2 be the eigenvalues of a 2×2 matrix A. *(If A has just one eigenvalue take $\lambda_1 = \lambda_2$.)* Use Exercise 5 to show $tr(A) = \lambda_1 + \lambda_2$ and $\det(A) = \lambda_1\lambda_2$.

7. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and suppose λ_1 is an eigenvalue of A. Use Exercise 6 to show that if $b \neq 0$ or $\lambda_1 \neq a$, then $\begin{bmatrix} -b \\ a - \lambda_1 \end{bmatrix}$ is an eigenvector of A associated with the eigenvalue λ_1. Find an eigenvector associated with λ_1 if $b = 0$ and $\lambda_1 = a$. *(Consider two cases, $c \neq 0$ and $c = 0$.)*

8. True or false.

(a) If $\lambda + 3$ is a factor of the characteristic polynomial of A, then 3 is an eigenvalue of A.

(b) The row replacement elementary row operation does not change the determinant of a square matrix.

(c) The row replacement elementary row operation does not change the characteristic polynomial of a square matrix.

(d) The row replacement elementary row operation does not change the eigenvalues of a square matrix.

(e) A square matrix and its transpose have the same eigenvalues.

(f) A square matrix and its transpose have the same eigenvectors.

9. Let A be an $n \times n$ matrix in which the column sums all equal the same scalar s. Use Exercise 3 from this section and Exercise 9 from Section 6.1 to show that s is an eigenvalue of A.

6.3 Diagonalization

Diagonal matrices are the easiest to analyze in terms of eigenvalues and eigenvectors. Since diagonal matrices are triangular, the entries on their main diagonal are their eigenvalues. If

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix},$$

then the eigenvectors of D are also easy to spot. Since $D\mathbf{e}_j = \lambda_j \mathbf{e}_j$, $\mathbf{e}_j$ is an eigenvector of D corresponding to the eigenvalue λ_j. So for any $n \times n$ diagonal matrix, the standard basis $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is a basis of eigenvectors.

Matrices that are similar to a diagonal matrix D have the same characteristic polynomial and eigenvalues as D but $\mathcal{S}_n$ need not be a basis of eigenvectors. However, we will see that the matrices that are similar to diagonal matrices are precisely those that have *some* basis of eigenvectors.

Definition 6.4. An $n \times n$ matrix A is **diagonalizable** if A is similar to a diagonal matrix. That is, A is diagonalizable if there is a diagonal matrix D and an invertible matrix P such that $D = P^{-1}AP$. In this case we say P **diagonalizes** A to D.

Theorem 6.8. Let A be an $n \times n$ matrix. The matrix A is diagonalizable if and only if $\mathbb{R}^n$ has a basis of eigenvectors of A.

Proof Throughout this proof, it is useful to reference the diagram in Figure 6.1a. It will help you keep track of the location of each vector.

To begin, suppose $\mathbb{R}^n$ has a basis of eigenvectors of A. Call it $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. Let λ_i be the eigenvalue associated with the eigenvector $\mathbf{v}_i$. Note that the list $\lambda_1, \cdots, \lambda_n$ may have repeated values. Let $P = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$. We show that

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Recall that P and P^{-1} are change of basis matrices and that $P^{-1}\mathbf{w} = [\mathbf{w}]_{\mathcal{B}}$ and $P[\mathbf{w}]_{\mathcal{B}} = \mathbf{w}$ for all $\mathbf{w} \in \mathbb{R}^n$ so

$$\begin{aligned} (P^{-1}AP)\mathbf{e}_j &= (P^{-1}A)(P\mathbf{e}_j) \\ &= (P^{-1}A)P[\mathbf{v}_j]_{\mathcal{B}} \\ &= (P^{-1}A)\mathbf{v}_j \\ &= P^{-1}(A\mathbf{v}_j) \\ &= P^{-1}(\lambda_j\mathbf{v}_j) \\ &= \lambda_j(P^{-1}\mathbf{v}_j) \\ &= \lambda_j[\mathbf{v}_j]_{\mathcal{B}} \\ &= \lambda_j\mathbf{e}_j \end{aligned}$$

for $j = 1, \cdots, n$. But $(P^{-1}AP)\mathbf{e}_1, \cdots, (P^{-1}AP)\mathbf{e}_n$ are the columns of $P^{-1}AP$ so that

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

To prove the other direction, suppose A is diagonalizable. We prove that $\mathbb{R}^n$ has a basis of eigenvectors of A. Since A is diagonalizable, there exists a diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

and an invertible matrix P such that $D = P^{-1}AP$. Since P is invertible, the columns of P form a basis for $\mathbb{R}^n$. Suppose $P = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$ and let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. We show that $\mathcal{B}$ is a basis of eigenvectors of A.

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "A"] \arrow[d, shift right, "P^{-1}"'] & \mathbb{R}^n \arrow[d, shift left, "P^{-1}"] \\
\mathbb{R}^n \arrow[r, dotted, "D"'] \arrow[u, shift right, "P"'] & \mathbb{R}^n \arrow[u, shift left, "P"]
\end{tikzcd}$$

(a)

$$\begin{tikzcd}
\mathbf{v}_j \arrow[r, "A"] \arrow[d, shift right, "P^{-1}"'] & \lambda_j \mathbf{v}_j \arrow[d, shift left, "P^{-1}"] \\
\mathbf{e}_j = [\mathbf{v}_j]_{\mathcal{B}} \arrow[r, dotted, "D"'] \arrow[u, shift right, "P"'] & \lambda_j [\mathbf{v}_j]_{\mathcal{B}} = \lambda_j \mathbf{e}_j \arrow[u, shift left, "P"]
\end{tikzcd}$$

(b)

Figure 6.1 *Diagrams useful in following the proof of Theorem 6.8.*

Since $D = P^{-1}AP$, we have $PDP^{-1} = A$ so that

$$\begin{aligned}
A\mathbf{v}_j &= PDP^{-1}\mathbf{v}_j \\
&= (PD)(P^{-1}\mathbf{v}_j) \\
&= (PD)([\mathbf{v}_j]_{\mathcal{B}}) \\
&= PD\mathbf{e}_j \\
&= P(D\mathbf{e}_j) \\
&= P(\lambda_j \mathbf{e}_j) \\
&= \lambda_j (P\mathbf{e}_j) \\
&= \lambda_j (P[\mathbf{v}_j]_{\mathcal{B}}) \\
&= \lambda_j \mathbf{v}_j
\end{aligned}$$

for $j = 1, \cdots, n$. So $\mathcal{B}$ is a basis of eigenvectors of A. ■

As this proof shows, an invertible matrix P diagonalizes A if and only if the columns of P constitute a basis of $\mathbb{R}^n$ of eigenvectors of A.

Corollary 6.9. Let A be an $n \times n$ matrix. If A has n distinct eigenvalues, then A is diagonalizable.

Proof Suppose $\lambda_1, \cdots, \lambda_n$ are the n distinct eigenvalues of A. By the definition of eigenvalue, each eigenvalue λ_j has a corresponding eigenvector $\mathbf{v}_j$. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. By Theorem 6.3 of section 6.1, $\mathcal{B}$ is linearly independent. Since $\mathcal{B}$ is linearly independent and contains n vectors, $\mathcal{B}$ is a basis for $\mathbb{R}^n$ of eigenvectors of A. Therefore, A is diagonalizable by Theorem 6.8. ■

Example 6.6

Determine whether A and B defined below are diagonalizable.

$$A = \begin{bmatrix} 5 & 2 & -1 \\ 0 & 4 & 2 \\ 0 & 0 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 2 \\ 1 & -1 & 2 \end{bmatrix}$$

Solution The matrix A is triangular, so its eigenvalues appear on its main diagonal. Since A is 3×3 and has 3 distinct eigenvalues, A is diagonalizable. The matrix

$$D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

is a diagonal matrix that is similar to A.

Since B is not triangular, we use the characteristic polynomial to find the eigenvalues of B.

$$\begin{aligned} \det(B - \lambda I) &= \begin{vmatrix} 1-\lambda & 0 & 0 \\ 2 & -1-\lambda & 2 \\ 1 & -1 & 2-\lambda \end{vmatrix} \\ &= (1-\lambda)\begin{vmatrix} -1-\lambda & 2 \\ -1 & 2-\lambda \end{vmatrix} \\ &= (1-\lambda)\left[(-1-\lambda)(2-\lambda)+2\right] \\ &= -(\lambda-1)\left[\lambda^2-\lambda-2+2\right] \\ &= -(\lambda-1)(\lambda^2-\lambda) \\ &= -\lambda(\lambda-1)^2. \end{aligned}$$

The eigenvalues of B are $\lambda = 0, 1$. Since B has only two eigenvalues we are not yet sure whether B is diagonalizable. We need to know whether we have a basis of eigenvectors.

▶ For $\lambda = 0$, we solve $(B - 0I)\mathbf{x} = \mathbf{0}$.

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

Letting $z = t$ gives $x = 0, y = 2t$, and $z = t$. We can describe the solutions by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

so that

$$E_0 = span\left\{\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right\}.$$

▶ For $\lambda = 1$, we solve $(B - 1I)\mathbf{x} = \mathbf{0}$.

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

Letting $y = s$ and $z = t$ gives $x = s - t, y = s$, and $z = t$. We can describe the solutions by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = s\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

so that

$$E_1 = span\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\-1\end{bmatrix}\right\}.$$

It is easy to see that

$$\left\{\begin{bmatrix}0\\2\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\-1\end{bmatrix}\right\}$$

is a linearly independent set by noting that there are three pivot columns in the matrix they form after row reduction:

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

Thus $\mathbb{R}^3$ has a basis of eigenvectors of B and B is diagonalizable. In fact, B is similar to the diagonal matrix

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

Notice that the eigenvalue $\lambda = 1$ is a double root of the characteristic equation $-\lambda(\lambda-1)^2 = 0$ and the eigenspace E_1 is two dimensional. On the other hand, the eigenvalue $\lambda = 0$ is a single root of the characteristic equation and the eigenspace E_0 is one dimensional. This is not a coincidence, but the relationship is not as simple as this example suggests.

Definition 6.5. Let A be an $n \times n$ matrix and λ_0 an eigenvalue of A. The **algebraic multiplicity** of λ_0 is the number of times $\lambda - \lambda_0$ appears as a factor in the characteristic polynomial $\det(A - \lambda I)$. The **geometric multiplicity** of λ_0 is the dimension of the eigenspace E_{λ_0}.

Example 6.7

In Example 6.6, B has two eigenvalues 0 and 1. The eigenvalue 0 has both algebraic and geometric multiplicity of 1. The eigenvalue 1 has both algebraic and geometric multiplicity of 2.

Theorem 6.10 is not proved here, but you can use it to reduce the amount of work required to answer many questions.

Theorem 6.10. Let A be an $n \times n$ matrix. The geometric multiplicity of an eigenvalue of A is always greater than or equal to 1 and less than or equal to its algebraic multiplicity.

Theorem 6.3 in section 6.1 is a special case of a more general result. We state that result next but we do not prove it. Again, this theorem can be used to reduce the work required to answer many questions.

Theorem 6.11. Let A be an $n \times n$ matrix. Suppose $\lambda_1, \cdots, \lambda_r$ are distinct eigenvalues of A and $\mathcal{S}_1, \cdots, \mathcal{S}_r$ are linearly independent sets of eigenvectors corresponding to $\lambda_1, \cdots, \lambda_r$ respectively. Then $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_r$ is linearly independent.

In Theorem 6.3 of section 6.1, each set $\mathcal{S}_i$ contains a single eigenvector. We could have used these theorems to avoid some work in determining whether B is diagonalizable in Example 6.6. Since the characteristic polynomial of B is $-\lambda(\lambda-1)^2$, we know the geometric multiplicity of $\lambda = 0$ must be 1 and the geometric multiplicity of $\lambda = 1$ is 1 or 2 by Theorem 6.10. Once we check to see that the eigenvalue $\lambda = 1$ has a geometric multiplicity of 2, we know that we have 3 linearly independent eigenvectors, hence a basis of eigenvectors. This tells us B is diagonalizable. We did four calculations for B:

1. We determined the characteristic polynomial.

2. We determined a basis for the eigenspace E_0.

3. We determined a basis for the eigenspace E_1.

4. We checked the union of these bases for linear independence.

In light of Theorems 6.10 and 6.11, only calculations (1) and (3) were necessary.

What can go wrong that prevents a matrix from being diagonalizable? There are a couple of things. The following examples illustrate.

Example 6.8

Is

$$A = \begin{bmatrix} 3 & -2 \\ 5 & -3 \end{bmatrix}$$

diagonalizable?

Solution The characteristic polynomial of A is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 3-\lambda & -2 \\ 5 & -3-\lambda \end{vmatrix} \\ &= (3-\lambda)(-3-\lambda) + 10 \\ &= \lambda^2 + 1. \end{aligned}$$

Solving the characteristic equation $\lambda^2 + 1 = 0$ we get $\lambda = \pm i$. This tells us that there are no real numbers λ and nonzero $\mathbf{x} \in \mathbb{R}^2$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Hence A has no real eigenvalues.

This problem goes away if we expand our horizons a bit. If we allow complex entries in matrices and vectors, then there is no reason we can't diagonalize A. We must realize, however, that in that case we are not working in the vector space $\mathbb{R}^2$. In that case we are working in $\mathbb{C}^2$ (where $\mathbb{C}$ represents the set of complex numbers) and our set of scalars is $\mathbb{C}$ rather than $\mathbb{R}$.

Complex vector spaces are studied in great detail just as real vector spaces are. Much of the theory is identical, but in a few places like here they differ. We will not study complex vector spaces here, but you should be aware of their existence because they come up in many applications. If we carried out the complex arithmetic necessary to find eigenvectors of A we would find that

$$\begin{bmatrix} 2 \\ 3-i \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 3+i \end{bmatrix}$$

are eigenvectors of A corresponding to the eigenvalues of i and $-i$ respectively, and the matrix A is similar to

$$D = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

In this text, we simply say A is not diagonalizable over the real numbers because it has imaginary eigenvalues.

You may have guessed the other problem that prevents some matrices from being diagonalized. There are times when the geometric multiplicity of an eigenvalue is strictly less than its algebraic multiplicity. When that happens, there is no basis of eigenvectors so the matrix is not diagonalizable. Example 6.9 illustrates.

Example 6.9

Is

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

diagonalizable?

Solution The characteristic polynomial of A is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -1-\lambda & 4 \\ -1 & 3-\lambda \end{vmatrix} \\ &= (-1-\lambda)(3-\lambda) + 4 \\ &= \lambda^2 - 2\lambda + 1 \\ &= (\lambda - 1)^2. \end{aligned}$$

So A has one eigenvalue, namely $\lambda = 1$, and it has an algebraic multiplicity of 2. To determine the eigenspace, we solve $(A - 1I)\mathbf{x} = \mathbf{0}$. From row reduction,

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

let $y = t$ so that $x = -2t$. Then

$$E_1 = span \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}.$$

The eigenvalue $\lambda = 1$ has a geometric multiplicity of 1. Since $\mathbb{R}^2$ does not have a basis of eigenvectors of A, A is not diagonalizable.

Though complex vector spaces are not studied in this text, we prove the last theorem in this section that can involve imaginary eigenvalues. To understand this result we must realize that the eigenvalues (real and imaginary) are just the roots of the characteristic equation.

The fundamental theorem of algebra tells us that all polynomials of degree n can be factored into n linear factors, but those factors may contain imaginary numbers and some of the factors may be repeated. So if A is an $n \times n$ matrix, then its characteristic polynomial factors $p(\lambda) = \det(A - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda)$. Note that on the one hand, $p(0) = \det(A - 0I) = \det A$, and on the other hand $p(0) = (\lambda_1 - 0)\cdots(\lambda_n - 0) = \lambda_1\cdots\lambda_n$. Thus $\det A = \lambda_1\cdots\lambda_n$. This proves the following theorem.

Theorem 6.12. Let A be an $n \times n$ matrix. The determinant of A equals the product of its eigenvalues counting both real and imaginary eigenvalues and counting algebraic multiplicity.

Problem Set 6.3

1. Determine whether the following matrices are diagonalizable, and if so, provide a diagonal matrix similar to the given matrix.

(a) $\begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$ **(b)** $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ **(c)** $\begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 5 & -1 \end{bmatrix}$ **(e)** $\begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ **(f)** $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$

(g) $\begin{bmatrix} -3 & 4 & -4 \\ 0 & -1 & 0 \\ 2 & -4 & 3 \end{bmatrix}$ **(h)** $\begin{bmatrix} -5 & 8 & -6 \\ -1 & 1 & -1 \\ 2 & -4 & 3 \end{bmatrix}$

2. Suppose A, D, and P are square matrices and P is invertible. Prove $D = P^{-1}AP$ if and only if $AP = PD$. *(This result allows you to check whether $D = P^{-1}AP$ without calculating P^{-1}. You simply check to see whether $AP = PD$)*

3. For each diagonalizable matrix A and the diagonal matrix D you found from Exercise 1, find a matrix P that diagonalizes A to D and use Exercise 2 to show $D = P^{-1}AP$.

4. Suppose A is an $n \times n$ matrix that an invertible matrix P diagonalizes to D. Prove that $A = PDP^{-1}$.

5. Let $D_1 = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ and $D_2 = \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{bmatrix}$ be diagonal $n \times n$ matrices.

(a) Express $D_1 D_2$ as an $n \times n$ diagonal matrix.

(b) Let k be a positive integer. Express D_1^k as an $n \times n$ diagonal matrix.

(c) If D_1 is invertible, express D_1^{-1} as an $n \times n$ diagonal matrix.

(d) Prove $D_1 D_2 = D_2 D_1$.

6. Suppose A is an $n \times n$ matrix and that an invertible matrix P diagonalizes A to D.

(a) Prove that for k a positive integer, P diagonalizes A^k to D^k.

(b) Prove that if, in addition, A is invertible, then P diagonalizes A^{-1} to D^{-1}.

7. Use Exercises 4, 5, and 6 to express $\begin{bmatrix} 8 & -18 \\ 3 & -7 \end{bmatrix}^k$ as a 2×2 matrix.

8. Let $B = \begin{bmatrix} -16 & 100 \\ -5 & 29 \end{bmatrix}$. Use Exercises 4, 5, and 6 to find a 2×2 matrix A such that $A^2 = B$.

9. Suppose A and B are $n \times n$ matrices that are both diagonalized by the same invertible matrix P. Use Exercises 4 and 5 to prove that $AB = BA$.

10. True or False. A, P, and D are $n \times n$ matrices.

(a) A is diagonalizable if and only if there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

(b) If A is diagonalizable, then A has n distinct eigenvalues.

(c) If A has n distinct eigenvalues, then A is diagonalizable.

(d) If A has n distinct eigenvectors, then A is diagonalizable.

(e) If A has n linearly independent eigenvectors, then A is diagonalizable.

(f) If A is diagonalizable, then A has n linearly independent eigenvectors.

(g) If A has n eigenvalues counting algebraic multiplicity, then A is diagonalizable.

(h) If A has n eigenvalues counting geometric multiplicity, then A is diagonalizable.

(i) If A is invertible, then A is diagonalizable.

(j) If A is diagonalizable, then A is invertible.

11. For each of the following questions **(a)** through **(g)**, give one of the following answers:
(i) A is diagonalizable.
(ii) A is not diagonalizable.
(iii) A may be diagonalizable, but it is not certain.
(iv) The situation described is impossible.

(a) A is a 6×6 matrix with 5 distinct eigenvalues.

(b) A is a 5×5 matrix with a 3-dimensional eigenspace and a 2-dimensional eigenspace.

(c) A is a 3×3 matrix with exactly two distinct eigenvalues and each one has an algebraic multiplicity of 1.

(d) A is a 3×3 matrix with exactly two distinct eigenvalues and each one has a geometric multiplicity of 1.

(e) A is a 4×4 matrix with exactly two distinct eigenvalues. Both eigenvalues have an algebraic multiplicity of 2, but the geometric multiplicities are 1 and 2.

(f) A is a 4×4 matrix with exactly two distinct eigenvalues. Both eigenvalues have a geometric multiplicity of 2, but the algebraic multiplicities are 1 and 2.

(g) A is a 3×3 matrix with three distinct eigenvalues. Its eigenspaces have dimensions 0, 1, and 2.

12. Let A be a diagonalizable $n \times n$ matrix and P an invertible matrix that diagonalizes A. Prove that $(P^{-1})^T$ diagonalizes A^T.

13. Let A be a diagonalizable $n \times n$ matrix, and suppose B is similar to A. Prove that B is diagonalizable.

14. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ $(n \geq 2)$, and suppose $\mathbf{u}$ and $\mathbf{v}$ are nonzero and not orthogonal *(see Definition 2.6 in section 2.2)*. Let $A = \mathbf{u}\mathbf{v}^T$. The matrix A is an $n \times n$ matrix.

(a) What is the rank and nullity of A?

(b) Is A invertible?

(c) Is 0 an eigenvalue of A? If so, what is the dimension of the eigenspace associated with 0?

(d) Show that $\mathbf{u}$ is an eigenvector of A.

(e) What is the eigenvalue associated with $\mathbf{u}$?

(f) Is A diagonalizable?

(g) What is the characteristic polynomial of A?

15. Repeat Exercise 14 replacing $\mathbf{u}$ and $\mathbf{v}$ with nonzero orthogonal vectors.

6.4 Eigenvalues and Linear Operators

In section 6.1 eigenvalues and eigenvectors are defined for $n \times n$ matrices A. They could just as well be defined for a linear operator $T : V \longrightarrow V$ from a vector space V to itself.

Definition 6.6. Let V be a vector space and $T : V \longrightarrow V$ a linear operator on V. If there exists a scalar λ and a nonzero vector $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$, then λ is an **eigenvalue** of T and $\mathbf{v}$ is an **eigenvector** of T associated with the eigenvalue λ.

It should be clear that if A is an $n \times n$ matrix, then the eigenvalues and eigenvectors of A are exactly the same as the eigenvalues and eigenvectors of the linear operator

$T_A : \mathbb{R}^n \to \mathbb{R}^n$ given by $T_A(\mathbf{x}) = A\mathbf{x}$. But this linear operator definition of eigenvalues and eigenvectors is more general because it applies to linear operators on vector spaces in general (e.g. infinite dimensional ones) and not just on $\mathbb{R}^n$. Example 6.10 illustrates.

Example 6.10

Let V be the vector space of all real-valued functions of a single variable that have derivatives of all orders. Define $D : V \longrightarrow V$ by $D(f) = f'$. In Example 5.3, we showed that the differential operator is linear.

If $f(x) = e^{3x}$, then $D(f(x)) = \frac{d}{dx}(e^{3x}) = 3e^{3x} = 3f(x)$. So 3 is an eigenvalue of D and $f(x) = e^{3x}$ is an associated eigenvector. Of course there is nothing special about 3 in this example. For any real number λ, $\frac{d}{dx}(e^{\lambda x}) = \lambda e^{\lambda x}$ making every single real number λ an eigenvalue of the differential operator D with $f(x) = e^{\lambda x}$ an associated eigenvector.

Since the vector space V in Example 6.10 is infinite dimensional, matrix techniques cannot be employed to describe D completely. However, if V is a finite dimensional vector space with a basis $\mathcal{B}$ and T is a linear operator on V, then T can be described with a square matrix A via coordinate vectors and the isomorphism $T_{\mathcal{B}}$. The eigenvalues and eigenvectors of T can then be analyzed through A.

Example 6.11

Let $\mathbb{P}_3$ be the vector space of all polynomials of degree 3 or less and define $T : \mathbb{P}_3 \longrightarrow \mathbb{P}_3$ by $T(p(t)) = p(4-t)$. Show that T is indeed a linear operator and find the eigenvalues and corresponding eigenspaces of T.

Solution To show that T is linear observe that $T((p+q)(t)) = (p+q)(4-t) = p(4-t) + q(4-t) = T(p(t)) + T(q(t))$ and $T((cp)(t)) = (cp)(4-t) = cp(4-t) = cT(p(t))$. Since the degree of a composition of two polynomials equals the product of their degrees, the degree of $p(4-t)$ is the same as the degree of $p(t)$ since the degree of $r(t) = 4-t$ is 1. So T is indeed a linear operator on $\mathbb{P}_3$.

The set $\mathcal{B} = \{1, t, t^2, t^3\}$ is an ordered basis for $\mathbb{P}_3$ and

$$\begin{array}{rclcl} T(1) & = & 1 & = & 1 \\ T(t) & = & 4-t & = & 4-t \\ T(t^2) & = & (4-t)^2 & = & 16-8t+t^2 \\ T(t^3) & = & (4-t)^3 & = & 64-48t+12t^2-t^3 \end{array}$$

So the matrix of this transformation can be described by coordinate vectors by

$$\begin{array}{rcl} A & = & \left[[T(1)]_{\mathcal{B}}\ [T(t)]_{\mathcal{B}}\ [T(t^2)]_{\mathcal{B}}\ [T(t^3)]_{\mathcal{B}} \right] \\ & = & \begin{bmatrix} 1 & 4 & 16 & 64 \\ 0 & -1 & -8 & -48 \\ 0 & 0 & 1 & 12 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \end{array}$$

Since A is triangular, we see right away that its characteristic polynomial is $\det(A-\lambda I) = (1-\lambda)^2(-1-\lambda)^2 = (\lambda-1)^2(\lambda+1)^2$ so A has two eigenvalues $\lambda = 1$ and $\lambda = -1$ each having algebraic multiplicity 2. We proceed to find the eigenspaces.

▶ For $\lambda = 1$, we solve $(A - 1I)\mathbf{x} = \mathbf{0}$.

$$\begin{bmatrix} 0 & 4 & 16 & 64 \\ 0 & -2 & -8 & -48 \\ 0 & 0 & 0 & 12 \\ 0 & 0 & 0 & -2 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Letting $x_1 = \alpha$ and $x_3 = \beta$ gives $x_1 = \alpha, x_2 = -4\beta, x_3 = \beta$ and $x_4 = 0$. We can describe these solutions by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}$$

so that

$$E_1 = span\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -4 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

▶ For $\lambda = -1$, we solve $(A - (-1)I)\mathbf{x} = \mathbf{0}$.

$$\begin{bmatrix} 2 & 4 & 16 & 64 \\ 0 & 0 & -8 & -48 \\ 0 & 0 & 2 & 12 \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 8 & 32 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & -16 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Letting $x_2 = \alpha$ and $x_4 = \beta$ gives $x_1 = -2\alpha + 16\beta, x_2 = \alpha, x_3 = -6\beta$ and $x_4 = \beta$. We can describe these solutions by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \alpha \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 16 \\ 0 \\ -6 \\ 1 \end{bmatrix}$$

so that

$$E_{-1} = span\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 16 \\ 0 \\ -6 \\ 1 \end{bmatrix} \right\}.$$

So A is diagonalizable and

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 16 \\ 0 \\ -6 \\ 1 \end{bmatrix} \right\}$$

is a basis of eigenvectors. Translating back to $\mathbb{P}_3$ and T we have $p_1(t) = 1$ and $p_2(t) = t^2 - 4t$ form a basis for the eigenspace E_1 of T and $p_3(t) = t - 2$ and $p_4(t) = t^3 - 6t^2 + 16$ form a basis for the eigenspace E_{-1} of T.

Problem Set 6.4

1. In each part (a) - (c) below, you are given a linear operator T defined on $\mathbb{P}_3$.
(i) Find the matrix representation, A, of T relative to the ordered basis $\mathcal{B} = \{1, t, t^2, t^3\}$.
(ii) Find the eigenvalues, λ, and bases, $\mathcal{B}_\lambda$, for the associated eigenspaces of A.
(iii) Find the eigenvalues, λ, and bases, $\mathcal{C}_\lambda$, for the associated eigenspaces of T.
(iv) Does $\mathbb{P}_3$ have a basis of eigenvectors of T?

(a) $T(p(t)) = p'(t)$.

(b) $T(p(t)) = (t^2 - 1)p'(2)$.

(c) $T(p(t)) = (t + 1)p'(t)$.

2. In each part (a) - (c) below, you are given an $n \times n$ matrix B and a vector space V. In each case define the linear operator $T : V \to V$ by $T(X) = BX$ where $X \in V$.
(i) Find a matrix representation, A, of T.
(ii) Find the eigenvalues, λ, and bases, $\mathcal{B}_\lambda$, for the associated eigenspaces of A.
(iii) Find the eigenvalues, λ, and bases, $\mathcal{C}_\lambda$, for the associated eigenspaces of T.
(iv) Does V have a basis of eigenvectors of T?

(a) $B = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$, $V = M_{2,2}$, the vector space of all 2×2 matrices. *(Hint:* $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ *is an ordered basis of* $M_{2,2}$.*)*

(b) $B = \begin{bmatrix} 4 & 9 & -9 \\ -1 & -1 & 3 \\ 0 & 1 & 1 \end{bmatrix}$, $V = span\{I, B, B^2\}$. *(Hint:* $B^3 = 2I - 5B + 4B^2$.*)*

(c) $B = \begin{bmatrix} 5 & -4 & -4 \\ 0 & 1 & 0 \\ 2 & -2 & -1 \end{bmatrix}$, $V = span\{I, B\}$. *(Hint:* $B^2 = -3I + 4B$.*)*

7 (Real) Inner Product Spaces

7.1 Introduction

In chapter 2 we found geometric interpretations of vectors. Important in those geometric interpretations were the notions of length, distance, and angle. In chapter 2, when we were dealing with just two and three dimensions, we found and proved that these quantities are connected to the dot product. Now we want to extend these notions to other vector spaces. It turns out that this can be done in many different ways that produce different lengths, distances, and angles, but they all still conform to what are considered basic properties that length, distance, and angle should have. For example, some differences amount to nothing more than a change of scale. It is as though we are measuring in inches rather than centimeters, but some differences are more drastic so that what we consider perpendicular actually changes. We start with the most common extension for vectors in $\mathbb{R}^n$.

Definition 7.1. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ for some $u_1, \cdots, u_n, v_1, \cdots, v_n \in \mathbb{R}$. The **dot product** or **Euclidean inner product** of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n.$$

Other equivalent ways of expressing the dot product are

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

and

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$$

where $\mathbf{u}^T \mathbf{v}$ is matrix multiplication.

Definition 7.2. The vector space $\mathbb{R}^n$ together with the dot product is called **Euclidean n-space**.

Theorem 7.1 contains some important properties of the dot product.

Theorem 7.1. Suppose $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and c is a scalar.

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

(c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$

(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Proof Exercises. ■

It turns out that these four properties are key to develop notions of length, distance, and angle, so it is useful to define the more abstract notion of inner product.

Definition 7.3. Suppose V is a (real) vector space. A **(real) inner product on** V is a function that assigns a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$. This function must have the following four properties to be an inner product.
For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all scalars c:

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

(b) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

(c) $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

(d) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

A (real) vector space together with a (real) inner product is called a **(real) inner product space**.

At this point we drop the adjective real and assume that every time we mention a vector space, an inner product, or an inner product space, we are talking about a real one.

In chapter 6 we mentioned complex vector spaces. Indeed there are complex inner products and complex inner product spaces and even other types besides the real and complex varieties. They are not studied here but only mentioned so that you are not left with the false impression that the real ones are the only ones.

Theorem 7.1 proves that the dot product is an example of an inner product. In this course you see examples of a variety of inner products, but most of the examples involve the dot product in $\mathbb{R}^n$. Theorems that hold for inner products in general are proved that way, but our main focus is on the dot product on $\mathbb{R}^n$.

Example 7.1

Let A be an invertible $n \times n$ matrix. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, define $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$. It is easy to prove this is an inner product. Since it is defined in terms of matrix multiplication and the familiar dot product, we have a lot with which to work. Properties (a), (b), and

(c) follow very easily, in fact A need not even be invertible to satisfy these properties. We focus on property (d) where the necessity of A being invertible becomes clear.

Since $\langle \mathbf{u}, \mathbf{u} \rangle = (A\mathbf{u}) \cdot (A\mathbf{u})$ is a dot product of $A\mathbf{u}$ with itself, $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ by part (d) of Theorem 7.1. Also by Theorem 7.1, $\langle \mathbf{u}, \mathbf{u} \rangle = (A\mathbf{u}) \cdot (A\mathbf{u}) = 0$ if and only if $A\mathbf{u} = \mathbf{0}$. But $A\mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u} = \mathbf{0}$ since A is invertible, so $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

Example 7.2

Let $C[a, b]$ be the vector space of all continuous functions on $[a, b]$. For $f, g \in C[a, b]$, define

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

Since the product of two continuous functions is continuous and all continuous functions are integrable, this definition makes sense for all $f, g \in C[a, b]$.

The fact that this satisfies the four axioms of inner product is really just the result of familiar properties of the definite integral. For example,

$$\begin{aligned}
\langle f + g, h \rangle &= \int_a^b (f(x) + g(x))\, h(x)\ dx \\
&= \int_a^b (f(x)h(x) + g(x)h(x))\ dx \\
&= \int_a^b f(x)h(x)\ dx + \int_a^b g(x)h(x)\ dx \\
&= \langle f, h \rangle + \langle g, h \rangle.
\end{aligned}$$

The others are handled similarly.

Here are some additional properties of inner product that follow quickly from the four axioms and familiar properties of vectors and real numbers.

Theorem 7.2. If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in an inner product space, and c is a scalar, then

(a) $\langle \mathbf{0}, \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{0} \rangle = 0$

(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

(c) $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

(d) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$

(e) $\langle \mathbf{u}, \mathbf{v} - \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{w} \rangle$

Proof We prove part (e) as an example and leave the rest as exercises.

Justify each equal sign below with the correct axiom or property of vectors or real numbers:

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v}-\mathbf{w}\rangle &= \langle \mathbf{v}-\mathbf{w}, \mathbf{u}\rangle \\ &= \langle \mathbf{v}+(-1)\mathbf{w}, \mathbf{u}\rangle \\ &= \langle \mathbf{v}, \mathbf{u}\rangle + \langle (-1)\mathbf{w}, \mathbf{u}\rangle \\ &= \langle \mathbf{v}, \mathbf{u}\rangle + (-1)\langle \mathbf{w}, \mathbf{u}\rangle \\ &= \langle \mathbf{v}, \mathbf{u}\rangle - \langle \mathbf{w}, \mathbf{u}\rangle \\ &= \langle \mathbf{u}, \mathbf{v}\rangle - \langle \mathbf{u}, \mathbf{w}\rangle\end{aligned}$$

■

In chapter 2 we defined the norm of a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ in $\mathbb{R}^2$ to be $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$. In $\mathbb{R}^3$ we have $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ and $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$.

This gives us the length of the directed line segment that starts at the origin and ends at the point (v_1, v_2) in $\mathbb{R}^2$ and (v_1, v_2, v_3) in $\mathbb{R}^3$. Similarly, in chapter 2 we defined the distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ as $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$ in $\mathbb{R}^2$ and $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}$ in $\mathbb{R}^3$.

Note that the formula for norms can be rewritten in terms of the dot product and the formula for distances can be rewritten in terms of norms by $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ and $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$. Because the dimensions are too high, we don't have the geometric interpretations of vectors in $\mathbb{R}^n$ for $n > 3$ as we do in $\mathbb{R}^2$ and $\mathbb{R}^3$, but we can extend the definition analogously to vectors in $\mathbb{R}^n$ for all n and thereby impose some geometric notions on these otherwise algebraic entities.

Definition 7.4. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in Euclidean n-space ($\mathbb{R}^n$ with dot product). Define the **norm** of $\mathbf{v}$ to be

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

This is known as the 2-norm or the **Euclidean norm** in $\mathbb{R}^n$. Define the **distance** between $\mathbf{u}$ and $\mathbf{v}$ to be

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Moving to higher levels of abstraction we define norm and distance analogously for any inner product space.

Definition 7.5. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Define the **norm** of $\mathbf{v}$ to be

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

This is known as the norm of $\mathbf{v}$ relative to the given inner product. Define the **distance** between $\mathbf{u}$ and $\mathbf{v}$ to be

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Example 7.3

Let $\mathbf{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$, $f(x) = x^3$, and $g(x) = x$.

(a) The 2-norm is based on the dot product in $\mathbb{R}^2$ (Euclidean 2-space).

$$\|\mathbf{u}\| = \sqrt{2^2 + 3^2} = \sqrt{13}$$

and

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \left\| \begin{bmatrix} -3 \\ 5 \end{bmatrix} \right\| = \sqrt{\begin{bmatrix} -3 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 5 \end{bmatrix}} = \sqrt{34}.$$

(b) Relative to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$,

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{(A\mathbf{u}) \cdot (A\mathbf{u})} = \sqrt{\begin{bmatrix} 5 \\ 8 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 8 \end{bmatrix}} = \sqrt{5^2 + 8^2} = \sqrt{89}$$

and

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \left\| \begin{bmatrix} -3 \\ 5 \end{bmatrix} \right\| = \sqrt{\left\langle \begin{bmatrix} -3 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 5 \end{bmatrix} \right\rangle} = \sqrt{\left(A \begin{bmatrix} -3 \\ 5 \end{bmatrix} \right) \cdot \left(A \begin{bmatrix} -3 \\ 5 \end{bmatrix} \right)}$$

$$= \sqrt{\begin{bmatrix} 2 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 7 \end{bmatrix}} = \sqrt{2^2 + 7^2} = \sqrt{53}.$$

(c) Relative to the inner product $\langle f, g \rangle = \int_0^1 f(x) g(x)\, dx$,

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_0^1 x^6\, dx} = \sqrt{\frac{1}{7}}$$

and

$$d(f, g) = \|f - g\| = \|x^3 - x\| = \sqrt{\int_0^1 (x^3 - x)^2\, dx} = \sqrt{\frac{1}{7} - \frac{2}{5} + \frac{1}{3}} = \sqrt{\frac{8}{105}}.$$

You can see that the two different inner products on $\mathbb{R}^2$ given in parts (a) and (b) in Example 7.3 result in different norms and distances. This is common. Norms and distances typically depend on the inner product. Though the answers are different, both sets of norm and distance formulas are internally consistent. What do we mean by internally consistent? We mean that both pairs of norm and distance formulas behave the way we want norm and distance formulas to behave. How do we want norm and distance formulas to behave? Well, over the years mathematicians have decided on the properties all norms and distances should have. Definitions 7.6 and 7.7 spell these out.

Definition 7.6. let V be a vector space. A **norm** on V is a real-valued function that assigns the real number $\|\mathbf{v}\|$ to the vector $\mathbf{v}$. This function must have the following properties to be a norm. If $\mathbf{u}, \mathbf{v} \in V$ and c is a scalar, then

(a) $\|\mathbf{u}\| \geq 0$

(b) $\|\mathbf{u}\| = 0$ if and only if $\mathbf{u} = \mathbf{0}$

(c) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$

(d) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

Property (d) is called the **triangle inequality**.

Definition 7.7. Let V be a set. A **distance function** defined on V is a function that assigns a real number $d(\mathbf{u}, \mathbf{v})$ to each ordered pair of elements of $\mathbf{u}, \mathbf{v} \in V$. This function must have the following properties to be a distance function. If $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, then

(a) $d(\mathbf{u}, \mathbf{v}) \geq 0$

(b) $d(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$

(c) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$

(d) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$ (triangle inequality)

In section 7.2 we prove that all norms defined relative to an inner product $\left(\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}\right)$ satisfy the four properties of norms listed in Definition 7.6. We also prove that all distance formulas defined relative to a norm $(d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|)$ satisfy the four properties of distance formulas listed in Definition 7.7.

There are norms that do not come out of inner products in this way and there are distance formulas that do not come out of norms. They are not studied here. But if you have an inner product you get a norm and a distance formula that come along with it (for free). An inner product gives you something else. It gives you a way to define the angle between two vectors. This is examined in section 7.2 too.

Problem Set 7.1

1. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$, and $A = \begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix}$.

(a) Using the Euclidean inner product, find

(i) $\mathbf{u} \cdot \mathbf{v}$. **(ii)** $\mathbf{v} \cdot \mathbf{w}$. **(iii)** $\|\mathbf{u}\|$. **(iv)** $d(\mathbf{u}, \mathbf{v})$.

(b) Using the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$, find

(i) $\langle \mathbf{u}, \mathbf{v} \rangle$. **(ii)** $\langle \mathbf{v}, \mathbf{w} \rangle$. **(iii)** $\|\mathbf{u}\|$. **(iv)** $d(\mathbf{u}, \mathbf{v})$.

2. Let $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ -3 \\ 4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}$, and $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

(a) Using the Euclidean inner product, find

(i) $\mathbf{u} \cdot \mathbf{v}$. **(ii)** $\mathbf{v} \cdot \mathbf{w}$. **(iii)** $\|\mathbf{u}\|$. **(iv)** $d(\mathbf{u}, \mathbf{v})$.

(b) Using the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$, find

(i) $\langle \mathbf{u}, \mathbf{v} \rangle$. **(ii)** $\langle \mathbf{v}, \mathbf{w} \rangle$. **(iii)** $\|\mathbf{u}\|$. **(iv)** $d(\mathbf{u}, \mathbf{v})$.

3. In the inner product space $C[0, 2\pi]$ with inner product $\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)\, dx$, let $f(x) = \sin x$ and $g(x) = \cos x$. Calculate

(a) $\langle f, g \rangle$. **(b)** $\|f\|$. **(c)** $d(f, g)$.

(Hint: Use an integral table or a computer algebra system if necessary.)

4. Recall from chapter 2 the point-normal form $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) = 0$ for the equation of a line in Euclidean 2-space. This generalizes to the equation $\langle \mathbf{n}, \mathbf{x} - \mathbf{x_0} \rangle = 0$ in inner product spaces. Find the slope-intercept form for the line that satisfies this equation under the Euclidean inner product, and find the slope-intercept form for the line that satisfies this equation under the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ with $A = \begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix}$ for $\mathbf{n} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\mathbf{x_0} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

5. The **weighted Euclidean inner product** on $\mathbb{R}^n$ with weights $w_1, w_2, \ldots, w_n$ that are positive scalars has the formula $\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + \cdots + w_n u_n v_n$. Show that the weighted inner products form a special case of $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ by constructing an invertible $n \times n$ matrix A that produces the weighted Euclidean inner product formula.

6. Prove Theorem 7.1.

7. From Example 7.1, prove that $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ satisfies properties (a), (b), and (c) in the definition of inner product (Definition 7.3).

8. From Example 7.2, prove that $\langle f, g \rangle = \int_a^b f(x)g(x)\, dx$ satisfies properties (a) and (c) in the definition of inner product (Definition 7.3). *(Note: Part (d) requires an ϵ-δ proof from calculus.)*

9. Prove parts (a), (b), (c), and (d) of Theorem 7.2.

10. Show that the following identity holds for vectors in any inner produce space.

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2$$

11. (**Parallelogram Law**) Show that the following identity holds for vectors in any inner product space.

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$

7.2 Angle and Orthogonality

In chapter 2 we learned that $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$ for θ the angle between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$. This was proved using algebraic properties of the dot product and the geometric law of cosines. To calculate the angle between two nonzero vectors, we simply divide both sides of the equation by $\|\mathbf{u}\|\|\mathbf{v}\|$ yielding $\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$. Finally,

$$\theta = \cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right).$$

We wish to generalize to all inner product spaces to impose a notion of angle that is consistent with what we have with the dot product in $\mathbb{R}^2$ and $\mathbb{R}^3$.

The most natural extension would be for the angle between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V to be the angle θ, $0 \le \theta \le \pi$, for which

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

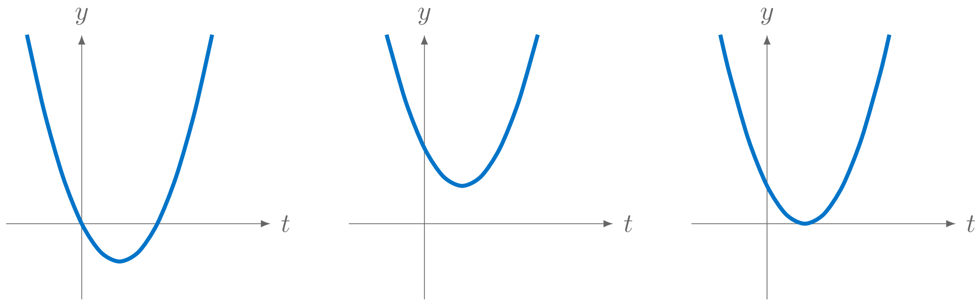
But in order for this to even make sense we need

$$-1 \le \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \le 1$$

or equivalently $|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\|\|\mathbf{v}\|$. This would match our geometric intuition even better if we could show for nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ that $\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|\|\mathbf{v}\|$ if and only if $\mathbf{v}$ is a positive multiple of $\mathbf{u}$ and $\langle \mathbf{u}, \mathbf{v} \rangle = -\|\mathbf{u}\|\|\mathbf{v}\|$ if and only if $\mathbf{v}$ is a negative multiple of $\mathbf{u}$ since $\cos 0 = 1$ (positive multiple) and $\cos\pi = -1$ (negative multiple). This is precisely what the Cauchy-Schwarz inequality gives us.

Before presenting the Cauchy-Schwarz inequality we remind you that the graph of an equation of the form $y = at^2 + bt + c$, for $a > 0$ is a parabola that opens up. The quadratic formula

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$



(a) $b^2 - 4ac > 0$ **(b)** $b^2 - 4ac < 0$ **(c)** $b^2 - 4ac = 0$

Figure 7.1 *The graph of* $y = at^2 + bt + c$.

tells us where the parabola crosses the t-axis. The expression under the radical, $b^2 - 4ac$, is called the **discriminant**. If the discriminant is positive, the parabola crosses the t-axis in two places. If it is negative, the parabola stays above the t-axis. If it is zero, the parabola comes down and touches the t-axis at just one point, the vertex of the parabola (see Figure 7.1).

Theorem 7.3 (Cauchy-Schwarz Inequality)**.** For all $\mathbf{u}, \mathbf{v}$ in an inner product space V,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|.$$

In addition for nonzero $\mathbf{u}, \mathbf{v}$, $\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\|$ if and only if one of the vectors is a positive multiple of the other and $\langle \mathbf{u}, \mathbf{v} \rangle = -\|\mathbf{u}\| \|\mathbf{v}\|$ if and only if one is a negative multiple of the other.

Proof To begin, note that the theorem is true if $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$. We may assume, therefore, that $\mathbf{u}$ and $\mathbf{v}$ are nonzero.

Consider the norm squared of $t\mathbf{u} - \mathbf{v}$:

$$\begin{aligned}
\|t\mathbf{u} - \mathbf{v}\|^2 &= \langle t\mathbf{u} - \mathbf{v}, t\mathbf{u} - \mathbf{v} \rangle \\
&= \langle t\mathbf{u}, t\mathbf{u} \rangle - 2\langle t\mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&= t^2 \langle \mathbf{u}, \mathbf{u} \rangle - 2t\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&= \|\mathbf{u}\|^2 t^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle t + \|\mathbf{v}\|^2
\end{aligned}$$

Since $\|\mathbf{u}\|^2 > 0$, the graph of $y = \|t\mathbf{u} - \mathbf{v}\|^2$ is a parabola that opens up. Since $\|t\mathbf{u} - \mathbf{v}\|^2 \geq 0$, the parabola never dips below the t axis, so the discriminant $4\langle \mathbf{u}, \mathbf{v} \rangle^2 - 4\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \leq 0$. Simplifying we get $\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$, and taking the positive square root of both sides we get

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|.$$

Next, suppose $\mathbf{u}$ and $\mathbf{v}$ are nonzero and $|\langle \mathbf{u}, \mathbf{v} \rangle| = \|\mathbf{u}\| \|\mathbf{v}\|$. We show that $\mathbf{v}$ is a multiple of $\mathbf{u}$.

By expanding the following inner product (Definition 7.3, Theorem 7.2, and Definition 7.4) we see

$$\begin{aligned}\left\langle \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|^2}\mathbf{u}-\mathbf{v}, \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|^2}\mathbf{u}-\mathbf{v}\right\rangle &= \frac{\langle \mathbf{u},\mathbf{v}\rangle^2}{\|\mathbf{u}\|^4}\langle \mathbf{u},\mathbf{u}\rangle - 2\frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|^2}\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\rangle^2 \\ &= \frac{\langle \mathbf{u},\mathbf{v}\rangle^2}{\|\mathbf{u}\|^4}\|\mathbf{u}\|^2 - 2\frac{\langle \mathbf{u},\mathbf{v}\rangle^2}{\|\mathbf{u}\|^2} + \|\mathbf{v}\|^2.\end{aligned}$$

And since $|\langle \mathbf{u},\mathbf{v}\rangle| = \|\mathbf{u}\|\|\mathbf{v}\|$, $\langle \mathbf{u},\mathbf{v}\rangle^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2$. We continue simplifying.

$$\begin{aligned} &= \frac{\|\mathbf{u}\|^2\|\mathbf{v}\|^2}{\|\mathbf{u}\|^4}\|\mathbf{u}\|^2 - 2\frac{\|\mathbf{u}\|^2\|\mathbf{v}\|^2}{\|\mathbf{u}\|^2} + \|\mathbf{v}\|^2 \\ &= \|\mathbf{v}\|^2 - 2\|\mathbf{v}\|^2 + \|\mathbf{v}\|^2 \\ &= 0.\end{aligned}$$

By the definition of inner product (Definition 7.3(d)), therefore, $\frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|^2}\mathbf{u} - \mathbf{v} = \mathbf{0}$. Thus, $\mathbf{v} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|^2}\mathbf{u}$, so $\mathbf{v}$ is a multiple of $\mathbf{u}$.

Now, suppose $\mathbf{v} = c\mathbf{u}$. We show that $\langle \mathbf{u},\mathbf{v}\rangle = \|\mathbf{u}\|\|\mathbf{v}\|$ if $c \geq 0$ and that $\langle \mathbf{u},\mathbf{v}\rangle = -\|\mathbf{u}\|\|\mathbf{v}\|$ if $c < 0$.

To begin, $\langle \mathbf{u},\mathbf{v}\rangle = \langle \mathbf{u},c\mathbf{u}\rangle = c\langle \mathbf{u},\mathbf{u}\rangle = c\|\mathbf{u}\|^2$. If $c \geq 0$, then $\langle \mathbf{u},\mathbf{v}\rangle = c\|\mathbf{u}\|^2 = \|\mathbf{u}\|(c\|\mathbf{u}\|) = \|\mathbf{u}\|\left(c\sqrt{\langle \mathbf{u},\mathbf{u}\rangle}\right) = \|\mathbf{u}\|\sqrt{c^2\langle \mathbf{u},\mathbf{u}\rangle} = \|\mathbf{u}\|\sqrt{\langle c\mathbf{u},c\mathbf{u}\rangle} = \|\mathbf{u}\|\sqrt{\langle \mathbf{v},\mathbf{v}\rangle} = \|\mathbf{u}\|\|\mathbf{v}\|$. If $c < 0$, then $c = -\sqrt{c^2}$ and so $\langle \mathbf{u},\mathbf{v}\rangle = c\|\mathbf{u}\|^2 = -\|\mathbf{u}\|\left(\sqrt{c^2}\sqrt{\langle \mathbf{u},\mathbf{u}\rangle}\right) = -\|\mathbf{u}\|\sqrt{c^2\langle \mathbf{u},\mathbf{u}\rangle} = -\|\mathbf{u}\|\sqrt{\langle c\mathbf{u},c\mathbf{u}\rangle} = -\|\mathbf{u}\|\sqrt{\langle \mathbf{v},\mathbf{v}\rangle} = -\|\mathbf{u}\|\|\mathbf{v}\|$.

■

We can now show that the norm of an inner product space, $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v},\mathbf{v}\rangle}$, satisfies the four properties we want all norms to have (see Definition 7.6).

Theorem 7.4. The norm of an inner product space V defined by $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v},\mathbf{v}\rangle}$ satisfies the following four properties. If $\mathbf{u},\mathbf{v} \in V$ and c is a scalar, then

(a) $\|\mathbf{u}\| \geq 0$

(b) $\|\mathbf{u}\| = 0$ if and only if $\mathbf{u} = \mathbf{0}$

(c) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$

(d) $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (triangle inequality)

Proof

(a) By property (d) of the definition of inner product, $\langle \mathbf{u},\mathbf{u}\rangle \geq 0$, so $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u},\mathbf{u}\rangle} \geq 0$ also.

(b) Also by property (d), $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}$, so

$$\begin{aligned} \|\mathbf{u}\| = 0 \quad &\iff \quad \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = 0 \\ &\iff \quad \langle \mathbf{u}, \mathbf{u} \rangle = 0 \\ &\iff \quad \mathbf{u} = \mathbf{0}. \end{aligned}$$

(c) $\|c\mathbf{u}\| = \sqrt{\langle c\mathbf{u}, c\mathbf{u} \rangle} = \sqrt{c^2 \langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{c^2}\sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = |c|\|\mathbf{u}\|$ by property (c) of the definition of inner product.

(d)
$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| + \|\mathbf{v}\|^2 \text{since } \langle \mathbf{u}, \mathbf{v} \rangle \leq |\langle \mathbf{u}, \mathbf{v} \rangle| \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \text{ by the Cauchy-Schwarz inequality} \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \end{aligned}$$

So $\|\mathbf{u}+\mathbf{v}\|^2 \leq (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$. Taking the positive square root of both sides we get $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

■

Definition 7.8. As in $\mathbb{R}^2$ and $\mathbb{R}^3$, a **unit vector** in an inner product space is a vector with a norm of 1.

Using property Theorem 7.4(c), it is easy to show that the unit vector that is a positive multiple of a nonzero $\mathbf{v}$ is $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$. It is called the **unit vector in the direction of v**.

All normed vector spaces have an automatic distance function defined by $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$. We show that this distance function has the four propeties we want all distance functions to have.

Theorem 7.5. The distance function of a normed vector space V, $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$, satisfies the following four properties. If $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ then

(a) $d(\mathbf{u}, \mathbf{v}) \geq 0$

(b) $d(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$

(c) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$

(d) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

Proof

(a) $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| \geq 0$ by property (a) of a norm (Definition 7.6).

(b) $d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = 0 \iff \mathbf{u}-\mathbf{v} = \mathbf{0} \iff \mathbf{u} = \mathbf{v}$ by property (b) of a norm.

(c) $d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = \|(-1)(\mathbf{v}-\mathbf{u})\| = |-1|\|\mathbf{v}-\mathbf{u}\| = \|\mathbf{v}-\mathbf{u}\| = d(\mathbf{v},\mathbf{u})$.

(d)
$$\begin{aligned} d(\mathbf{u},\mathbf{v}) &= \|\mathbf{u}-\mathbf{v}\| \\ &= \|(\mathbf{u}-\mathbf{w})+(\mathbf{w}-\mathbf{v})\| \\ &\leq \|\mathbf{u}-\mathbf{w}\| + \|\mathbf{w}-\mathbf{v}\| \text{ by the triangle inequality of norms (Definition 7.6)} \\ &= d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v}). \end{aligned}$$

∎

The Cauchy-Schwarz inequality provides us with exactly what we need for the following definition to make sense.

Definition 7.9. Let V be an inner product space. We define the angle between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ to be
$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right).$$

The angle between two nonzero vectors $\mathbf{u},\mathbf{v} \in V$ is that angle θ such that $0 \leq \theta \leq \pi$ and
$$\cos\theta = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$
Clearing the denominator we get
$$\langle \mathbf{u},\mathbf{v}\rangle = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta.$$
We continue to define terms for inner product spaces in the way that is analogous to the definitions from Euclidean 2 and 3 space.

Definition 7.10. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** in an inner product space V if $\langle \mathbf{u},\mathbf{v}\rangle = 0$.

Since $\cos\theta = 0$ and $0 \leq \theta \leq \pi$ if and only if $\theta = \frac{\pi}{2}$, nonzero vectors are orthogonal when the angle between them is $\frac{\pi}{2}$ (that is, when the are perpendicular).

The zero vector is the only vector that is orthogonal to every other vector. It is the only vector that is orthogonal to itself.

Example 7.4

The angle between vectors
$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix}$$

in Euclidean 4-space is

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right) = \cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right) = \cos^{-1}\left(\frac{3}{\sqrt{3}\sqrt{6}}\right) = \cos^{-1}\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4} \text{ or } 45°.$$

Example 7.5

In Euclidean 2-space, the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are orthogonal. The angle between them is $\frac{\pi}{2}$, but in the inner product space $\mathbb{R}^2$ with the inner product $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u})\cdot(A\mathbf{u})$ where

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

The angle between them is

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{e}_1, \mathbf{e}_2\rangle}{\|\mathbf{e}_1\|\|\mathbf{e}_2\|}\right)$$

where

$$\langle \mathbf{e}_1, \mathbf{e}_2\rangle = (A\mathbf{e}_1)\cdot(A\mathbf{e}_2) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 3,$$

$$\|\mathbf{e}_1\| = \sqrt{\langle \mathbf{e}_1, \mathbf{e}_1\rangle} = \sqrt{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \sqrt{2},$$

and

$$\|\mathbf{e}_2\| = \sqrt{\langle \mathbf{e}_2, \mathbf{e}_2\rangle} = \sqrt{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 2 \end{bmatrix}} = \sqrt{5}$$

so

$$\theta = \cos^{-1}\left(\frac{3}{\sqrt{2}\sqrt{5}}\right) \approx 0.32175$$

or about 18.435°.

Example 7.6

The functions $f(t) = t^2 - t$ and $g(t) = 2t - 1$ are in $C[0,1]$. The standard inner product for this space is $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$. Calculating the inner product of these two functions we get

$$\begin{aligned}
\langle f, g\rangle &= \int_0^1 (t^2 - t)(2t - 1)\,dt \\
&= \int_0^1 (2t^3 - 3t^2 + t)\,dt \\
&= \frac{t^4}{2} - t^3 + \frac{t^2}{2}\Big|_0^1 \\
&= \frac{1}{2} - 1 + \frac{1}{2} \\
&= 0
\end{aligned}$$

So these functions are orthogonal in $C[0,1]$.

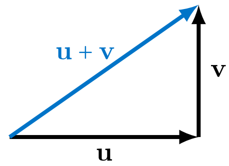


Figure 7.2 *The Pythagorean Theorem in an inner product space:*
$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Theorem 7.6 (Pythagorean Theorem)**.** If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in an inner product space V, then

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof
$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \text{ by the definition of inner product norm}\\ &= \langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle\\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \text{ since } \langle \mathbf{u}, \mathbf{v}\rangle = 0.\end{aligned}$$

■

Next, we define what it means for a vector to be orthogonal to a set of vectors.

Definition 7.11. Let S be a nonempty set of vectors in an inner product space V. A vector $\mathbf{v} \in V$ is **orthogonal to** S, written $\mathbf{v} \perp S$, if $\langle \mathbf{v}, \mathbf{s}\rangle = 0$ for every $\mathbf{s} \in S$. The **orthogonal complement** of S is $S^\perp = \{\mathbf{v} \in V : \mathbf{v} \perp S\}$.

The above definition says that $S^\perp$ (read S-perp) is the set of all vectors in V that are orthogonal to all vectors in S.

Theorem 7.7. Let V be an inner product space, S a nonempty subset of V, $\mathcal{S}$ a nonempty finite subset of V, and W a subspace of V.

(a) $S^\perp$ is a subspace of V

(b) $W \cap W^\perp = \{\mathbf{0}\}$

(c) $\mathcal{S}^\perp = (\mathit{span}\ \mathcal{S})^\perp$

(d) $S \subseteq (S^\perp)^\perp$

(e) If W is finite dimensional, then $W = (W^\perp)^\perp$.

Proof

(a) We use the subspace test.

1. Show $S^\perp \neq \varnothing$.
   ▶ Since $\langle \mathbf{0}, \mathbf{s}\rangle = 0$ for all $\mathbf{s} \in S$, we have $\mathbf{0} \in S^\perp$ so that $S^\perp \neq \varnothing$.
2. Suppose $\mathbf{u}, \mathbf{v} \in S^\perp$. Show $\mathbf{u} + \mathbf{v} \in S^\perp$.
   ▶ For all $\mathbf{s} \in S$, $\langle \mathbf{u} + \mathbf{v}, \mathbf{s}\rangle = \langle \mathbf{u}, \mathbf{s}\rangle + \langle \mathbf{v}, \mathbf{s}\rangle = 0 + 0 = 0$ since $\mathbf{u}, \mathbf{v} \in S^\perp$. Thus $\mathbf{u} + \mathbf{v} \in S^\perp$.
3. Suppose $\mathbf{v} \in S^\perp$ and c is a scalar. Show $c\mathbf{v} \in S^\perp$.
   ▶ For all $\mathbf{s} \in S$, $\langle c\mathbf{v}, \mathbf{s}\rangle = c\langle \mathbf{v}, \mathbf{s}\rangle = c(0) = 0$ since $\mathbf{v} \in S^\perp$. So $c\mathbf{v} \in S^\perp$.

By the subspace test, $S^\perp$ is a subspace of V.

(b) Since W and $W^\perp$ are subspaces of V, $\mathbf{0} \in W$ and $\mathbf{0} \in W^\perp$, so $\mathbf{0} \in W \cap W^\perp$. If $\mathbf{u} \in W \cap W^\perp$, then $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ since $\mathbf{u} \in W^\perp$ and $\mathbf{u} \in W$. But $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ implies $\mathbf{u} = \mathbf{0}$ so that $W \cap W^\perp = \{\mathbf{0}\}$.

(c) Since $\mathcal{S} \subseteq span\ \mathcal{S}$, every vector in $(span\ \mathcal{S})^\perp$ must be orthogonal to every vector in $\mathcal{S}$, so $(span\ \mathcal{S})^\perp \subseteq \mathcal{S}^\perp$. To see that $\mathcal{S}^\perp \subseteq (span\ \mathcal{S})^\perp$, let $\mathbf{v} \in \mathcal{S}^\perp$ and show $\mathbf{v} \in (span\ \mathcal{S})^\perp$. To show $\mathbf{v} \in (span\ \mathcal{S})^\perp$, let $\mathbf{w} \in span\ \mathcal{S}$ and show $\langle \mathbf{v}, \mathbf{w}\rangle = 0$. Since $\mathcal{S}$ is finite, let $\mathcal{S} = \{\mathbf{s}_1, \cdots, \mathbf{s}_n\}$. Since $\mathbf{w} \in span\ \mathcal{S}$, there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{w} = c_1\mathbf{s}_1 + \cdots + c_n\mathbf{s}_n$. So

$$\begin{aligned}\langle \mathbf{v}, \mathbf{w}\rangle &= \langle \mathbf{v}, c_1\mathbf{s}_1 + \cdots + c_n\mathbf{s}_n\rangle \\ &= c_1\langle \mathbf{v}, \mathbf{s}_1\rangle + \cdots + c_n\langle \mathbf{v}, \mathbf{s}_n\rangle \\ &= 0\end{aligned}$$

since $\mathbf{v} \in \mathcal{S}^\perp$. Therefore $\mathbf{v} \in (span\ \mathcal{S})^\perp$ and $\mathcal{S}^\perp = (span\ \mathcal{S})^\perp$.

(d) To show $S \subseteq (S^\perp)^\perp$, let $\mathbf{s} \in S$ and show $\mathbf{s} \in (S^\perp)^\perp$. To show $\mathbf{s} \in (S^\perp)^\perp$, let $\mathbf{v} \in S^\perp$ and show $\langle \mathbf{s}, \mathbf{v}\rangle = 0$. But, since $\mathbf{v} \in S^\perp$ and $\mathbf{s} \in S$, $\langle \mathbf{v}, \mathbf{s}\rangle = 0$, so $\langle \mathbf{s}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{s}\rangle = 0$. Thus $\mathbf{s} \in (S^\perp)^\perp$ and $S \subseteq (S^\perp)^\perp$.

(e) Proved later.

■

The proof of part (e) in Theorem 7.7 is delayed because we need results from the next two sections. In fact, it is possible to have $W \subsetneqq (W^\perp)^\perp$ if W is infinite dimensional. By the end of this section we are able to show the special case that $W = (W^\perp)^\perp$ in Euclidean n-space. We step carefully through the development of this intricate topic. For a clear understanding you must follow it carefully. Theorem 7.8 is the next step in that development.

Theorem 7.8. Let V be an inner product space and W a subspace of V. If $V = W \oplus W^\perp$, then $W = (W^\perp)^\perp$.

Proof By Theorem 7.7(d), $W \subseteq (W^\perp)^\perp$ so we need only show $(W^\perp)^\perp \subseteq W$. To show $(W^\perp)^\perp \subseteq W$, let $\mathbf{v} \in (W^\perp)^\perp$ and show $\mathbf{v} \in W$. Since $\mathbf{v} \in (W^\perp)^\perp \subseteq V = W \oplus W^\perp$, there exists $\mathbf{w} \in W$ and $\mathbf{u} \in W^\perp$ such that $\mathbf{v} = \mathbf{w} + \mathbf{u}$. We look at $\langle \mathbf{v}, \mathbf{u}\rangle$.

Since $\mathbf{v} \in (W^\perp)^\perp$ and $\mathbf{u} \in W^\perp$, $\langle \mathbf{v}, \mathbf{u} \rangle = 0$. On the other hand, since $\mathbf{v} = \mathbf{w} + \mathbf{u}$,

$$\langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{w} + \mathbf{u}, \mathbf{u} \rangle = \langle \mathbf{w}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{u} \rangle.$$

Since $\mathbf{u} \in W^\perp$ and $\mathbf{w} \in W$, $\langle \mathbf{w}, \mathbf{u} \rangle = 0$ so $\langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle$. Equating we get $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ which implies that $\mathbf{u} = \mathbf{0}$. Thus $\mathbf{v} = \mathbf{w} + \mathbf{u} = \mathbf{w} + \mathbf{0} = \mathbf{w}$, so $\mathbf{v} \in W$. Therefore $W = (W^\perp)^\perp$. ■

One wonders whether Theorem 7.8 could indeed be strengthened to "if and only if." Again, the answer is no because it is possible to have $W = (W^\perp)^\perp$ but $W \oplus W^\perp \subsetneq V$ if W is infinite dimensional.

Let A be an $m \times n$ matrix. Throughout this text we have seen several different interpretations of the solution set of a system $A\mathbf{x} = \mathbf{b}$. In this section we introduce a new interpretation of the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$. That is, a new interpretation of the null space of A.

Let $\mathbf{a}_1, \cdots, \mathbf{a}_m$ be the rows (not columns) of the matrix A. We rewrite the system $A\mathbf{x} = \mathbf{0}$ to look like

$$\begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note that in the process of the matrix multiplication $A\mathbf{x}$ we take the dot product of each row of A with the vector $\mathbf{x}$. So through the act of solving $A\mathbf{x} = \mathbf{0}$ we are finding all vectors $\mathbf{x}$ such that

$$\begin{aligned} \mathbf{a}_1 \cdot \mathbf{x} &= 0 \\ \vdots \quad & \quad \vdots \\ \mathbf{a}_m \cdot \mathbf{x} &= 0. \end{aligned}$$

In other words, the null space of A is the orthogonal complement of $\mathcal{S} = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$. But by Theorem 7.7(c), $\mathcal{S}^\perp = (span\ \mathcal{S})^\perp$, and since $span\ \mathcal{S}$ is the row space of A we have

$$null\ A = (row\ A)^\perp$$

in Euclidean n-space. In addition, since $rank\ A + nullity\ A = n$ (Definition 1.8) and $row\ A \cap null\ A = \{\mathbf{0}\}$ (Theorem 7.7(b))

$$\mathbb{R}^n = row\ A \oplus null\ A \text{ by Theorem 4.34.}$$

Next, we show that $(null\ A)^\perp = row\ A$. Again by Theorem 7.7(b), $null\ A \cap (null\ A)^\perp = \{\mathbf{0}\}$, so $null\ A \oplus (null\ A)^\perp$ is a subspace of $\mathbb{R}^n$ of dimension $dim(null\ A) + dim(null\ A)^\perp$ (Theorem 4.33). Thus,

$$dim(null\ A) + dim(null\ A)^\perp \leq n.$$

But, $dim(null\ A) = nullity\ A = n - rank\ A$, so

$$dim(null\ A)^\perp \leq rank\ A.$$

Also, $row\ A \subseteq ((row\ A)^\perp)^\perp = (null\ A)^\perp$ and $dim(row\ A) = rank\ A$, so

$$rank\ A \leq dim(null\ A)^\perp.$$

Therefore, $dim(null\ A)^{\perp} = rank\ A$. So $row\ A$ is a subspace of $(null\ A)^{\perp}$. But these have the same dimension, so

$$(null\ A)^{\perp} = row\ A.$$

We summarize with Theorem 7.9.

Theorem 7.9. Let A be an $m \times n$ matrix, $row\ A$ the row space of A, and $null\ A$ the null space of A in Euclidean n-space. Then, the following hold.

(a) $null\ A = (row\ A)^{\perp}$

(b) $row\ A = (null\ A)^{\perp}$

(c) $\mathbb{R}^n = row\ A \oplus null\ A$

This shows us how to use matrix techniques to find orthogonal complements in Euclidean n-space.

Example 7.7

Let W be the span of

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\}$$

in Euclidean 5-space. Find a basis for $W^{\perp}$.

Solution We enter the spanning set as rows in the matrix A and solve $A\mathbf{x} = \mathbf{0}$ to find a basis for the null space of A.

▶ Row reduce.

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

▶ Assign free variables to nonpivot columns and state the parameterized solution.

Let $x_3 = r, x_4 = s$, and $x_5 = t$. Then

$$\begin{array}{rcrcrcr} x_1 & = & -2r & - & s & & \\ x_2 & = & r & - & s & - & 2t \\ x_3 & = & r & & & & \\ x_4 & = & & & s & & \\ x_5 & = & & & & & t \end{array}$$

So

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = r \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Thus

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $W^{\perp}$.

Recall that the dimension of the row space of A equals the rank of A and the dimension of the null space equals the nullity of A which equals $n - rank\ A$. In Example 7.7, since $rank\ A = 2$, the dimension of $W^{\perp} = 5 - 2 = 3$.

We finish off this section with a proof of Theorem 7.7(e) in the case where V is Euclidean n-space.

Theorem 7.10. Let W be a subspace of Euclidean n-space.

(a) $\mathbb{R}^n = W \oplus W^{\perp}$

(b) $W = (W^{\perp})^{\perp}$

Proof

(a) Let $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ be a basis for W and let

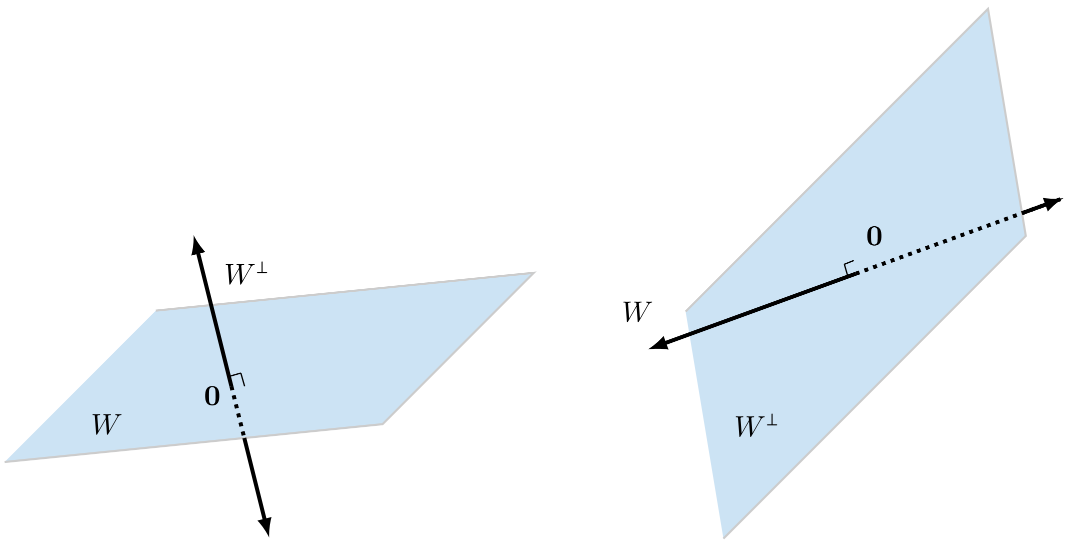
$$A = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_r \end{bmatrix}$$

be the $r \times n$ matrix with *rows* the vectors $\mathbf{w}_1, \cdots, \mathbf{w}_r$. Then $W = row\ A$ so $W^{\perp} = null\ A$ and $\mathbb{R}^n = W \oplus W^{\perp}$ by Theorem 7.9.

(b) Part (a) above and Theorem 7.8 imply $W = (W^{\perp})^{\perp}$.

■

Later in this chapter we are able to show that this theorem holds for all finite dimensional subspaces W and not just subspaces of Euclidean n-space. Examples exist, however, in infinite dimensional vector spaces in which W is infinite dimensional and is a proper subset of $(W^{\perp})^{\perp}$. These examples are not discussed in this text.



(a) In $\mathbb{R}^3$, if W is a plane then $W^{\perp}$ is a line.

(b) In $\mathbb{R}^3$, if W is a line then $W^{\perp}$ is a plane.

Figure 7.3 *Some diagrams illustrating Theorem 7.10 in Euclidean 3-space. In addition, we have* $\left(\mathbb{R}^3\right)^{\perp} = \{\mathbf{0}\}$ *and* $\{\mathbf{0}\}^{\perp} = \mathbb{R}^3$.

Problem Set 7.2

1. Determine whether the following pairs of vectors are orthogonal in the given inner product space.

(a) $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$ in Euclidean 3-space.

(b) $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$ in $\mathbb{R}^3$ with inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ where $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}$.

(c) $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$ in $\mathbb{R}^3$ with the weighted Euclidean inner product (see section 7.1, Exercise 5) with weights $w_1 = 1$, $w_2 = 2$, $w_3 = 3$.

(d) $f(x) = x$, $g(x) = \cos x$ in $C[-\pi/2, \pi/2]$ with inner product $\langle f, g \rangle = \int_{-\pi/2}^{\pi/2} f(x)g(x)dx$.

2. Find the unit vector in the direction of the given vector in the given inner product space.

(a) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, in Euclidean 2-space.

(b) $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, in Euclidean 4-space.

(c) $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, in $\mathbb{R}^4$ with inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u})\cdot(A\mathbf{v})$ where $A = \begin{bmatrix} 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

(d) $f(x) = x^2$ in $C[0,1]$ with inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.

3. Find the angle between the two vectors in the given inner product space.

(a) $\begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}, \begin{bmatrix} -\sqrt{3} \\ 3 \end{bmatrix}$, in Euclidean 2-space.

(b) $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, in Euclidean 4-space.

(c) $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, in $\mathbb{R}^4$ with inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u})\cdot(A\mathbf{v})$ where $A = \begin{bmatrix} 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

(d) $f(x) = x$, $g(x) = x^2$ in $C[0,1]$ with inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.

4. Determine whether $\mathbf{u}$ is orthogonal to S in Euclidean 3-space.

(a) $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 5 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix} \right\}$.

(b) $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $S = span \left\{ \begin{bmatrix} 5 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix} \right\}$.

(c) $\mathbf{u} = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right\}$.

5. Find two vectors of norm 1 in Euclidean 4-space that are orthogonal to $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix} \right\}$.

6. Let $W = span\ \mathcal{S}$. Find a basis for $W^\perp$ in Euclidean 4-space.

(a) $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ -7 \\ -4 \end{bmatrix} \right\}$.

(b) $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 5 \\ 1 \end{bmatrix} \right\}$.

7. Verify that $dim\ W + dim\ W^\perp = 4$ in both parts of Exercise 6 by finding the dimensions of W and $W^\perp$.

8. Let W be the plane $x + 2y - 3z = 0$ in Euclidean 3-space. Find parametric equations for $W^\perp$.

9. Let W be the line $\frac{x}{3} = \frac{y}{4} = \frac{z}{5}$ in Euclidean 3-space. Find the equation for $W^\perp$.

10. Let $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ and $A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

(a) Find a basis for the orthogonal complement of $\mathcal{S}$ in Euclidean 3-space.

(b) Find a basis for the orthogonal complement of $\mathcal{S}$ in the inner product space $\mathbb{R}^3$ with the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$.

11. Suppose $\mathbf{u}$ and $\mathbf{v}$ are orthogonal with $\|\mathbf{u}\| = 2$ and $\|\mathbf{v}\| = 3$ in an inner product space V. Find $d(\mathbf{u}, \mathbf{v})$ in V.

12. Use the Cauchy-Schwarz inequality to prove that for all real numbers a, b, θ we have

$$(a\cos\theta + b\sin\theta)^2 \leq a^2 + b^2.$$

13. Prove that for any positive integers m and n such that $m \neq n$, the functions $f_m(x) = \cos mx$ and $f_n(x) = \cos nx$ are orthogonal in the inner product space $C[0, \pi]$ with inner product $\langle f, g \rangle = \int_0^\pi f(x)g(x)dx$. *(Hint:* $\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)]$*)*

7.3 Orthogonal and Orthonormal Bases

Definition 7.12. A set of vectors in an inner product space is called an **orthogonal set** if all pairs of distinct vectors from that set are orthogonal. An orthogonal set in which each vector has a norm of 1 is called an **orthonormal set**.

Example 7.8

Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix},$$

and $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in Euclidean 3-space. Since $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, $\mathbf{v}_1 \cdot \mathbf{v}_3 = 0$, and $\mathbf{v}_2 \cdot \mathbf{v}_3 = 0$, $\mathcal{S}$ is an orthogonal set.

Each of these vectors has norm of 3. Normalizing each $\left(\frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i\right)$, let $\mathbf{u}_1 = \frac{1}{3}\mathbf{v}_1$, $\mathbf{u}_2 = \frac{1}{3}\mathbf{v}_2$, $\mathbf{u}_3 = \frac{1}{3}\mathbf{v}_3$, and $\mathcal{T} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. Then $\mathcal{T}$ is an orthonormal set.

The first thing we observe is this natural generalization of the Pythagorean Theorem.

Theorem 7.11. If $\mathbf{u} = \mathbf{v}_1 + \cdots + \mathbf{v}_n$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an orthogonal set in an inner product space V, then

$$\|\mathbf{u}\|^2 = \|\mathbf{v}_1\|^2 + \cdots + \|\mathbf{v}_n\|^2.$$

Proof

$$\begin{aligned}
\|\mathbf{u}\|^2 &= \langle \mathbf{u}, \mathbf{u} \rangle \\
&= \langle \mathbf{v}_1 + \cdots + \mathbf{v}_n, \mathbf{u} \rangle \\
&= \langle \mathbf{v}_1, \mathbf{u} \rangle + \cdots + \langle \mathbf{v}_n, \mathbf{u} \rangle \\
&= \langle \mathbf{v}_1, \mathbf{v}_1 + \cdots + \mathbf{v}_n \rangle + \cdots + \langle \mathbf{v}_n, \mathbf{v}_1 + \cdots + \mathbf{v}_n \rangle \\
&= (\langle \mathbf{v}_1, \mathbf{v}_1 \rangle + 0 + \cdots + 0) + \cdots + (0 + \cdots + 0 + \langle \mathbf{v}_n, \mathbf{v}_n \rangle) \\
&= \|\mathbf{v}_1\|^2 + \cdots + \|\mathbf{v}_n\|^2
\end{aligned}$$

■

Next, we observe that a finite orthogonal set of nonzero vectors must be linearly independent.

Theorem 7.12. If $\mathcal{S} = \{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is a finite orthogonal set of nonzero vectors in an inner product space V, then $\mathcal{S}$ is linearly independent.

Proof Suppose $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}$. We show $c_1 = \cdots = c_r = 0$. Since $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}$, $\langle \mathbf{v}_1, c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r \rangle = \langle \mathbf{v}_1, \mathbf{0} \rangle = 0$. But

$$\begin{aligned}
\langle \mathbf{v}_1, c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r \rangle &= c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle + c_2\langle \mathbf{v}_1, \mathbf{v}_2 \rangle + \cdots + c_r\langle \mathbf{v}_1, \mathbf{v}_r \rangle \\
&= c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle + 0 + \cdots + 0 \\
&= c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle
\end{aligned}$$

since $\langle \mathbf{v}_1, \mathbf{v}_k \rangle = 0$ for all $k \neq 1$.

Thus $c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = 0$ so that either $c_1 = 0$ or $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = 0$. But $\mathbf{v}_1$ nonzero implies $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle \neq 0$. Thus $c_1 = 0$. Similarly, taking the inner product of both sides of $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}$ with $\mathbf{v}_k$ for $k = 2, 3, \cdots, r$ shows $c_k = 0$ for $k = 2, 3, \cdots, r$. It follows that $\mathcal{S}$ is linearly independent. ■

Definition 7.13. An **orthogonal basis** of an inner product space V is a basis for V that is also an orthogonal set. An **orthonormal basis** is a basis that is also an orthonormal set.

Example 7.9

In Example 7.8, $\mathcal{S}$ is an orthogonal set of three nonzero vectors in Euclidean 3-space. By Theorem 7.12, $\mathcal{S}$ is linearly independent, so $\mathcal{S}$ is an orthogonal basis for Euclidean 3-space. Similarly, $\mathcal{T}$ from Example 7.8 is an orthonormal basis for Euclidean 3-space.

Example 7.10

The standard basis $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is an orthonormal basis for Euclidean n-space.

Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V. Back in chapter 4 we learned that for each $\mathbf{w} \in V$ there exist unique scalars $c_1, \cdots, c_n$ such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Determining the values of these scalars $c_1, \cdots, c_n$ usually involved solving a system of linear equations. But if $\mathcal{B}$ is an orthogonal basis, there is an easy way to find those scalars without the need to solve a linear system.

Theorem 7.13. Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be an orthogonal basis for an inner product space V. For each $\mathbf{w} \in V$ the scalars $c_1, \cdots, c_n$ in

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

are given by

$$c_j = \frac{\langle \mathbf{w}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}$$

for $j = 1, \cdots, n$. So,

$$\mathbf{w} = \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \cdots + \frac{\langle \mathbf{w}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n.$$

Proof

$$\begin{aligned} \langle \mathbf{w}, \mathbf{v}_j \rangle &= \langle c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n, \mathbf{v}_j \rangle \\ &= c_1\langle \mathbf{v}_1, \mathbf{v}_j \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_j \rangle \\ &= c_j\langle \mathbf{v}_j, \mathbf{v}_j \rangle \end{aligned}$$

since $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for every $i \neq j$. But since $\mathbf{v}_j$ is a basis element, $\mathbf{v}_j \neq \mathbf{0}$, so $\langle \mathbf{v}_j, \mathbf{v}_j \rangle \neq 0$. Dividing both sides of

$$\langle \mathbf{w}, \mathbf{v}_j \rangle = c_j\langle \mathbf{v}_j, \mathbf{v}_j \rangle$$

by $\langle \mathbf{v}_j, \mathbf{v}_j \rangle$ we get

$$c_j = \frac{\langle \mathbf{w}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}.$$

Thus,

$$\mathbf{w}_j = \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \cdots + \frac{\langle \mathbf{w}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n.$$

■

Example 7.11

In Examples 7.8 and 7.9, we saw that if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix},$$

and $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, then $\mathcal{S}$ is an orthogonal basis. Let $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Find the coordinate vector $[\mathbf{w}]_{\mathcal{S}}$.

Solution Theorem 7.13 tells us that we can write $\mathbf{w}$ as a linear combination of basis vectors $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$ where

$$c_1 = \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} = \frac{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}} = \frac{1+4+6}{1+4+4} = \frac{11}{9},$$

$$c_2 = \frac{\langle \mathbf{w}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} = \frac{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}}{\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}} = \frac{2+2-6}{4+1+4} = \frac{-2}{9},$$

$$c_3 = \frac{\langle \mathbf{w}, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_3, \mathbf{v}_3 \rangle} = \frac{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}}{\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}} = \frac{2-4+3}{4+4+1} = \frac{1}{9}.$$

So, $\mathbf{w} = \frac{11}{9}\mathbf{v}_1 - \frac{2}{9}\mathbf{v}_2 + \frac{1}{9}\mathbf{v}_3$ and

$$[\mathbf{w}]_{\mathcal{S}} = \begin{bmatrix} 11/9 \\ -2/9 \\ 1/9 \end{bmatrix}.$$

When dealing with an orthonormal basis, things are even easier.

Theorem 7.14. Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V. For each $\mathbf{w} \in V$ the scalars $c_1, \cdots, c_n$ in

$$\mathbf{w} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$$

are given by

$$c_j = \langle \mathbf{w}, \mathbf{u}_j \rangle$$

for $j = 1, \cdots, n$. So,

$$\mathbf{w} = \langle \mathbf{w}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{w}, \mathbf{u}_n \rangle \mathbf{v}_n.$$

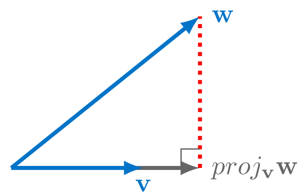


Figure 7.4 *Like Euclidean 2 and 3-space,* $proj_{\mathbf{v}}\mathbf{w} = \frac{\langle \mathbf{w},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ *in inner product spaces.*

Proof By Theorem 7.13,

$$c_j = \frac{\langle \mathbf{w}, \mathbf{u}_j\rangle}{\langle \mathbf{u}_j, \mathbf{u}_j\rangle}.$$

But since $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal set, $\langle \mathbf{u}_j, \mathbf{u}_j\rangle = 1$. So $c_j = \langle \mathbf{w}, \mathbf{u}_j\rangle$. ■

Recall from chapter 2 the orthogonal projection of $\mathbf{w}$ onto $\mathbf{v} \neq \mathbf{0}$ for vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ (see Figure 7.4). Orthogonal projections can now be defined for all inner product spaces and not just vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Theorem 7.15. Let $\mathbf{w}$ be a vector in an inner product space V and $\mathbf{v}$ a nonzero vector in V. There is a unique decomposition of $\mathbf{w}$ into the sum of two vectors

$$\mathbf{w} = \widehat{\mathbf{w}} + \widehat{\widehat{\mathbf{w}}}$$

where $\widehat{\mathbf{w}}$ is a multiple of $\mathbf{v}$ and $\widehat{\widehat{\mathbf{w}}}$ is orthogonal to $\mathbf{v}$.

Proof We want to know whether there are such pairs $\widehat{\mathbf{w}}$ and $\widehat{\widehat{\mathbf{w}}}$ and if so, we want to know how many such pairs.

Since $\widehat{\mathbf{w}}$ must be a multiple of $\mathbf{v}$, it has the form $x\mathbf{v}$ where x is a scalar. Since $\mathbf{w} = \widehat{\mathbf{w}} + \widehat{\widehat{\mathbf{w}}}$, $\widehat{\widehat{\mathbf{w}}} = \mathbf{w} - \widehat{\mathbf{w}}$ so that $\widehat{\widehat{\mathbf{w}}}$ must have the form $\mathbf{w} - x\mathbf{v}$. Since $\widehat{\widehat{\mathbf{w}}}$ must be orthogonal to $\mathbf{v}$, we need to find all values of x such that $\langle \mathbf{w} - x\mathbf{v}, \mathbf{v}\rangle = 0$. We solve:

$$\langle \mathbf{w} - x\mathbf{v}, \mathbf{v}\rangle = 0$$

$$\langle \mathbf{w}, \mathbf{v}\rangle - x\langle \mathbf{v}, \mathbf{v}\rangle = 0$$

$$x = \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}$$

since $\langle \mathbf{v}, \mathbf{v}\rangle \neq 0$. So there is only one such pair $\widehat{\mathbf{w}}$, $\widehat{\widehat{\mathbf{w}}}$, namely $\widehat{\mathbf{w}} = \frac{\langle \mathbf{w},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ and $\widehat{\widehat{\mathbf{w}}} = \mathbf{w} - \frac{\langle \mathbf{w},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$. ■

Definition 7.14. If V is an inner product space, $\mathbf{w} \in V$, and $\mathbf{v}$ is nonzero in V, then the **orthogonal projection** of $\mathbf{w}$ onto $\mathbf{v}$ is

$$proj_{\mathbf{v}}\mathbf{w} = \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}$$

and the **component of w orthogonal to v** is $\mathbf{w} - proj_{\mathbf{v}}\mathbf{w}$.

This gives us a nice rewording of Theorem 7.13 that provides some good geometric intuition.

Theorem 7.13 (restated)**.** Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an orthogonal basis for an inner project space V. For each $\mathbf{w} \in V$,

$$\mathbf{w} = proj_{\mathbf{v}_1}\mathbf{w} + \cdots + proj_{\mathbf{v}_n}\mathbf{w}.$$

Next we begin to extend the idea of orthogonal projections from onto a vector to onto a subspace.

Theorem 7.16. Let W be a subspace of an inner product space V with $\mathcal{B} = \{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ an orthogonal basis for W. Then $V = W \oplus W^{\perp}$ and $W = (W^{\perp})^{\perp}$.

Proof Since $W \cap W^{\perp} = \{\mathbf{0}\}$ (Theorem 7.7), the sum $W + W^{\perp}$ is, in fact, the direct sum $W \oplus W^{\perp}$. So $W \oplus W^{\perp}$ is a subspace of V. To show that it equals all of V, let $\mathbf{v} \in V$ and show that $\mathbf{v} \in W \oplus W^{\perp}$. Let

$$\widehat{\mathbf{v}} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_r \rangle}{\langle \mathbf{w}_r, \mathbf{w}_r \rangle}\mathbf{w}_r$$

and $\widehat{\widehat{\mathbf{v}}} = \mathbf{v} - \widehat{\mathbf{v}}$. Since $\widehat{\mathbf{v}}$ is a linear combination of basis vectors of W, $\widehat{\mathbf{v}} \in W$.

Next, we show that $\widehat{\widehat{\mathbf{v}}} \in \mathbf{W}^{\perp}$. For each $j = 1, \cdots, r$,

$$\begin{aligned}
\langle \widehat{\widehat{\mathbf{v}}}, \mathbf{w_j} \rangle &= \langle \mathbf{v} - \widehat{\mathbf{v}}, \mathbf{w}_j \rangle \\
&= \langle \mathbf{v}, \mathbf{w}_j \rangle - \langle \widehat{\mathbf{v}}, \mathbf{w}_j \rangle \\
&= \langle \mathbf{v}, \mathbf{w}_j \rangle - \left\langle \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_r \rangle}{\langle \mathbf{w}_r, \mathbf{w}_r \rangle}\mathbf{w}_r, \mathbf{w}_j \right\rangle \\
&= \langle \mathbf{v}, \mathbf{w}_j \rangle - \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\langle \mathbf{w}_1, \mathbf{w}_j \rangle - \cdots - \frac{\langle \mathbf{v}, \mathbf{w}_r \rangle}{\langle \mathbf{w}_r, \mathbf{w}_r \rangle}\langle \mathbf{w}_r, \mathbf{w}_j \rangle \\
&= \langle \mathbf{v}, \mathbf{w}_j \rangle - \frac{\langle \mathbf{v}, \mathbf{w}_j \rangle}{\langle \mathbf{w}_j, \mathbf{w}_j \rangle}\langle \mathbf{w}_j, \mathbf{w}_j \rangle \text{ since } \langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0 \text{ for all } i \neq j \\
&= \langle \mathbf{v}, \mathbf{w}_j \rangle - \langle \mathbf{v}, \mathbf{w}_j \rangle \\
&= 0.
\end{aligned}$$

So $\widehat{\widehat{\mathbf{v}}}$ is orthogonal to $\mathbf{w}_1, \mathbf{w}_2, \cdots,$ and $\mathbf{w}_r$. Thus $\widehat{\widehat{\mathbf{v}}} \in \mathcal{B}^\perp$, and since *span* $\mathcal{B} = W$, $\mathcal{B}^\perp = W^\perp$ by Theorem 7.7(c). So $\widehat{\widehat{\mathbf{v}}} \in \mathbf{W}^\perp$.

Since $\widehat{\widehat{\mathbf{v}}} = \mathbf{v} - \widehat{\mathbf{v}}$, we have $\mathbf{v} = \widehat{\mathbf{v}} + \widehat{\widehat{\mathbf{v}}}$, and since $\widehat{\mathbf{v}} \in W$ and $\widehat{\widehat{\mathbf{v}}} \in \mathbf{W}^\perp$, we know $\mathbf{v} \in W \oplus W^\perp$. Thus $V = W \oplus W^\perp$. By Theorem 7.8, $W = (W^\perp)^\perp$. ■

Let's refresh our memories a bit about sums and direct sums of subspaces. Suppose V is a vector space with U and W subspaces of V. If $V = U + W$, that means that every vector in V can be written as a sum of a vector from U plus a vector from W. That is, for every $\mathbf{v} \in V$ there exists $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. If we have a direct sum $V = U \oplus W$, that means $V = U + W$ and $U \cap W = \{\mathbf{0}\}$. That added condition $(U \cap W = \{\mathbf{0}\})$ has some important implications. The implication that is most important to us right now is that the representation $\mathbf{v} = \mathbf{u} + \mathbf{w}$ is unique. That is, there is only one pair of vectors $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$ (see Theorem 4.21).

Definition 7.15. If V is an inner product space and W is a subspace of V such that $V = W \oplus W^\perp$, then each vector $\mathbf{v} \in V$ has a unique decomposition $\mathbf{v} = \widehat{\mathbf{v}} + \widehat{\widehat{\mathbf{v}}}$ where $\widehat{\mathbf{v}} \in W$ and $\widehat{\widehat{\mathbf{v}}} \in \mathbf{W}^\perp$. We define the **orthogonal projection** of $\mathbf{v}$ onto W as $proj_W \mathbf{v} = \widehat{\mathbf{v}}$. The vector $\widehat{\widehat{\mathbf{v}}}$ is called the **component of v orthogonal** to W.

If W has a finite orthogonal basis, then the proof of Theorem 7.16 tells us exactly how to calculate $\widehat{\mathbf{v}}$ and $\widehat{\widehat{\mathbf{v}}}$.

Corollary 7.17. If V is an inner product space and W is a subspace of V with an orthogonal basis $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$, then for all $\mathbf{v} \in V$,

$$proj_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_r \rangle}{\langle \mathbf{w}_r, \mathbf{w}_r \rangle}\mathbf{w}_r$$

and the component of $\mathbf{v}$ orthogonal to W is

$$\mathbf{v} - proj_W \mathbf{v}.$$

Example 7.12

Let

$$\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix},$$

$\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2\}$ and $W = span\ \mathcal{B}$. Note that $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0$ so $\mathcal{B}$ is an orthogonal basis for the plane W in Euclidean 3-space spanned by $\mathbf{w}_1$ and $\mathbf{w}_2$. Find the orthogonal projection of $\mathbf{v}$ onto W and the component of $\mathbf{v}$ orthogonal to W.

Solution

$$\frac{\mathbf{v} \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \frac{7}{6} \text{ and } \frac{\mathbf{v} \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2} = \frac{4}{3}$$

So

$$proj_W\mathbf{v} = \frac{7}{6}\begin{bmatrix}1\\1\\2\end{bmatrix} + \frac{4}{3}\begin{bmatrix}1\\1\\-1\end{bmatrix} = \begin{bmatrix}5/2\\5/2\\1\end{bmatrix}$$

and

$$\mathbf{v} - proj_W\mathbf{v} = \begin{bmatrix}3\\2\\1\end{bmatrix} - \begin{bmatrix}5/2\\5/2\\1\end{bmatrix} = \begin{bmatrix}1/2\\-1/2\\0\end{bmatrix}.$$

Problem Set 7.3

1. For each part (a) - (d) complete the following:
(i) Verify that the set $\mathcal{S}$ is an orthogonal set in the given inner product space.
(ii) Construct an orthonormal set by normalizing the vectors in $\mathcal{S}$.
(iii) Determine whether $\mathcal{S}$ is a basis for the given inner product space.
(Hint: Theorems 7.12 & 4.32 make this task easier.)

(a) $\mathcal{S} = \left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\-4\\3\end{bmatrix}\right\}$ in Euclidean 3-space.

(b) $\mathcal{S} = \left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-7\\11\end{bmatrix}\right\}$ in $\mathbb{R}^2$ with inner product $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u})\cdot(A\mathbf{v})$ where $A = \begin{bmatrix}1 & 1\\2 & 1\end{bmatrix}$.

(c) $\mathcal{S} = \left\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}\right\}$ in Euclidean 4-space.

(d) $\mathcal{S} = \{p(x), q(x)\}$ where $p(x) = x^2 - x$, and $q(x) = 2x - 1$ in $\mathbb{P}_2$ with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$.

2. Use Theorem 7.13 to find the coordinate vector $[\mathbf{w}]_{\mathcal{B}}$ for the vector $\mathbf{w}$ and the orthogonal basis $\mathcal{B}$ in the given inner product space.

(a) $\mathbf{w} = \begin{bmatrix}1\\2\\1\end{bmatrix}$, $\mathcal{B} = \left\{\begin{bmatrix}1\\2\\2\end{bmatrix}, \begin{bmatrix}2\\1\\-2\end{bmatrix}, \begin{bmatrix}2\\-2\\1\end{bmatrix}\right\}$ in Euclidean 3-space.

(b) $\mathbf{w} = \begin{bmatrix}1\\2\\1\\2\end{bmatrix}$, $\mathcal{B} = \left\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}\right\}$ in Euclidean 4-space.

(c) $\mathbf{w} = \begin{bmatrix}3\\1\end{bmatrix}$, $\mathcal{B} = \left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-7\\11\end{bmatrix}\right\}$ in $\mathbb{R}^2$ with inner product $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u})\cdot(A\mathbf{v})$ where $A = \begin{bmatrix}1 & 1\\2 & 1\end{bmatrix}$.

(d) $\mathbf{w} = x^2 + 1$, $\mathcal{B} = \{1, 2x - 1, 6x^2 - 6x + 1\}$ in $\mathbb{P}_2$ with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$.

3. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}$, $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$, and $\mathbf{w}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$, and let $W_1 = span\{\mathbf{w}_1\}$, $W_2 = span\{\mathbf{w}_1, \mathbf{w}_2\}$, and $W_3 = span\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ in Euclidean 4-space.

(a) Find the orthogonal projection of $\mathbf{v}$ onto $\mathbf{w}_1$ and the component of $\mathbf{v}$ orthogonal to $\mathbf{w_1}$.

(b) Find the orthogonal projection of $\mathbf{v}$ onto W_1 and the component of $\mathbf{v}$ orthogonal to W_1.

(c) Find the orthogonal projection of $\mathbf{v}$ onto W_2 and the component of $\mathbf{v}$ orthogonal to W_2.

(d) Find the orthogonal projection of $\mathbf{v}$ onto W_3 and the component of $\mathbf{v}$ orthogonal to W_3.

4. Let W be a subspace of $\mathbb{R}^n$. This exercise develops an interesting new way to find the $n \times n$ orthogonal projection matrix P with the property that $P\mathbf{x} = proj_W\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ if you have an orthogonal basis for W.

(a) For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, note that $\mathbf{v}^T\mathbf{w} = \mathbf{v} \cdot \mathbf{w}$ is a scalar and $\mathbf{v}\mathbf{w}^T$ is an $n \times n$ matrix. Let $\mathbf{w}$ be a nonzero vector in $\mathbb{R}^n$ and let $P = (\frac{1}{\mathbf{w}^T\mathbf{w}})\mathbf{w}\mathbf{w}^T$. Show that $P\mathbf{x} = proj_{\mathbf{w}}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

(b) Let $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be an orthogonal basis for W, and let

$$P = (\frac{1}{\mathbf{w}_1^T\mathbf{w}_1})\mathbf{w}_1\mathbf{w}_1^T + \cdots + (\frac{1}{\mathbf{w}_k^T\mathbf{w}_k})\mathbf{w}_k\mathbf{w}_k^T.$$

Show that $P\mathbf{x} = proj_W\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

(c) Let $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Using the method described in part (a), find the orthogonal projection matrix P such that $P\mathbf{x} = proj_{\mathbf{w}}\mathbf{x}$.

(d) Let $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}\right\}$ and $W = span\mathcal{B}$. Using the method described in part (b), find the orthogonal projection matrix P such that $P\mathbf{x} = proj_W\mathbf{x}$.

5. Let W be a subspace of $\mathbb{R}^n$. For each $\mathbf{x} \in \mathbb{R}^n$ we know that there exist unique vectors $\hat{\mathbf{x}}$ and $\hat{\hat{\mathbf{x}}}$ such that $\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{x}$ onto W and $\hat{\hat{\mathbf{x}}}$ is the component of $\mathbf{x}$ orthogonal to W, so $\mathbf{x} = \hat{\mathbf{x}} + \hat{\hat{\mathbf{x}}}$. We define the **orthogonal transformation of x across W** to be $ot_W\mathbf{x} = \hat{\mathbf{x}} - \hat{\hat{\mathbf{x}}}$. This exercise develops an interesting new way to find R, the $n \times n$ orthogonal transformation matrix across W, with the property that $R\mathbf{x} = ot_W\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ if you have an orthogonal basis for W.

(a) Show that $ot_{W^\perp}\mathbf{x} = -ot_W\mathbf{x}$.

(b) Show that $ot_W\mathbf{x} = 2proj_W\mathbf{x} - \mathbf{x}$ as defined in Exercise 4.

(c) Show that $R = 2P - I_n$ as defined in Exercise 4.

(d) Suppose $\mathbf{w}$ is nonzero and $W = span\{\mathbf{w}\}$. Show that the orthogonal transformation matrix across W is $R = \left(\frac{2}{\mathbf{w}^T\mathbf{w}}\right)\mathbf{w}\mathbf{w}^T - I_n$. *(Note that* $-R = I_n - \left(\frac{2}{\mathbf{w}^T\mathbf{w}}\right)\mathbf{w}\mathbf{w}^T$ *is the orthogonal transformation matrix across* $W^\perp$*. The matrix* $-R$ *is known as the Householder matrix.)*

(e) Let W be the subspace of $\mathbb{R}^3$ from Exercise 4(d). Calculate the orthogonal transformation matrix in two ways; first by applying part (c) of this exercise, and second by noting that $\left\{\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}\right\}$ is a basis for $W^\perp$ and applying part (d).

6. Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis of an inner product space V, and suppose $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Prove that $\|\mathbf{u}\| = \sqrt{c_1^2 + \cdots + c_n^2}$.

7. True or false.

(a) Every finite orthogonal set in an inner product space is linearly independent.

(b) Every finite orthonormal set in an inner product space is linearly independent.

(c) Every linearly independent set in an inner product space is an orthogonal set.

(d) If the vectors in an orthogonal set of nonzero vectors are normalized, the resulting vectors need not be orthogonal.

(e) The orthogonal projection of a given vector onto a nonzero vector $\mathbf{v}$ is the same as the orthogonal projection of that vector onto a nonzero multiple of $\mathbf{v}$.

(f) Every orthogonal set of n nonzero vectors in $\mathbb{R}^n$ is a basis for $\mathbb{R}^n$.

(g) Every $n \times n$ matrix with nonzero orthogonal columns is invertible.

(h) Every $n \times n$ matrix with nonzero orthogonal rows is invertible.

7.4 Gram-Schmidt

In this section, we develop a process – the Gram-Schmidt process – that starts with an arbitrary basis $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ for an n-dimensional inner product space V and generates an orthogonal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ for V. It does this in such a way so that the sets $\{\mathbf{v}_1\}$, $\{\mathbf{v}_1, \mathbf{v}_2\}, \cdots, \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ are orthogonal sets and $span\{\mathbf{v}_1\} = span\{\mathbf{w}_1\}$, $span\{\mathbf{v}_1, \mathbf{v}_2\} = span\{\mathbf{w}_1, \mathbf{w}_2\}, \cdots, span\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = span\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$.

The process is recursive. We start by simply letting $\mathbf{v}_1 = \mathbf{w}_1$ and note that trivially $\{\mathbf{v}_1\}$ is an orthogonal set and $span\{\mathbf{v}_1\} = span\{\mathbf{w}_1\}$.

Suppose we have defined $\mathbf{v}_1, \cdots, \mathbf{v}_j$ for j some integer with $1 \le j < n$, with $\{\mathbf{v}_1, \cdots, \mathbf{v}_j\}$ an orthogonal set, and with $span\{\mathbf{v}_1, \cdots, \mathbf{v}_j\} = span\{\mathbf{w}_1, \cdots, \mathbf{w}_j\}$. We show how to construct $\mathbf{v}_{j+1}$.

Let $W_j = span\{\mathbf{v}_1, \cdots, \mathbf{v}_j\}$. Since $\{\mathbf{v}_1, \cdots, \mathbf{v}_j\}$ is an orthogonal basis for W_j, by Corollary 7.17 we can construct the orthogonal projection of $\mathbf{w}_{j+1}$ onto W_j by

$$proj_{W_j}\mathbf{w}_{j+1} = \frac{\langle \mathbf{w}_{j+1}, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 + \cdots + \frac{\langle \mathbf{w}_{j+1}, \mathbf{v}_j\rangle}{\langle \mathbf{v}_j, \mathbf{v}_j\rangle}\mathbf{v}_j.$$

We let

$$\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - proj_{W_j}\mathbf{w}_{j+1}.$$

We know that by this construction, $\mathbf{v}_{j+1} \in W_j^\perp$ so that $\{\mathbf{v}_1, \cdots, \mathbf{v}_j, \mathbf{v}_{j+1}\}$ is an orthogonal set.

To show that $span\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\} = span\{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\}$, we take an arbitrary element in each span and show it is in the other.

▶ Suppose $\mathbf{u} \in span\{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\}$ and show $\mathbf{u} \in span\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\}$.

Since $\mathbf{u} \in span\{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\}$, there exist scalars $c_1, \cdots, c_{j+1}$ such that

$$\begin{aligned} \mathbf{u} &= c_1\mathbf{w} + \cdots + c_j\mathbf{w}_j + c_{j+1}\mathbf{w}_{j+1} \\ &= (c_1\mathbf{w}_1 + \cdots + c_j\mathbf{w}_j) + c_{j+1}\mathbf{w}_{j+1}. \end{aligned}$$

Since $\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - proj_{W_j}\mathbf{w}_{j+1}$, we have $\mathbf{w}_{j+1} = \mathbf{v}_{j+1} + proj_{W_j}\mathbf{v}_{j+1}$. Thus

$$\begin{aligned} \mathbf{u} &= (c_1\mathbf{w}_1 + \cdots + c_j\mathbf{w}_j) + c_{j+1}\left(proj_{W_j}\mathbf{v}_{j+1} + \mathbf{v}_{j+1}\right) \\ &= (c_1\mathbf{w}_1 + \cdots + c_j\mathbf{w}_j + c_{j+1}proj_{W_j}\mathbf{v}_{j+1}) + c_{j+1}\mathbf{v}_{j+1}. \end{aligned}$$

Since both $c_1\mathbf{w}_1 + \cdots + c_j\mathbf{w}_j$ and $c_{j+1}proj_{W_j}\mathbf{v}_{j+1}$ are in $W_j = span\{\mathbf{v}_1, \cdots, \mathbf{v}_j\}$, their sum is too. So, there exist scalars $d_1, \cdots, d_j$ such that $\mathbf{u} = (d_1\mathbf{v}_1 + \cdots d_j\mathbf{v}_j) + c_{j+1}\mathbf{v}_{j+1} \in span\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\}$. So $span\ \{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\} \subseteq span\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\}$.

▶ Suppose $\mathbf{u} \in span\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\}$ and show $\mathbf{u} \in span\{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\}$.

Since $\mathbf{u} \in \{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\}$, there exist scalars $c_1, \cdots, c_{j+1}$ such that

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + c_{j+1}\mathbf{v}_{j+1}.$$

Since $\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - proj_{W_j}\mathbf{w}_{j+1}$,

$$\begin{aligned} \mathbf{u} &= c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + c_{j+1}\left(\mathbf{w}_{j+1} - proj_{W_j}\mathbf{w}_{j+1}\right) \\ &= \left(c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j - c_{j+1}proj_{W_j}\mathbf{w}_{j+1}\right) + c_{j+1}\mathbf{w}_{j+1}. \end{aligned}$$

Since both $c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j$ and $c_{j+1}proj_{W_j}\mathbf{w}_{j+1}$ are in $W_j = span\{\mathbf{w}_1, \cdots, \mathbf{w}_j\}$, their difference is too. So there exist scalars $d_1, \cdots, d_j$ such that $\mathbf{u} = (d_1\mathbf{w}_1 + \cdots + d_j\mathbf{w}_j) + c_{j+1}\mathbf{w}_{j+1} \in span\{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\}$.

It follows that $span\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\} = span\{\mathbf{w}_1, \cdots, \mathbf{w}_{j+1}\}$.

By induction, $\{\mathbf{v}_1, \cdots, \mathbf{v}_j\}$ is an orthogonal set and $span\{\mathbf{v}_1, \cdots, \mathbf{v}_j\} = span\{\mathbf{w}_1, \cdots, \mathbf{w}_j\}$ for $j = 1, \cdots, n$. In particular, for $j = n$, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an orthogonal basis for V. This proves Theorem 7.18.

Theorem 7.18 (Gram-Schmidt process). Suppose V is an n-dimensional inner product space with $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ a basis for V. The following process generates an orthogonal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ for V in such a way that the sets $\{\mathbf{v}_1\}$, $\{\mathbf{v}_1, \mathbf{v}_2\}, \cdots, \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ are orthogonal sets and $\mathit{span}\,\{\mathbf{v}_1\} = \mathit{span}\,\{\mathbf{w}_1\}$, $\mathit{span}\,\{\mathbf{v}_1, \mathbf{v}_2\} = \mathit{span}\,\{\mathbf{w}_1, \mathbf{w}_2\}, \cdots, \mathit{span}\,\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = \mathit{span}\,\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$. Start by letting $\mathbf{v}_1 = \mathbf{w}_1$. If $\mathbf{v}_1, \cdots, \mathbf{v}_j$ have been defined, define

$$\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \left(\frac{\langle \mathbf{w}_{j+1}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \cdots + \frac{\langle \mathbf{w}_{j+1}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \mathbf{v}_j \right).$$

The Gram-Schmidt process is quite tedious to carry out by hand. You should carry out the process a small number of times for a relatively small value of n to see that you understand the process. After that, software like *Maple* can carry out the tedious calculations for you.

Part of the problem with the messy and tedious calculations is that they involve fractions. When working by hand you can make the calculations easier by replacing each $\mathbf{v}_j$ with a nonzero multiple of $\mathbf{v}_j$ that eliminates all fractions in the entries. We illustrate in the following example.

Example 7.13

Let

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

and $W = \mathit{span}\,\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. Use the Gram-Schmidt process to find an orthogonal basis for the 3-dimensional subspace W in Euclidean 4-space.

Solution

▶ Let

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}.$$

▶ Compute

$$\mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix} - \frac{3}{6} \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/2 \\ 2 \\ -1/2 \end{bmatrix}$$

To avoid fractions (i.e. for simplicity) in our calculation, we define $\mathbf{v}_2$ by

$$\mathbf{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \end{bmatrix}.$$

▶ Compute

$$\begin{aligned}
\mathbf{w}_3 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 &= \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} - \frac{(-2)}{6}\begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix} - \frac{6}{18}\begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ -6/3 \\ 3/3 \\ 0 \end{bmatrix} + \begin{bmatrix} 2/3 \\ 1/3 \\ 0 \\ -1/3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1/3 \\ -4/3 \\ 1/3 \end{bmatrix} \\
&= \begin{bmatrix} 2/3 \\ -4/3 \\ -1/3 \\ 0 \end{bmatrix}
\end{aligned}$$

Again, for simplicity, we avoid fractions by defining

$$\mathbf{v}_3 = \begin{bmatrix} 2 \\ -4 \\ -1 \\ 0 \end{bmatrix}.$$

▶ An orthogonal basis for W is

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

For an orthonormal basis, simply normalize each vector in the basis:

$$\left\{ \frac{1}{\sqrt{6}}\begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{18}}\begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{21}}\begin{bmatrix} 2 \\ -4 \\ -1 \\ 0 \end{bmatrix} \right\}$$

It is important to realize that the results of the Gram-Schmidt process depend on which order you proceed through the original basis. A different orthogonal basis for W results if we begin with $\mathbf{w}_2$ or $\mathbf{w}_3$.

The important theoretical result that comes from the Gram-Schmidt process is that it shows us that all finite-dimensional subspaces W of an inner product space V have orthogonal bases. We can, therefore, improve the wording of Theorem 7.16.

Theorem 7.19. If W is a finite-dimensional subspace of an inner product space V, then $V = W \oplus W^{\perp}$ and $W = (W^{\perp})^{\perp}$.

Theorem 7.19 tells us that we can take the orthogonal projection of any vector in V onto W if W is finite dimensional.

Problem Set 7.4

1. The set $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix} \right\}$ is a basis for Euclidean 2-space.

(a) Apply the Gram-Schmidt process to $\mathcal{B}$ to obtain an orthogonal basis for Euclidean 2-space.

(b) Normalize the vectors in the new basis obtained in (a) to form an orthonormal basis for Euclidean 2-space.

(c) Apply the Gram-Schmidt process to $\mathcal{B}$ to obtain an orthogonal basis for Euclidean 2-space but start the process with the vector you did not start with in (a). Note that this results in a a different orthogonal basis of Euclidean 2-space.

2. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 7 \\ 4 \end{bmatrix} \right\}$, and let $W = span\{\mathcal{B}\}$ in Euclidean 3-space. Apply the Gram-Schmidt process on $\mathcal{B}$ to construct an orthonormal basis for W.

3. $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$, and let $W = span\{\mathcal{B}\}$ in Euclidean 4-space. Apply the Gram-Schmidt process on $\mathcal{B}$ to construct an orthogonal basis for W.

4. Let $\mathcal{B} = \{x - 1, x^2 + x + 1\}$ in the inner product space $\mathbb{P}_2$ with inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$, and let $W = span\mathcal{B}$.

(a) Apply the Gram-Schmidt process to $\mathcal{B}$ to obtain an orthogonal basis for the subspace W.

(b) Normalize the vectors in the new basis obtained in (a) to form an orthonormal basis for W.

5. True or false.

(a) Every finite-dimensional subspace of inner product space has an orthonormal basis.

(b) Depending on the order in which the Gram-Schmidt process is applied to the vectors of a basis, the resulting vectors need not be linearly independent.

(c) Depending on the order in which the Gram-Schmidt process is applied to the vectors of a basis, the resulting vectors need not be orthogonal.

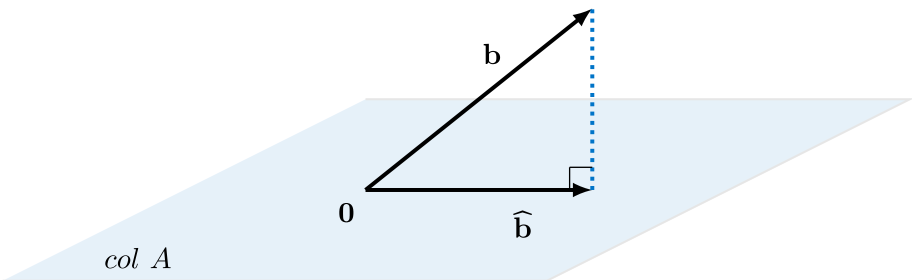
(d) Depending on the order in which the Gram-Schmidt process is applied to the vectors of a basis of a subspace, the resulting vectors need not span the same subspace as the original basis.

(e) Depending on the order in which the Gram-Schmidt process is applied to the vectors of a basis, it can result in different orthogonal bases of the same inner product space.

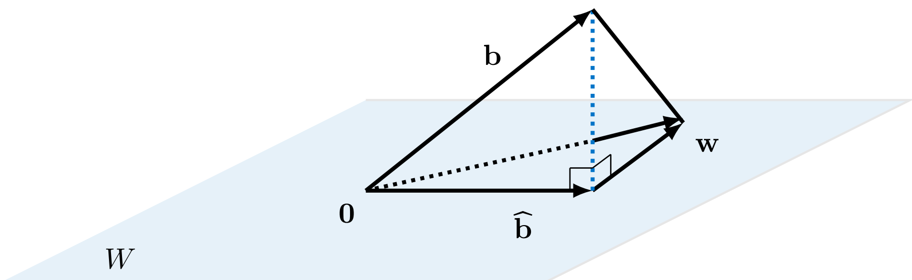
7.5 Least Squares

It frequently arises in applications that you need to find a "solution" to an inconsistent system $A\mathbf{x} = \mathbf{b}$. We place quotation marks around the word solution because, of course, an inconsistent system has no solution. When a genuine solution isn't possible we want something that is close.

Recall from chapter 4 that the column space of a matrix A equals the set of all vectors $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ is consistent. Whenever we have an inconsistent system $A\mathbf{x} = \mathbf{b}$, it is because $\mathbf{b}$ is not in the column space of A. The idea behind the method of least squares is to change the system $A\mathbf{x} = \mathbf{b}$ to $A\mathbf{x} = \widehat{\mathbf{b}}$ where $\widehat{\mathbf{b}}$ is chosen from the column space of A to be as close to $\mathbf{b}$ as possible. Figure 7.5a illustrates the situation that arises when A is a 3×2 matrix with a two-dimensional column space but $\mathbf{b}$ is not in *col* A.



(a)



(b)

Figure 7.5 $\widehat{\mathbf{b}}$ *is chosen from the column space of* A *to be as close to* $\mathbf{b}$ *as possible.*

It appears from this diagram that we wish to choose $\widehat{\mathbf{b}}$ to be the orthogonal projection of $\mathbf{b}$ onto *col* A, because $\widehat{\mathbf{b}}$ looks like the vector in *col* A that is closest to $\mathbf{b}$.

This is a good example to show how geometric intuition can point you in the right direction and even suggest a good proof.

Theorem 7.20. Let W be a finite dimensional subspace of an inner product space V. For each $\mathbf{b} \in V$, let $\widehat{\mathbf{b}}$ be the orthogonal projection of $\mathbf{b}$ onto W. For all $\mathbf{w} \in W$ such that $\mathbf{w} \neq \widehat{\mathbf{b}}$,

$$\|\mathbf{b} - \widehat{\mathbf{b}}\| < \|\mathbf{b} - \mathbf{w}\|.$$

That is, $\widehat{\mathbf{b}}$ is the unique vector in W that is closest to $\mathbf{b}$.

Proof Since $\widehat{\mathbf{b}} = proj_W \mathbf{b}$, $\mathbf{b} - \widehat{\mathbf{b}}$ is the component of $\mathbf{b}$ orthogonal to W so $\mathbf{b} - \widehat{\mathbf{b}} \in W^\perp$. Let $\mathbf{w} \in W$ such that $\mathbf{w} \neq \widehat{\mathbf{b}}$. Clearly

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \widehat{\mathbf{b}}) + (\widehat{\mathbf{b}} - \mathbf{w}).$$

Since $\widehat{\mathbf{b}}, \mathbf{w} \in W$, their difference $\widehat{\mathbf{b}} - \mathbf{w} \in W$, and since $\mathbf{b} - \widehat{\mathbf{b}} \in W^\perp$, the two vectors $\mathbf{b} - \widehat{\mathbf{b}}$ and $\widehat{\mathbf{b}} - \mathbf{w}$ are orthogonal (see Figure 1.3b). By the Pythagorean theorem

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \widehat{\mathbf{b}}\|^2 + \|\widehat{\mathbf{b}} - \mathbf{w}\|^2.$$

Since $\mathbf{w} \neq \widehat{\mathbf{b}}$, $\|\widehat{\mathbf{b}} - \mathbf{w}\|^2 > 0$. Thus

$$\|\widehat{\mathbf{b}} - \mathbf{w}\|^2 > \|\mathbf{b} - \widehat{\mathbf{b}}\|^2$$

which, in turn, implies

$$\|\mathbf{b} - \widehat{\mathbf{b}}\| < \|\mathbf{b} - \mathbf{w}\|.$$

■

Definition 7.16. Let $A\mathbf{x} = \mathbf{b}$ be an $m \times n$ system of equations. A **least squares solution** to $A\mathbf{x} = \mathbf{b}$ is a solution to the system $A\mathbf{x} = \widehat{\mathbf{b}}$ where $\widehat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto the column space of A in Euclidean m-space.

If the columns of A form an orthgonal set, then the nonzero columns of A form an orthogonal basis for $col\ A$. We could then find $\widehat{\mathbf{b}}$ and solve $A\mathbf{x} = \widehat{\mathbf{b}}$. If the columns of A are not mutually orthogonal, we could still find a basis for $col\ A$ in the columns of A and use the Gram-Schmidt process to find an orthogonal basis for $col\ A$. Though this approach works, the Gram-Schmidt process is very tedious. Fortunately, there is a better way.

Recall that the column space of A equals the row space of A^T ($col\ A = row\ A^T$), and that the null space of A^T is the orthogonal complement of the row space of A^T ($null\ A^T = (row\ A^T)^\perp$), so $null\ A^T = (col\ A)^\perp$.

Theorem 7.21. Let $A\mathbf{x} = \mathbf{b}$ be an $m \times n$ system of equations. A vector $\widehat{\mathbf{x}}$ is a least squares solution to $A\mathbf{x} = \mathbf{b}$ if and only if $\widehat{\mathbf{x}}$ is a solution to the system $A^T A\mathbf{x} = A^T\mathbf{b}$.

Proof Let $\widehat{\mathbf{b}}$ be the orthogonal projection of $\mathbf{b}$ onto *col* A.

$$\begin{aligned}
\text{The vector } \widehat{\mathbf{x}} \text{ is a least} \\
\text{squares solution to } A\mathbf{x} = \mathbf{b} &\iff A\widehat{\mathbf{x}} = \widehat{\mathbf{b}} \\
&\iff \mathbf{b} - A\widehat{\mathbf{x}} = \mathbf{b} - \widehat{\mathbf{b}} \\
&\iff \mathbf{b} - A\widehat{\mathbf{x}} \in (col\ A)^{\perp} \text{ (since } \mathbf{b} - \widehat{\mathbf{b}} \in (col\ A)^{\perp}) \\
&\iff \mathbf{b} - A\widehat{\mathbf{x}} \in null\ A^T \text{(since } null\ A^T = (col\ A)^{\perp}) \\
&\iff A^T(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0} \\
&\iff A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}.
\end{aligned}$$

∎

Definition 7.17. The system $A^T A\mathbf{x} = A^T\mathbf{b}$ is called the **normal system** or the **normal equations** of the system $A\mathbf{x} = \mathbf{b}$.

Example 7.14

Find the least squares solution to the system

$$\begin{array}{rcrcr} x & + & 2y & = & 1 \\ 3x & - & y & = & 2 \\ 2x & + & y & = & 3 \end{array}$$

Solution The matrix form is

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

The normal system is

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} 14 & 1 \\ 1 & 6 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 13 \\ 3 \end{bmatrix}.$$

This normal system is probably most easily solved using Cramer's rule.

$$x = \frac{\begin{vmatrix} 13 & 1 \\ 3 & 6 \end{vmatrix}}{\begin{vmatrix} 14 & 1 \\ 1 & 6 \end{vmatrix}} = \frac{78 - 3}{84 - 1} = \frac{75}{83}$$

$$y = \frac{\begin{vmatrix} 14 & 13 \\ 1 & 3 \end{vmatrix}}{\begin{vmatrix} 14 & 1 \\ 1 & 6 \end{vmatrix}} = \frac{42 - 13}{84 - 1} = \frac{29}{83}$$

So the least squares solution to the system is $x = \frac{75}{83}, y = \frac{29}{83}$. We can check to see whether it is a solution to the original system,

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 75/83 \\ 29/83 \end{bmatrix} = \frac{1}{83}\begin{bmatrix} 75+58 \\ 225-29 \\ 150+29 \end{bmatrix} = \frac{1}{83}\begin{bmatrix} 133 \\ 196 \\ 179 \end{bmatrix},$$

and see that it is not. However, it can be checked that $\frac{1}{83}\begin{bmatrix} 133 \\ 196 \\ 179 \end{bmatrix}$ is the projection of $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ onto the column space of $\begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & -1 \end{bmatrix}$.

The term "least squares," in the least squares solution to $A\mathbf{x} = \mathbf{b}$, comes from the fact that $\widehat{\mathbf{b}} = proj_{col\ A}\mathbf{b}$ is the closest vector in the column space of A to $\mathbf{b}$ under the Euclidean inner product.

Suppose $\widehat{\mathbf{x}}$ is a least squares solution to the system $A\mathbf{x} = \mathbf{b}$. Then $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$. For any $\mathbf{x} \in \mathbb{R}^n$ that is not a least squares solution, let $\mathbf{w} = A\mathbf{x}$. By Theorem 7.20 we know that $\|\mathbf{b} - \widehat{\mathbf{b}}\| < \|\mathbf{b} - \mathbf{w}\|$. Under the Euclidean inner product this translates to

$$\sqrt{(b_1 - \widehat{b}_1)^2 + \cdots + (b_m - \widehat{b}_m)^2} < \sqrt{(b_1 - w_1)^2 + \cdots + (b_m - w_m)^2}.$$

Squaring both sides gives

$$(b_1 - \widehat{b}_1)^2 + \cdots + (b_m - \widehat{b}_m)^2 < (b_1 - w_1)^2 + \cdots + (b_m - w_m)^2.$$

So the sum of these squared differences is least when $\widehat{\mathbf{x}}$ is used as an approximate solution to $A\mathbf{x} = \mathbf{b}$, hence "least squares."

Linear Regression and Least Squares

Imagine some scientific experiment that produces n ordered pairs $(x_1, y_1), \cdots, (x_n, y_n)$ with the x coordinate independent and the y coordinate dependent. Suppose that when graphed, the points tend to line up. Perhaps they don't line up exactly due to some round-off or other experimental error. Still, we wish to use this experimental data to find m and b in an equation $y = mx + b$ of a line that comes close to these data points. Statisticians use the term **linear regression** to describe a variety of techniques for finding such approximating lines, but the method of least squares is the method most used to accomplish this task.

If the points lined up exactly in a nonvertical line then there are real numbers m and b that would make each of the following equations true.

$$\begin{aligned} mx_1 + b &= y_1 \\ &\vdots \\ mx_n + b &= y_n \end{aligned}$$

Since the ordered pairs $(x_1, y_1), \cdots, (x_n, y_n)$ are known but m and b are unknown in this case, these equations form an $n \times 2$ system of linear equations with matrix form

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

If the points truly do line up, then this system could be solved for m and b, though it would be easier to find m and b using just two of the points. If, however, the points do not exactly line up, then the system is inconsistent and if you just used two points to determine m and b, your answers would vary depending on which pair of points you choose.

You could, however, use the method of least squares to solve the above system. It produces values for m and b that minimize the sum of the squares of the vertical distances between the line $y = mx + b$ and the data points, as illustrated in Figure 7.6. This is sometimes called the **line of best fit**.

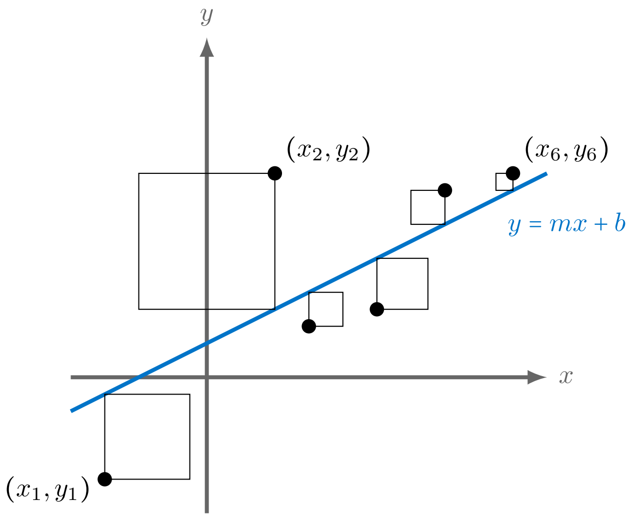


Figure 7.6 *Values of m and b in the line $y = mx + b$ minimize the sum of the squared vertical distances between the line and the data points $(x_1, y_1), \cdots, (x_6, y_6)$.*

Let $\mathbf{x}$ and $\mathbf{y}$ be the column vectors of all n of the x and y coordinates from the n data points and let $\mathbf{1}$ be the column vector of n 1's. Let A be the $n \times 2$ matrix $\begin{bmatrix} \mathbf{x} & \mathbf{1} \end{bmatrix}$. We wish to find the least squares solution to the system

$$A \begin{bmatrix} m \\ b \end{bmatrix} = \mathbf{y}.$$

Its normal system is $A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \mathbf{y}$ which can also be seen as

$$\begin{bmatrix} \mathbf{x} \cdot \mathbf{x} & \mathbf{x} \cdot \mathbf{1} \\ \mathbf{1} \cdot \mathbf{x} & \mathbf{1} \cdot \mathbf{1} \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{x} \cdot \mathbf{y} \\ \mathbf{1} \cdot \mathbf{y} \end{bmatrix}.$$

Using Cramer's rule, we solve for m and b:

$$m = \frac{\begin{vmatrix} \mathbf{x}\cdot\mathbf{y} & \mathbf{x}\cdot\mathbf{1} \\ \mathbf{1}\cdot\mathbf{y} & \mathbf{1}\cdot\mathbf{1} \end{vmatrix}}{\begin{vmatrix} \mathbf{x}\cdot\mathbf{x} & \mathbf{x}\cdot\mathbf{1} \\ \mathbf{1}\cdot\mathbf{x} & \mathbf{1}\cdot\mathbf{1} \end{vmatrix}} = \frac{(\mathbf{x}\cdot\mathbf{y})(\mathbf{1}\cdot\mathbf{1}) - (\mathbf{x}\cdot\mathbf{1})(\mathbf{1}\cdot\mathbf{y})}{(\mathbf{x}\cdot\mathbf{x})(\mathbf{1}\cdot\mathbf{1}) - (\mathbf{x}\cdot\mathbf{1})(\mathbf{1}\cdot\mathbf{x})} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{\begin{vmatrix} \mathbf{x}\cdot\mathbf{x} & \mathbf{x}\cdot\mathbf{y} \\ \mathbf{1}\cdot\mathbf{x} & \mathbf{1}\cdot\mathbf{y} \end{vmatrix}}{\begin{vmatrix} \mathbf{x}\cdot\mathbf{x} & \mathbf{x}\cdot\mathbf{1} \\ \mathbf{1}\cdot\mathbf{x} & \mathbf{1}\cdot\mathbf{1} \end{vmatrix}} = \frac{(\mathbf{x}\cdot\mathbf{x})(\mathbf{1}\cdot\mathbf{y}) - (\mathbf{x}\cdot\mathbf{y})(\mathbf{1}\cdot\mathbf{x})}{(\mathbf{x}\cdot\mathbf{x})(\mathbf{1}\cdot\mathbf{1}) - (\mathbf{x}\cdot\mathbf{1})(\mathbf{1}\cdot\mathbf{x})} = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Though the formulas get too messy for Cramer's rule, this same idea can be used with two independent variables and one dependent variable to find a plane of best fit. Indeed, given data with k independent variables and one dependent variable, we can use the method of least squares to find the hyperplane of best fit with the equation

$$y = m_1 x_1 + \cdots + m_k x_k + b.$$

Though exceptions exist and are easily constructed, we have good reason to expect that an $m \times n$ matrix A with entries that are somehow randomly selected has columns that span $\mathbb{R}^m$ if $m \le n$ and that are linearly independent if $m \ge n$. This tends to produce systems, $A\mathbf{x} = \mathbf{b}$, with infinitely many solutions if the system is underdetermined ($m < n$, i.e. A is short and fat), with one unique solution if the system is square ($m = n$), and with no solution if the system is overdetermined ($m > n$, i.e. A is tall and skinny).

We do not go into those reasons here, but because of them it is most frequently the case that the method of least squares is applied to overdetermined systems. It is frequently the case, therefore, when seeking a least squares solution to a system $A\mathbf{x} = \mathbf{b}$, that the columns of A are linearly independent.

Theorem 7.22. Let A be an $m \times n$ matrix. The matrix $A^T A$ is invertible if and only if the columns of A are linearly independent.

Proof Note that $A^T A$ is an $n \times n$ matrix. Note too that

$$\begin{aligned} A^T A\mathbf{x} = \mathbf{0} \quad &\iff \quad A\mathbf{x} \in \mathit{null}\ A^T \\ &\iff \quad A\mathbf{x} \in \left(\mathit{row}\ A^T\right)^{\perp} \\ &\iff \quad A\mathbf{x} \in (\mathit{col}\ A)^{\perp}. \end{aligned}$$

But $A\mathbf{x} \in \mathit{col}\ A$ and $(\mathit{col}\ A) \cap (\mathit{col}\ A)^{\perp} = \{\mathbf{0}\}$, so $A^T A\mathbf{x} = \mathbf{0} \iff A\mathbf{x} = \mathbf{0}$.

The matrix $A^T A$ is invertible $\iff$ $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution
$\iff$ $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (from above)
$\iff$ the columns of A are linearly independent.

■

The implications of Theorem 7.22 on least squares solutions are immediate.

Corollary 7.23. Let A be an $m \times n$ matrix. The system $A\mathbf{x} = \mathbf{b}$ has one unique least squares solution if and only if the columns of A are linearly independent.

Since all linear systems have at least one least squares solution, the system $A\mathbf{x} = \mathbf{b}$ has infinitely many least squares solutions if and only if the columns of A are linearly dependent. But because mostly we use the method of least squares on overdetermined systems ($m > n$) we can reasonably expect to have one unique least squares solution most of the time.

We end this section with a method for finding orthogonal projection matrices.

Definition 7.18. Let W be a subspace of Euclidean m space. An $m \times m$ matrix P is an **orthogonal projection matrix** onto W if $P\mathbf{v}$ equals the orthogonal projection of $\mathbf{v}$ onto W for all $\mathbf{v} \in \mathbb{R}^m$.

Let $\widehat{\mathbf{x}}$ be a least squares solution to the $m \times n$ system $A\mathbf{x} = \mathbf{b}$. Recall that this means $\widehat{\mathbf{x}}$ is a solution to the system $A\mathbf{x} = \widehat{\mathbf{b}}$ where $\widehat{\mathbf{b}} = proj_{col\ A}\mathbf{b}$ and $\widehat{\mathbf{x}}$ is a solution to the normal system $A^TA\mathbf{x} = A^T\mathbf{b}$.

Theorem 7.24. Let A be an $m \times n$ matrix. If the columns of A are linearly independent, then the matrix $P = A\left(A^TA\right)^{-1}A^T$ is an orthogonal projection matrix onto the column space of A.

Proof Let $\mathbf{v} \in \mathbb{R}^n$. We show that $P\mathbf{v} = proj_{col\ A}\mathbf{v}$. Since the columns of A are linearly independent, the system $A\mathbf{x} = \mathbf{v}$ has one unique least squares solution $\widehat{\mathbf{x}}$ and $A\widehat{\mathbf{x}} = \widehat{\mathbf{v}}$ where $\widehat{\mathbf{v}} = proj_{col\ A}\mathbf{v}$ and $A^TA\widehat{\mathbf{x}} = A^T\mathbf{v}$. Since the columns of A are linearly independent, A^TA is invertible so $\widehat{\mathbf{x}} = \left(A^TA\right)^{-1}A^T\mathbf{v}$. Multiplying both sides by A, we get

$$A\widehat{\mathbf{x}} = A\left(A^TA\right)^{-1}A^T\mathbf{v},$$

and since $A\widehat{\mathbf{x}} = \widehat{\mathbf{v}} = proj_{col\ A}\mathbf{v}$,

$$proj_{col\ A}\mathbf{v} = A\left(A^TA\right)^{-1}A^T\mathbf{v}.$$

Thus $P = A\left(A^TA\right)^{-1}A^T$ is an orthogonal projection matrix onto $col\ A$. ■

Example 7.15

Let W be the plane through the origin $3x + 2y + z = 0$ in Euclidean three space. Find a projection matrix onto W.

Solution Since $z = -3x - 2y$, we let $x = s$ and $y = t$ so

$$\begin{array}{rcrcr} x & = & s & & \\ y & = & & & t \\ z & = & -3s & - & 2t \end{array}.$$

Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -3 & -2 \end{bmatrix}$. The plane W is the column space of A, and

$$A^T A = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 6 \\ 6 & 5 \end{bmatrix}.$$

So

$$\left(A^T A\right)^{-1} = \frac{1}{14} \begin{bmatrix} 5 & -6 \\ -6 & 10 \end{bmatrix}.$$

Thus

$$\begin{aligned} P &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -3 & -2 \end{bmatrix} \left(\frac{1}{14} \begin{bmatrix} 5 & -6 \\ -6 & 10 \end{bmatrix} \right) \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -2 \end{bmatrix} \\ &= \frac{1}{14} \begin{bmatrix} 5 & -6 & -3 \\ -6 & 10 & -2 \\ -3 & -2 & 13 \end{bmatrix}. \end{aligned}$$

Problem Set 7.5

1. Find the normal system (in simplified matrix form) for each of the following systems.

(a) $\begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 1 & -1 \\ -2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 1 \end{bmatrix}$

(c) $\begin{array}{rcrcr} x & + & y & = & 3 \\ 2x & + & y & = & 2 \\ x & + & 2y & = & -1 \\ x & - & y & = & 1 \end{array}$

(d) $\begin{array}{rcrcrcr} x & + & y & - & z & = & 1 \\ 2x & & & + & z & = & 0 \\ x & - & y & + & z & = & -1 \\ x & + & 2y & - & z & = & 1 \end{array}$

2. Find the least squares solution to each system given in Exercise 1.

3. Find the least squares solution to the system $A\mathbf{x} = \mathbf{b}$ where $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \\ 1 & 3 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. *(Hint: Use Maple to help solve the normal system.)*

4. Find the line of best fit for each of the following lists of points (x, y).

(a) $(1,2)$, $(2,4)$, $(3,5)$

(b) $(-1,4)$, $(0,2)$, $(2,1)$, $(5,0)$

5. Find the plane of best fit for the following list of points (x, y, z): $(1,1,0)$, $(1,0,1)$, $(0,1,2)$, $(1,2,-1)$.

6. Let $W = span\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \right\}$ in Euclidean 3-space. Find the orthogonal projection matrix onto W.

7. Find the orthogonal projection matrix onto the plane $x - y + 2z = 0$ in Euclidean 3-space.

8. Use the orthogonal projection matrix from Exercise 7 to find the distance between the point $(1,2,3)$ and the plane $x - y + 2z = 0$.

9. True or false.

(a) The least squares solution to $A\mathbf{x} = \mathbf{b}$ is the vector $\mathbf{x}$ that is as close to $\mathbf{b}$ as possible.

(b) The least squares solution to $A\mathbf{x} = \mathbf{b}$ is the vector $\mathbf{x}$ that makes $A\mathbf{x}$ as close to $\mathbf{b}$ as possible.

(c) Any solution to $A^T A\mathbf{x} = A^T\mathbf{b}$ is a least squares solution to $A\mathbf{x} = \mathbf{b}$.

(d) The vector $\hat{\mathbf{x}}$ is a least squares solution to $A\mathbf{x} = \mathbf{b}$ if and only if $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ onto the column space of A.

(e) The normal system of the system $A\mathbf{x} = \mathbf{b}$ is inconsistent if the column of A are linearly dependent.

(f) If the normal system of the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions, then the columns of A are linearly dependent.

(g) Normal systems are always consistent.

(h) If $A\mathbf{x} = \mathbf{b}$ is consistent then the solution set to the system and the set of least squares solutions to the system are identical.

7.6 Inner Product Space Isomorphisms and Orthogonal Matrices

Recall that a linear transformation is a function between two vector spaces that preserves vector addition and multiplication by scalars. An isomorphism is a linear transformation that is both one to one and onto. Next, we define an inner product space isomorphism as an isomorphism that also preserves inner products.

Definition 7.19. Let V and W be inner product spaces with inner products $\langle , \rangle_V$ and $\langle , \rangle_W$ respectively. An **inner product space isomorphism** (IPSI) is a one-to-one onto function $T : V \longrightarrow W$ that satisfies the following three properties.
For all $\mathbf{u}, \mathbf{v} \in V$ and for all scalars c,

1. $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
2. $T(c\mathbf{u}) = cT(\mathbf{u})$
3. $\langle T(\mathbf{u}), T(\mathbf{v})\rangle_W = \langle \mathbf{u}, \mathbf{v}\rangle_V$

Theorem 7.25. Let $T : V \longrightarrow W$ be a linear transformation from one inner product space V onto another W, and suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an orthonormal basis for V. The following are equivalent:

(a) T is an inner product space isomorphism.

(b) $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is an orthonormal basis for W.

(c) For all $\mathbf{u} \in V$, $\|T(\mathbf{u})\|_W = \|\mathbf{u}\|_V$.

Proof ((a) $\Longrightarrow$ (b)) If T is an IPSI, then T is an isomorphism, so $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is a basis for W by Theorem 5.11. The following two calculations show that $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is an orthonormal set.

(i). For each $i = 1, \cdots, n$, $\|T(\mathbf{v}_i)\|_W = \sqrt{\langle T(\mathbf{v}_i), T(\mathbf{v}_i)\rangle_W} = \sqrt{\langle \mathbf{v}_i, \mathbf{v}_i\rangle_V} = \|\mathbf{v}_i\|_V = 1$.

(ii). For each $i \neq j$, $\langle T(\mathbf{v}_i), T(\mathbf{v}_j)\rangle_W = \langle \mathbf{v}_i, \mathbf{v}_j\rangle_V = 0$.

Therefore, $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is an orthonormal basis for W.

((b) $\Longrightarrow$ (c)) Let $\mathbf{u} \in V$. Since $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V, there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, and since T is linear we also have $T(\mathbf{u}) = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)$. Since both $\{c_1\mathbf{v}_1, \cdots, c_n\mathbf{v}_n\}$ and $\{c_1T(\mathbf{v}_1), \cdots, c_nT(\mathbf{v}_n)\}$ are orthogonal sets, by Theorem 7.11, Theorem 7.4(c), and the definition of orthonormal set,

$$\begin{aligned} \|\mathbf{u}\|_V^2 &= \|c_1\mathbf{v}_1\|_V^2 + \cdots + \|c_n\mathbf{v}_n\|_V^2 \\ &= c_1^2\|\mathbf{v}_1\|_V^2 + \cdots + c_n^2\|\mathbf{v}_n\|_V^2 \\ &= c_1^2 + \cdots + c_n^2 \end{aligned}$$

and

$$\begin{aligned}\|T(\mathbf{u})\|_W^2 &= \|c_1T(\mathbf{v}_1)\|_W^2 + \cdots + \|c_nT(\mathbf{v}_n)\|_W^2 \\ &= c_1^2\|T(\mathbf{v}_1)\|_W^2 + \cdots + c_n^2\|T(\mathbf{v}_n)\|_W^2 \\ &= c_1^2 + \cdots + c_n^2.\end{aligned}$$

So, $\|T(\mathbf{u})\|_W = \|\mathbf{u}\|_V$.

((c) $\Longrightarrow$ (a)) Let $\mathbf{u} \in \mathit{ker}\ T$. Then $T(\mathbf{u}) = \mathbf{0}$ so $\|\mathbf{u}\|_V = \|T(\mathbf{u})\|_W = 0$. Thus $\mathbf{u} = \mathbf{0}$ and $\mathit{ker}\ T = \{\mathbf{0}\}$. This tells us that T is one-to-one by Theorem 5.10(a) and we are given that T is onto, so T is an isomorphism. To see that T preserves inner product note that in any inner product space, $\langle\mathbf{u},\mathbf{v}\rangle = \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2$ (section 7.1, exercise 10) so

$$\langle\mathbf{u},\mathbf{v}\rangle_V = \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|_V^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|_V^2$$

and

$$\langle T(\mathbf{u}),T(\mathbf{v})\rangle_W = \frac{1}{4}\|T(\mathbf{u})+T(\mathbf{v})\|_W^2 - \frac{1}{4}\|T(\mathbf{u})-T(\mathbf{v})\|_W^2.$$

But since T is linear $T(\mathbf{u}) \pm T(\mathbf{v}) = T(\mathbf{u} \pm \mathbf{v})$ and since T preserves norm,

$$\begin{aligned}\langle T(\mathbf{u}),T(\mathbf{v})\rangle_W &= \frac{1}{4}\|T(\mathbf{u})+T(\mathbf{v})\|_W^2 - \frac{1}{4}\|T(\mathbf{u})-T(\mathbf{v})\|_W^2 \\ &= \frac{1}{4}\|T(\mathbf{u}+\mathbf{v})\|_W^2 - \frac{1}{4}\|T(\mathbf{u}-\mathbf{v})\|_W^2 \\ &= \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|_V^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|_V^2 \\ &= \langle\mathbf{u},\mathbf{v}\rangle_V.\end{aligned}$$

Thus T is an inner product space isomorphism. ■

Theorem 7.25 gives us three equivalent ways of thinking about inner product space isomorphisms. That is, an inner product space isomorphism is a surjective linear transformation between inner product spaces that

(a) preserves inner product,

(b) maps an orthonormal basis to an orthonormal basis, and

(c) preserves norm.

We recall that we can construct a linear transformation between two finite-dimensional vector spaces by mapping a basis of the domain to elements in the codomain and then extending linearly (Theorem 5.2, Definition 5.5). Part (b), therefore, tells us how to construct an inner product space isomorphism between finite-dimensional inner product spaces. Simply map an orthonormal basis of the domain onto an orthonormal basis of the codomain in a one-to-one manner (they must be the same dimension to do this) and then extend linearly.

We have discussed coordinate vectors earlier. The assignment of coordinate vectors to an n-dimensional vector space V involves creating a mapping $T : V \longrightarrow \mathbb{R}^n$ by taking an ordered basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of V and assigning $T(\mathbf{v}_i) = \mathbf{e}_i$ from the standard basis $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ and then extending linearly. Thus, $T : V \longrightarrow \mathbb{R}^n$ given by $T(\mathbf{u}) = [\mathbf{u}]_{\mathcal{B}}$ is an isomorphism from V to $\mathbb{R}^n$. If, in addition, V is an inner product space, then Theorem 7.25 implies that T is an inner product space isomorphism between V and Euclidean n-space if and only if $\mathcal{B}$ is an orthonormal basis since $\mathcal{S}_n$ is an orthonormal basis of $\mathbb{R}^n$ under the dot product. The following is an immediate consequence of Theorem 7.25.

Theorem 7.26. Let V be an inner product space with $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ an ordered orthonormal basis of V. For all $\mathbf{u}, \mathbf{v} \in V$,

(a) $\langle \mathbf{u}, \mathbf{v} \rangle_V = [\mathbf{u}]_{\mathcal{B}} \cdot [\mathbf{v}]_{\mathcal{B}}$.

(b) $\|\mathbf{u}\|_V = \| [\mathbf{u}]_{\mathcal{B}} \|$.

(c) $\|\mathbf{u} - \mathbf{v}\|_V = \| [\mathbf{u}]_{\mathcal{B}} - [\mathbf{v}]_{\mathcal{B}} \|$.

Here we are most interested in the case where V is Euclidean n-space. In this case, finding coordinate vectors amounts to performing a change of basis. Recall that if $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an ordered basis of $\mathbb{R}^n$ and P is the $n \times n$ matrix $P = [\mathbf{v}_1 \cdots \mathbf{v}_n]$, then for all $\mathbf{u} \in \mathbb{R}^n$, $P[\mathbf{u}]_{\mathcal{B}} = \mathbf{u}$ and $P^{-1}\mathbf{u} = [\mathbf{u}]_{\mathcal{B}}$. the matrices P and P^{-1} are called change of basis matrices from $\mathcal{B}$ to $\mathcal{S}_n$ and from $\mathcal{S}_n$ to $\mathcal{B}$ respectively.

Now if $\mathcal{B}$ is an orthonormal basis for Euclidean n-space, then P and P^{-1} are very special matrices indeed.

Theorem 7.27. Suppose U is an $n \times n$ matrix. The following are equivalent.

(a) The columns of U form an orthonormal basis for Euclidean n-space.

(b) $U^T = U^{-1}$.

(c) The rows of U form an orthonormal basis for Euclidean n-space.

Proof Let $\mathbf{u}_1, \cdots, \mathbf{u}_n$ be the columns of U, so $U = [\mathbf{u}_1 \cdots \mathbf{u}_n]$. Note that

$$U^T U = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \vdots & & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \cdots & \mathbf{u}_n \cdot \mathbf{u}_n \end{bmatrix}.$$

Clearly, the columns of U form an orthonormal basis for Euclidean n-space if and only if $U^T U = I_n$. To see that part (a) is equivalent to part (b), $U^T = U^{-1} \iff U^T U = I_n \iff$ the columns of U form an orthonormal basis for Euclidean n-space. To see that part (c) is equivalent to part (b), $U^T = U^{-1} \iff UU^T = I_n \iff$ the rows of U form an orthonormal basis for Euclidean n-space. ■

Definition 7.20. An $n \times n$ matrix U is called an **orthogonal matrix** if $U^T = U^{-1}$.

In light of Theorem 7.27, it would seem to be more appropriate to name a matrix like this an orthonormal matrix. Traditionally, these matrices are called orthogonal matrices, so we stick with this tradition.

Example 7.16

The following are all examples of orthogonal matrices:

$$\begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{3}} \\ \frac{2}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix}$$

Example 7.17 Rotations and Reflections in $\mathbb{R}^2$

It is easy enough to characterize all orthogonal 2×2 matrices. Let U be an arbitrary orthogonal 2×2 matrix. The first column can be any unit vector. Since they all lie on the unit circle, they all have the form $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ where θ represents the angle they make with the positive x axis.

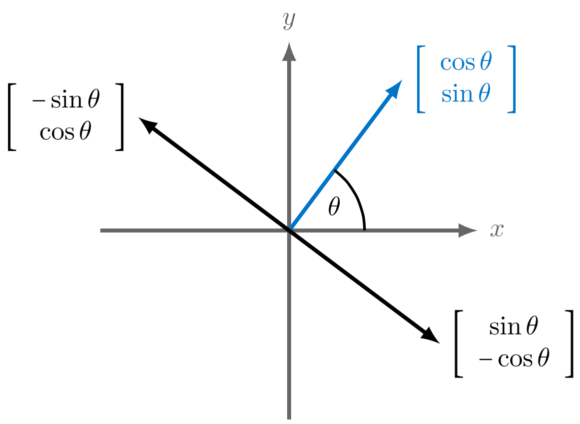


Figure 7.7 *There are two unit vectors orthogonal to an arbitrary unit vector in* $\mathbb{R}^2$.

Once that column of U is selected, there are only two choices left for the second column since it too must be a unit vector and it must also be orthogonal to the first. One such vector comes from rotating $\mathbf{e}_2$ by the angle θ, $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ and the other is its negative. So,

$$U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad U = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

The first you recognize as the rotation matrix R_θ. The second is a reflection matrix that reflects a vector across the line through the origin that makes an angle of $\theta/2$ with the positive x axis.

Theorem 7.28. Let A and B be orthogonal $n \times n$ matrices.

(a) A^{-1} is an orthogonal matrix.

(b) AB is an orthogonal matrix.

(c) $\det A = \pm 1$.

Proof

(a) To prove A^{-1} is orthogonal from the definition we show $\left(A^{-1}\right)^T = \left(A^{-1}\right)^{-1}$. But

$$\begin{aligned} \left(A^{-1}\right)^T &= \left(A^T\right)^{-1} \quad (\textit{Theorem } 1.12(c)) \\ &= \left(A^{-1}\right)^{-1} \quad \text{since } A \text{ is orthogonal.} \end{aligned}$$

(b) Since inverse and transpose have the socks-shoes property, $(AB)^T = B^T A^T = B^{-1}A^{-1} = (AB)^{-1}$. Therefore, AB is orthogonal.

(c) Since $\det A = \det A^T$ and $\det A^T A = \left(\det A^T\right)(\det A)$, $1 = \det I = \det\left(A^T A\right) = \left(\det A^T\right)(\det A) = (\det A)^2$. Since $(\det A)^2 = 1$, $\det A = \pm 1$.

■

Since the mapping $\mathbf{x} \mapsto A\mathbf{x}$ can be thought of as a change of basis mapping when A is square and invertible, and change of basis mappings are isomorphisms, the next theorem follows immediately from Theorem 7.25.

Theorem 7.29. The following are equivalent.

(a) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(b) A is an orthogonal matrix.

(c) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ under the Euclidean norm for every $\mathbf{x} \in \mathbb{R}^n$.

Example 7.18

The sets

$$\mathcal{A} = \left\{ \begin{bmatrix} 2/3 \\ 2/3 \\ -1/3 \end{bmatrix}, \begin{bmatrix} 2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \begin{bmatrix} -1/3 \\ 2/3 \\ 2/3 \end{bmatrix} \right\} \quad \text{and} \quad \mathcal{B} = \left\{ \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \right\}$$

are orthonormal bases for Euclidean 3-space (verify!). Find the change of basis matrix U such that $U[\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{A}}$ for all $\mathbf{v} \in \mathbb{R}^3$.

Solution We find U in two ways.

▶ Let

$$A = \begin{bmatrix} 2/3 & 2/3 & -1/3 \\ 2/3 & -1/3 & 2/3 \\ -1/3 & 2/3 & 2/3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}.$$

Then A and B are change of basis matrices such that $A[\mathbf{v}]_{\mathcal{A}} = \mathbf{v}$ and $B[\mathbf{v}]_{\mathcal{B}} = \mathbf{v}$. So $A^T\mathbf{v} = [\mathbf{v}]_{\mathcal{A}}$ and $B^T\mathbf{v} = [\mathbf{v}]_{\mathcal{B}}$. Thus $A^T B[\mathbf{v}]_{\mathcal{B}} = A^T(B[\mathbf{v}]_{\mathcal{B}}) = A^T\mathbf{v} = [\mathbf{v}]_{\mathcal{A}}$ so $A^T B$ is the change of basis matrix we seek.

$$U = A^T B = \begin{bmatrix} 2/3 & 2/3 & -1/3 \\ 2/3 & -1/3 & 2/3 \\ -1/3 & 2/3 & 2/3 \end{bmatrix}\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \end{bmatrix}.$$

▶ Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ represent the first, second, and third vectors in $\mathcal{A}$, so $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Similarly for $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$. Using Theorem 5.21 we see that

$$U = [[\mathbf{b}_1]_{\mathcal{A}}, [\mathbf{b}_2]_{\mathcal{A}}, [\mathbf{b}_3]_{\mathcal{A}}].$$

Since $\mathcal{A}$ is an orthonormal basis,

$$\begin{aligned}
\mathbf{b}_1 = (\mathbf{b}_1 \cdot \mathbf{a}_1)\mathbf{a}_1 + (\mathbf{b}_1 \cdot \mathbf{a}_2)\mathbf{a}_2 + (\mathbf{b}_1 \cdot \mathbf{a}_3)\mathbf{a}_3 &= \frac{1}{\sqrt{3}}\mathbf{a}_1 + \frac{1}{\sqrt{3}}\mathbf{a}_2 + \frac{1}{\sqrt{3}}\mathbf{a}_3 \\
\mathbf{b}_2 = (\mathbf{b}_2 \cdot \mathbf{a}_1)\mathbf{a}_1 + (\mathbf{b}_2 \cdot \mathbf{a}_2)\mathbf{a}_2 + (\mathbf{b}_2 \cdot \mathbf{a}_3)\mathbf{a}_3 &= \frac{1}{\sqrt{2}}\mathbf{a}_1 + 0\mathbf{a}_2 - \frac{1}{\sqrt{2}}\mathbf{a}_3 \\
\mathbf{b}_3 = (\mathbf{b}_3 \cdot \mathbf{a}_1)\mathbf{a}_1 + (\mathbf{b}_3 \cdot \mathbf{a}_2)\mathbf{a}_2 + (\mathbf{b}_3 \cdot \mathbf{a}_3)\mathbf{a}_3 &= -\frac{1}{\sqrt{6}}\mathbf{a}_1 + \frac{2}{\sqrt{6}}\mathbf{a}_2 - \frac{1}{\sqrt{6}}\mathbf{a}_3
\end{aligned}$$

So,

$$[\mathbf{b}_1]_{\mathcal{A}} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, [\mathbf{b}_2]_{\mathcal{A}} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, [\mathbf{b}_3]_{\mathcal{A}} = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}$$

yielding

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \end{bmatrix}.$$

Problem Set 7.6

1. Let $\mathcal{B} = \left\{\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}\right\}$, $\mathcal{B}' = \left\{\begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}, \begin{bmatrix} -3/5 \\ 4/5 \end{bmatrix}\right\}$, and $\mathcal{S}_2 = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$

(a) Verify that $\mathcal{B}$, $\mathcal{B}'$, and $\mathcal{S}_2$ are orthonormal bases for Euclidean 2-space.

(b) Find the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}_2$.

(c) Find the change of basis matrix from $\mathcal{S}_2$ to $\mathcal{B}$.

(d) Find the change of basis matrix from $\mathcal{B}'$ to $\mathcal{S}_2$.

(e) Find the change of basis matrix from $\mathcal{S}_2$ to $\mathcal{B}'$.

(f) Find the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$.

(g) Find the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$.

2. Repeat Exercise 1 with the following three sets. Replace $\mathcal{S}_2$ with $\mathcal{S}_3$.

$$\mathcal{B} = \left\{ \frac{1}{3}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \frac{1}{15}\begin{bmatrix} 2 \\ 11 \\ -10 \end{bmatrix}, \frac{1}{15}\begin{bmatrix} 14 \\ 2 \\ 5 \end{bmatrix} \right\},\ \mathcal{B}' = \left\{ \frac{1}{5}\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}, \frac{1}{25}\begin{bmatrix} -12 \\ 9 \\ 20 \end{bmatrix}, \frac{1}{25}\begin{bmatrix} -16 \\ 12 \\ -15 \end{bmatrix} \right\},$$

and $\mathcal{S}_3 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$

Any ordered orthonormal basis $\mathcal{B} = \{\mathbf{b}_1 \ldots \mathbf{b}_n\}$ of Euclidean n-space can be used to generate a rectangular coordinate system in $\mathbb{R}^n$. As noted earlier, the corresponding $n \times n$ matrix $U = [\mathbf{b}_1 \ldots \mathbf{b}_n]$ is an orthogonal matrix, so $\det(U) = \pm 1$ (Theorem 7.28(c)). This tells us that the ordered orthonormal bases of Euclidean n-space can be divided into two groups, those for which their corresponding orthogonal matrix has a determinant of 1, and those for which this determinant is -1. An ordered orthonormal basis from one group can easily be altered to form one from the other group simply be swapping the positions of two vectors in the order of the ordered basis or by replacing one vector in the basis with its negative.

In $\mathbb{R}^3$ this distinction manifests itself geometrically by producing right-handed and left-handed rectangular coordinate systems. A right-handed coordinate system is produced if the thumb of your right hand points in the direction of the third basis vector when your curled fingers move from the tip of the first basis vector at the knuckles 90° to the tip of the second vector at the finger tips. The standard ordered basis $\mathcal{S}_3 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ generates a right-handed coordinate system and $\det(I_3) = 1$, but $\{\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3\}$ generates a left-handed coordinate system and $det[\mathbf{e}_2\mathbf{e}_1\mathbf{e}_3] = -1$.

3. Identify the following ordered orthonormal bases of Euclidean 3-space as generating right-handed or left-handed coordinate systems.

(a) $\mathcal{B}_1 = \left\{ \frac{1}{7}\begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \frac{1}{7}\begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix}, \frac{1}{7}\begin{bmatrix} -6 \\ 2 \\ 3 \end{bmatrix} \right\}$

(b) $\mathcal{B}_2 = \left\{ \frac{1}{7}\begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \frac{1}{7}\begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix}, \frac{1}{7}\begin{bmatrix} 6 \\ -2 \\ -3 \end{bmatrix} \right\}$

(c) $\mathcal{B}_3 = \left\{ \frac{1}{7}\begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \frac{1}{7}\begin{bmatrix} -6 \\ 2 \\ 3 \end{bmatrix}, \frac{1}{7}\begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix} \right\}$

4. For a nonzero vector $\mathbf{n}$ in Euclidean 3-space, the Householder matrix, $I_3 - \frac{2}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}\mathbf{n}^T$, is an easy way to construct a standard reflection matrix across the plane $\mathbf{n}\cdot\mathbf{x} = 0$. Let $Q = I_3 - \frac{2}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}\mathbf{n}^T$.

(a) Show that Q is a symmetric matrix.

(b) Show that Q is an orthogonal matrix.

(c) Show that $Q\mathbf{n} = -\mathbf{n}$.

(d) Show that $Q\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in the plane $\mathbf{n}\cdot\mathbf{x} = 0$.

(e) Use parts (c) and (d) of this exercise to identify two of the eigenvalues of Q.

(f) Fill in the blanks in the following statement:
The eigenvalue identified in part (c) must have geometric multiplicity of at least __________, thus its algebraic multiplicity must be at least __________.

(g) Fill in the blanks in the following statement:
The eigenvalue identified in part (d) must have geometric multiplicity of at least __________, thus its algebraic multiplicity must be at least __________.

(h) Counting algebraic multiplicities, what is the maximum number of eigenvalues a 3×3 matrix can have, and how many of them have you identified above for Q?

(i) What are the eigenvalues of Q, and what are their algebraic and geometric multiplicities?

(j) What is the determinant of Q, and do the columns of Q generate a right-handed or a left-handed coordinate system for Euclidean 3-space? *(Hint: Use Theorem 6.12.)*

(k) Find the standard matrix that reflects a vector in Euclidean 3-space orthogonally across the plane $x + 2y + 3z = 0$.

5. A rotation function on Euclidean 3-space is a function that maps each vector $\mathbf{x}$ to the vector obtained by rotating $\mathbf{x}$ by a fixed angle θ about a fixed line through the origin (called the axis of rotation). Building on Definition 5.3, the rotation function that rotates each vector $\mathbf{x}$ counterclockwise (when looking toward the origin from a perspective on the positive x axis) by an angle θ about the x axis is, in fact, a matrix transformation, and its standard matrix is

$$R_\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

Rotation matrices that describe rotations around other axes can be obtained from this one by performing a similarity transformation (change of basis) on R_θ. Special care must be taken to accomplish this. Suppose you wish to find a rotation matrix A that rotates each vector $\mathbf{x}$ counterclockwise by an angle θ around the axis $span\{\mathbf{v}_0\}$ for some $\mathbf{v}_0 \neq 0$ from the perspective of looking towards the origin from a positive multiple of $\mathbf{v}_0$. Choose an ordered orthonormal basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ as follows:

(a) The vector $\mathbf{b}_1$ is a positive multiple of $\mathbf{v}_0$.

(b) The ordered orthonormal basis $\mathcal{B}$ must generate a right-handed coordinate system.

(c) Let $U = [\mathbf{b}_1 \mathbf{b}_2 \mathbf{b}_3]$.

Since U is an orthogonal matrix, $U^T = U^{-1}$, so $U^T A U = R_\theta$, thus $A = U R_\theta U^T$.

(a) Show that R_θ is an orthogonal matrix.

(b) Find the determinant of R_θ.

(c) Show that A is an orthogonal matrix.

(d) Find the determinant of A.

(e) Find the rotation matrix that rotates the vectors in Euclidean 3-space counterclockwise (when viewed from a perspective on the positive x axis looking toward the origin) about the x axis by 120°.

(f) Find the rotation matrix that rotates the vectors in Euclidean 3-space counterclockwise (when viewed from a perspective on a positive multiple of $\mathbf{v}_0$ axis looking toward the origin) about $span\{\mathbf{v}_0\}$ by 120° where $\mathbf{v}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. *Hint: Do the matrix multiplication on Maple. Finish by simplifying your answer on Maple* (Maple command: simplify($U.R.U^{\%T}$)).

(g) Explain geometrically what is happening that makes this result from (f) so simple.

7.7 Symmetric Matrices and Orthogonal Diagonalizability

Recall that if A is an $n \times n$ matrix, then A defines a linear operator, $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, defined by $T(\mathbf{x}) = A\mathbf{x}$ relative to the standard basis $\mathcal{S}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$. If $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is another basis for $\mathbb{R}^n$, then $P = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ is the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}_n$, and the matrix $M = P^{-1}AP$ is said to be similar to A and describes the same linear operator, T, as A but relative to the new basis $\mathcal{B}$. That is, $A\mathbf{x} = \mathbf{y}$, if and only if $M[\mathbf{x}]_{\mathcal{B}} = [\mathbf{y}]_{\mathcal{B}}$.

Recall too that $\mathcal{B}$ is a basis of $\mathbb{R}^n$ of eigenvectors of A if and only if M is a diagonal matrix. When such a basis $\mathcal{B}$ exists, we say A is diagonalizable.

In this section we are especially interested in when $\mathcal{B}$ is an orthonormal basis of eigenvectors of A. When this occurs, we say A is orthogonally diagonalizable. Though it may seem that this situation is quite esoteric, it turns out that this has important applications throughout the sciences and mathematics.

In a very striking result, we show that a matrix A is orthogonally diagonalizable if and only if A is symmetric ($A^T = A$). We develop this result in the remainder of this section.

Lemma 7.30. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors associated with distinct eigenvalues of a symmetric matrix A, then $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$.

Proof Let λ_1 and λ_2 be distinct eigenvalues of A associated with eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively. Since the eigenvalues are distinct, $\lambda_1 \neq \lambda_2$. Note that

$$\begin{aligned} \mathbf{v}_1^T A \mathbf{v}_2 &= \mathbf{v}_1^T (A\mathbf{v}_2) \\ &= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) \\ &= \lambda_2 \left(\mathbf{v}_1^T \mathbf{v}_2\right) \\ &= \lambda_2 (\mathbf{v}_1 \cdot \mathbf{v}_2) \end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbf{v}_1^T A \mathbf{v}_2 &= \left(\mathbf{v}_1^T A\right)\mathbf{v}_2 \\
&= (A\mathbf{v}_1)^T \mathbf{v}_2 \\
&= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 \\
&= \lambda_1 \left(\mathbf{v}_1^T \mathbf{v}_2\right) \\
&= \lambda_1 (\mathbf{v}_1 \cdot \mathbf{v}_2)
\end{aligned}$$

Thus, $\lambda_2 (\mathbf{v}_1 \cdot \mathbf{v}_2) = \lambda_1 (\mathbf{v}_1 \cdot \mathbf{v}_2)$. So

$$\begin{aligned}
0 &= \lambda_2 (\mathbf{v}_1 \cdot \mathbf{v}_2) - \lambda_1 (\mathbf{v}_1 \cdot \mathbf{v}_2) \\
&= (\lambda_2 - \lambda_1)(\mathbf{v}_1 \cdot \mathbf{v}_2)
\end{aligned}$$

Solving we get $\lambda_1 = \lambda_2$ or $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Since $\lambda_1 \neq \lambda_2$, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ hence $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$. ■

Lemmas 7.31 and 7.32 are presented without proof. Their proofs require the development of complex inner product spaces.

Lemma 7.31. The eigenvalues of a real symmetric matrix are all real numbers.

Lemma 7.32. The geometric multiplicity of each eigenvalue of a real symmetric matrix equals its algebraic multiplicity.

Theorem 7.33. A real $n \times n$ matrix is orthogonally diagonalizable if and only if it is symmetric.

Proof Suppose A is orthogonally diagonalizable. Then, there exists an orthogonal matrix U such that $U^T AU = D$ where D is a diagonal matrix. Solving for A, we get $A = UDU^T$ since U is orthogonal. Thus

$$A^T = \left(UDU^T\right)^T = UD^T U^T = UDU^T = A$$

since diagonal matrices are symmetric. Therefore A is symmetric.

Suppose A is symmetric. Since all the eigenvalues of A are real (Lemma 7.31), counting algebraic multiplicities, A has n real eigenvalues. Since the geometric multiplicities equal the algebraic multiplicities, A has n linearly independent eigenvectors (Lemma 7.32). Thus A is diagonalizable (Theorem 6.8). Each eigenspace is a finite-dimensional subspace of Euclidean n-space, so by applying the Gram-Schmidt process if necessary, each eigenspace has an orthonormal basis. Since eigenvectors associated with distinct eigenvalues are orthogonal (Lemma 7.30), the union of the orthonormal bases for each eigenspace forms an orthonormal basis for all of Euclidean n-space. Therefore, A is orthogonally diagonalizable. ■

Example 7.19

Which of the following matrices are orthogonally diagonalizable? For those that are orthogonally diagonalizable, find a diagonal matrix D and an orthogonal matrix U such that $D = U^T AU$.

(a)

$$\begin{bmatrix} -1 & 5 & 2 \\ 5 & -1 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 11 & 2 & 10 \\ 2 & 14 & -5 \\ 10 & -5 & -10 \end{bmatrix}$$

(c)

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

Solution

(a) Since this matrix is symmetric, it is orthogonally diagonalizable. We start by calculating its characteristic polynomial to find its eigenvalues.

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -1-\lambda & 5 & 2 \\ 5 & -1-\lambda & 2 \\ 2 & 2 & 2-\lambda \end{vmatrix} \\ &= (-1-\lambda)(-1-\lambda)(2-\lambda) + 20 + 20 - 4(-1-\lambda) - 25(2-\lambda) - 4(-1-\lambda) \\ &= -\lambda^3 + 36\lambda \\ &= -\lambda(\lambda^2 - 36) \\ &= \lambda(\lambda+6)(\lambda-6) \end{aligned}$$

There are three eigenvalues $\lambda = 0, -6, 6$. Because A is symmetric we know it is orthogonally diagonalizable. A diagonal matrix D we seek has 0, -6, and 6 as its diagonal entries in no particular order. To find an orthogonal matrix U that diagonalizes A, we must find eigenvectors corresponding to each eigenvalue.

▶$\lambda_1 = 6$

Row reduction

$$\begin{bmatrix} -7 & 5 & 2 \\ 5 & -7 & 2 \\ 2 & 2 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & -2 \\ 5 & -7 & 2 \\ -7 & 5 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & -2 \\ 0 & -12 & 12 \\ 0 & 12 & -12 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

gives that

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

is an eigenvector associated with eigenvalue $\lambda_1 = 6$.

▶$\lambda_2 = -6$

Row reduction

$$\begin{bmatrix} 5 & 5 & 2 \\ 5 & 5 & 2 \\ 2 & 2 & 8 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 4 \\ 5 & 5 & 2 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 4 \\ 0 & 0 & -18 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

gives that

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

is an eigenvector associated with eigenvalue $\lambda_2 = -6$.
▶$\lambda_3 = 0$
Row reduction

$$\begin{bmatrix} -1 & 5 & 2 \\ 5 & -1 & 2 \\ 2 & 2 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -5 & -2 \\ 0 & 24 & 12 \\ 0 & 12 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -5 & -2 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -5 & -2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$$

gives that

$$\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

is an eigenvector associated with eigenvalue $\lambda_3 = 0$.

Notice that the three eigenvectors are orthogonal. This must occur because they correspond to distinct eigenvalues of a symmtric matrix (Lemma 7.30). To find U we need only normalize. The normalized eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{u}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

so

$$U = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{bmatrix}.$$

With the eigenvectors in this order in U we obtain

$$U^T AU = \begin{bmatrix} 6 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(b) Again, with this symmetric matrix we find its eigenvalues

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 11-\lambda & 2 & 10 \\ 2 & 14-\lambda & -5 \\ 10 & -5 & -10-\lambda \end{vmatrix} \\ &= (11-\lambda)(14-\lambda)(-10-\lambda) - 100 - 100 - 100(14-\lambda) - 4(-10-\lambda) - 25(11-\lambda) \\ &= -1540 + 250\lambda - 10\lambda^2 - 154\lambda + 25\lambda^2 - \lambda^3 - 1835 + 129\lambda \\ &= (-\lambda^3 + 15\lambda^2) + (225\lambda - 3375) \\ &= -\lambda^2(\lambda - 15) + 225(\lambda - 15) \\ &= -(\lambda - 15)(\lambda^2 - 225) \\ &= -(\lambda - 15)^2(\lambda + 15) \end{aligned}$$

The eigenvalues are $\lambda = 15, -15$. The multiplicity of $\lambda = 15$ is two, so it has a two-dimensional eigenspace (Lemma 7.32). We now find a basis of eigenvectors.
▶$\lambda_1 = -15$
Row reduction

$$\begin{bmatrix} 26 & 2 & 10 \\ 2 & 29 & -5 \\ 10 & -5 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & -1 & 1 \\ 2 & 29 & -5 \\ 26 & 2 & 10 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & -1 & 1 \\ 0 & 30 & -6 \\ 0 & 15 & -3 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 2 & -1 & 1 \\ 0 & 5 & -1 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & -1 & 1 \\ 0 & 1 & -1/5 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 0 & 4/5 \\ 0 & 1 & -1/5 \\ 0 & 0 & 0 \end{bmatrix}$$

gives that

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$$

is an eigenvector associated with eigenvalue $\lambda_1 = -15$.
▶$\lambda_2 = 15$
Row reduction

$$\begin{bmatrix} -4 & 2 & 10 \\ 2 & -1 & -5 \\ 10 & -5 & -25 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & -1 & -5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

yields two free variables. Let $x = s$ and $z = t$. Then,

$$\begin{array}{rcrcr} x & = & s & & \\ y & = & 2s & - & 5t \\ z & = & & & t \end{array}$$

so that

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \text{ and } \mathbf{v}_3 = \begin{bmatrix} 0 \\ -5 \\ 1 \end{bmatrix}$$

are linearly independent eigenvectors associated with eigenvalue $\lambda_2 = 15$.

Notice that both eigenvectors associated with $\lambda_2 = 15$ are orthogonal to the eigenvector associated with $\lambda_1 = -15$. Again, this must occur because they are associated with distinct eigenvalues of a symmetric matrix. This time, however, we cannot get by with simply normalizing the three vectors to form U. This is because the two linearly independent eigenvectors we chose for a basis of the eigenspace for $\lambda_2 = 15$ do not happen to be orthogonal. We use the Gram-Schmidt process to replace them with an orthogonal basis for the same eigenspace.
▶ Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ -5 \\ 1 \end{bmatrix}.$$

Gram-Schmidt gives

$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \text{ and } \mathbf{w}_2 = \mathbf{v}_2 - \frac{(\mathbf{v}_2 \cdot \mathbf{w}_1)}{(\mathbf{w}_1 \cdot \mathbf{w}_1)}\mathbf{w}_1 = \begin{bmatrix} 0 \\ -5 \\ 1 \end{bmatrix} - \frac{-10}{5}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}.$$

So we choose $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$ as an orthogonal basis for this eigenspace. This gives us an orthogonal basis for Euclidean 3-space of eigenvectors of A:

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

We normalize each and make the normalized vectors the columns of U.

$$U = \begin{bmatrix} -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{6}} \\ \frac{5}{\sqrt{30}} & 0 & \frac{1}{\sqrt{6}} \end{bmatrix}$$

With this choice of U we get

$$U^T AU = \begin{bmatrix} -15 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 15 \end{bmatrix}.$$

(c) This matrix is not symmetric, so it is not orthogonally diagonalizable.

Problem Set 7.7

1. For each of the following symmetric matrices A, find a diagonal matrix D and an orthogonal matrix U such that $D = U^T AU$.

(a) $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ **(b)** $\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$ **(c)** $\begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix}$ **(d)** $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 2 & -2 \\ 2 & -1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$ **(f)** $\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$ **(g)** $\begin{bmatrix} 2 & 2 & -10 \\ 2 & 11 & 8 \\ -10 & 8 & 5 \end{bmatrix}$

(h) $\begin{bmatrix} 3 & 0 & -4 \\ 0 & 5 & 0 \\ -4 & 0 & -3 \end{bmatrix}$ **(i)** $\begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$

2. Let A be a symmetric $n \times n$ matrix, and let B be any $m \times n$ matrix. Prove that the following matrices are symmetric: $B^T B$, BB^T, BAB^T.

3. Let $\mathbf{u}$ be a unit vector in Euclidean n-space and let $P = \mathbf{u}\mathbf{u}^T$.

- **(a)** Show that $P\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{u}$. *(Hint: See Exercise 4(a) in section 7.3.)*
- **(b)** Show that P is a symmertic matrix.
- **(c)** Show that $P^2 = P$.
- **(d)** Show that $\mathbf{u}$ is an eigenvector of P by finding its corresponding eigenvalue.
- **(e)** Find the other eigenvalue of P, describe its eigenspace, and give the dimension of that eigenspace.

4. Let A be an $n \times n$ symmetric matrix. Prove that if $\mathbf{x}$ and $\mathbf{y}$ are in Euclidean n-space, then $(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$.

5. True or false. Let A be an $n \times n$ matrix with real entries.

(a) If A is an orthogonally diagonalizable matrix, then A must be symmetric.

(b) If A is a symmetric matrix, then A must be orthogonally diagonalizable.

(c) If A is symmetric, then all of the eigenvalues of A are real numbers.

(d) If all the eigenvalues of A are real, then A is symmetric.

(e) If A is symmetric, then all the eigenvalues of A are distinct.

(f) If all the eigenvalues of A are distinct and real, then A is symmetric.

(g) If $A^T = A$, $A\mathbf{u} = 2\mathbf{u}$, and $A\mathbf{v} = 3\mathbf{v}$, then $\mathbf{u} \cdot \mathbf{v} = 0$.

(h) If A is symmetric, then A cannot have 0 as an eigenvalue.

(i) If A is orthogonal, then A cannot have 0 as an eigenvalue.

(j) If A is symmetric, then the sum of the dimensions of all the eigenspaces of A must equal n.

8 Applications

8.1 Quadratic Forms

An algebraic expression of the form

$$ax^2 + by^2 + cxy$$

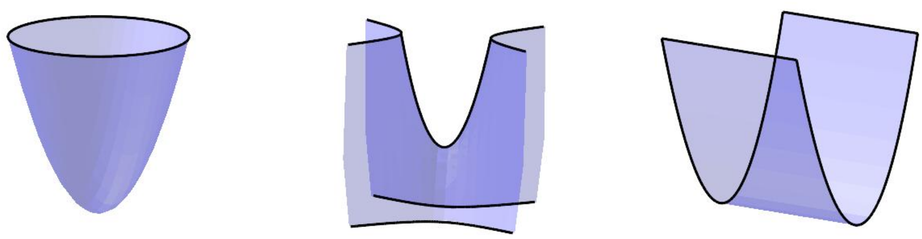
where x and y are variables and a, b, and c are constants is called a quadratic form in the variables x and y. Similarly, an expression of the form

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz$$

is a quadratic form in the variables x, y, and z. In general, a **quadratic form** in the variables $x_1, \cdots, x_n$ is an algebraic expression that can be written as a sum of terms of the form ax_ix_j where a is a constant and $i \le j$ (when $i = j$ the term is written ax_i^2).

Quadratic forms have many applications in geometry (conic sections), engineering (design criteria, optimization, and signal processing), statistics, physics, and economics.

Quadratic forms in two variables can be used to define functions $Q(x, y) = ax^2 + by^2 + cxy$ with graphs $z = Q(x, y)$ that can be viewed in three space. We shall see that the graphs fall into three basic categories depending on the constants a, b, and c (see Figure 8.1).



(a) Elliptic paraboloid. **(b)** Hyperbolic paraboloid. **(c)** Parabolic cylinder.

Figure 8.1

For various constants a, b, and c, the graphs may be turned upside down and for $a = b = c = 0$ the graph of $z = Q(x, y) = 0$ is the trivial (and uninteresting) xy–plane in $\mathbb{R}^3$. We don't have enough geometric dimensions to view graphs of quadratic forms in more than two variables.

About the only place in this course where we have encountered variables squared is with norms ($\|\mathbf{u}\|^2 = x_1^2 + \cdots + x_n^2$). But we shall see that matrices can be used to describe quadratic forms and that an analysis involving the orthogonal diagonalization of symmetric matrices can be used to understand quadratic forms and their applications. The following example introduces us to this inquiry.

Example 8.1

Let $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $B = \begin{bmatrix} 6 & 3 \\ 1 & 5 \end{bmatrix}$. The expression $\mathbf{x}^T B\mathbf{x}$ can be simplified as follows.

$$\begin{aligned} \mathbf{x}^T B\mathbf{x} &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 6 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 6x+3y \\ x+5y \end{bmatrix} \\ &= x(6x+3y) + y(x+5y) \\ &= 6x^2 + 3xy + xy + 5y^2 \\ &= 6x^2 + 5y^2 + 4xy \end{aligned}$$

Notice that the expression $\mathbf{x}^T B\mathbf{x}$ in Example 8.1 simplifies to the quadratic form $6x^2 + 5y^2 + 4xy$. We find it helpful to reverse this process. That is, to start with a quadratic form and end with an expression of the form $\mathbf{x}^T B\mathbf{x}$ where B is a square matrix.

By following the calculations in Example 8.1, it is easy to see that the coefficients of x^2 and y^2 (i.e. 6 and 5) come from the diagonal entries of B, and the coefficient of xy (i.e. 4) comes from the off-diagonal entries (i.e. $3+1=4$). With this in mind, it appears that there are many matrices besides $B = \begin{bmatrix} 6 & 3 \\ 1 & 5 \end{bmatrix}$ that result in the same quadratic form $6x^2 + 5y^2 + 4xy$. The matrices

$$\begin{bmatrix} 6 & 0 \\ 4 & 5 \end{bmatrix}, \begin{bmatrix} 6 & 1 \\ 3 & 5 \end{bmatrix}, \begin{bmatrix} 6 & 2 \\ 2 & 5 \end{bmatrix}, \text{ and } \begin{bmatrix} 6 & 7 \\ -3 & 5 \end{bmatrix}$$

would all work. Notice too that of all the matrices that work, only $\begin{bmatrix} 6 & 2 \\ 2 & 5 \end{bmatrix}$ is symmetric. Because of the amazing property of orthogonal diagonalizability we choose the symmetric matrix.

To express the quadratic form $6x^2 + 5y^2 + 4xy$ in matrix form $\mathbf{x}^T A\mathbf{x}$ with A symmetric, let $A = \begin{bmatrix} 6 & 2 \\ 2 & 5 \end{bmatrix}$.

Example 8.2

Express the quadratic form $x^2 - 2y^2 + 3z^2 + 4xy + 5xz - 6yz$ in the form $\mathbf{x}^T A\mathbf{x}$ where A is symmetric.

Solution Let

$$A = \begin{bmatrix} 1 & 2 & \frac{5}{2} \\ 2 & -2 & -3 \\ \frac{5}{2} & -3 & 3 \end{bmatrix}.$$

Note that the coefficients of x^2, y^2, and z^2 (1, -2, and 3) are the diagonal entries of A. The coefficient of xy is 4. Half of 4 is 2 and both the $(1,2)$ and the $(2,1)$ entries of A are 2. Similarly, both the $(1,3)$ and the $(3,1)$ entries of A are $\frac{5}{2}$ because the coefficient of xz is 5, and both the $(2,3)$ and $(3,2)$ entries of A are -3 because the coefficient of yz is -6. You can check to see that $\mathbf{x}^T A\mathbf{x}$ equals the given quadratic form.

The terms of a quadratic form that involve two different variables are called **cross-product terms**. To write a quadratic form $\mathbf{x}^T A\mathbf{x}$, where A is symmetric, each diagonal entry a_{ii} equals the coefficient of x_i^2 in the quadratic form, and each off-diagonal a_{ij} and a_{ji} $(i < j)$ equals half of the coefficient of $x_i x_j$ in the quadratic form.

Because of the power that orthogonal diagonalizability brings to bear on our quadratic forms, we always choose A symmetric when writing them as $\mathbf{x}^T A\mathbf{x}$.

Quadratic forms in which all cross-product terms are 0 are particularly easy to understand. For one thing, when written as $\mathbf{x}^T A\mathbf{x}$, the cross-product terms are all 0 if and only if A is, in fact, a diagonal matrix.

In the two-variable case it is easy to see that when the cross-product term is 0, the graph of $z = Q(x, y)$ where $Q(x, y) = ax^2 + by^2$ is

(a) an elliptical paraboloid if a and b have the same signs,

(b) a hyperbolic paraboloid if a and b have opposite signs, and

(c) a parabolic cylinder if either a or b (but not both) is 0.

The next example illustrates how orthogonal diagonalization can be used to determine the graph of a quadratic form.

Example 8.3

Determine the shape of the graph of $Q(x, y) = 3x^2 + 4xy$.

Solution Rewriting,

$$Q(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}.$$

Since A is symmetric (by design), A is orthogonally diagonalizable. To diagonalize A, we find the eigenvalues and eigenvectors of A. The characteristic polynomial of A is

$$\begin{aligned} p(\lambda) &= \det(A - \lambda I) \\ &= \begin{vmatrix} 3-\lambda & 2 \\ 2 & -\lambda \end{vmatrix} \\ &= (3-\lambda)(-\lambda) - 4 \\ &= \lambda^2 - 3\lambda - 4 \\ &= (\lambda - 4)(\lambda + 1) \end{aligned}$$

So, the eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = -1$.

We now find eigenvectors. For $\lambda_1 = 4$, we solve $(A - 4I)\mathbf{x} = \mathbf{0}$. By row reduction,

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

Letting $y = t$ gives $x = 2t$ and the eigenspace

$$E_4 = span\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}.$$

For $\lambda_2 = -1$, since A is symmetric, the eigenvectors for $\lambda_2 = -1$ must be orthogonal to the eigenvector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ in the other eigenspace. Thus

$$E_{-1} = span\left\{\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right\}.$$

The matrix

$$P = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

would diagonalize A, but for a matrix that *orthogonally* diagonalizes A, we must normalize the eigenvectors. So let

$$U = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$

Then

$$U^T AU = D = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}.$$

To use this information, we perform a change of basis. Let

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{u}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

form our new basis and let $\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix}$ be the new coordinates of $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ relative to the new ordered orthonomal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$. So $\mathbf{x} = U\mathbf{x}'$ and $\mathbf{x}' = U^T\mathbf{x}$. This amounts to a rotation of the original axes by $\theta = \tan^{-1}(1/2) \approx 26.6°$ (see Figure 8.2).

If $z = Q(x, y) = 3x^2 + 4xy$, to write z in terms of the new coordinate system we have

$$\begin{aligned}
z &= 3x^2 + 4xy \\
&= \begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\
&= \mathbf{x}^T A\mathbf{x} \\
&= (U\mathbf{x}')^T A(U\mathbf{x}') \quad \text{since } \mathbf{x} = U\mathbf{x}' \\
&= \left(\mathbf{x}'^T U^T\right) A(U\mathbf{x}') \\
&= \mathbf{x}'^T (U^T AU)\mathbf{x}' \\
&= \mathbf{x}'^T D\mathbf{x}' \quad \text{since } U^T AU = D \\
&= \begin{bmatrix} x' & y' \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} \\
&= 4(x')^2 - (y')^2
\end{aligned}$$

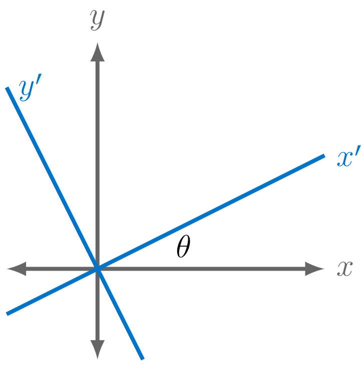


Figure 8.2 *A new coordinate system in which to view the quadratic form* $Q(x, y) = 3x^2 + 4xy$.

Since this is a quadratic form with a cross-product term of 0, we see that the graph is a hyperbolic paraboloid because the coefficients of $(x')^2$ and $(y')^2$ have opposite signs.

Example 8.3 is rather long and involved. It is presented that way in order to review concepts developed earlier and to present a thorough step-by-step explanation as to how each conclusion is obtained. Once that example is thoroughly studied and understood, it should be noted that the conclusions of this example can be made with considerably less work.

Example 8.3 (stripped down version)

Determine the shape of the graph of $Q(x, y) = 3x^2 + 4xy$.

Solution Since

$$Q(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

let

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}.$$

Being symmetric, A must have real eigenvalues. Therefore, under a new coordinate system that orthogonally diagonalizes A, $z = \mathbf{x}^T A\mathbf{x}$ can be rewritten $z = \lambda_1 x'^2 + \lambda_2 y'^2$ where λ_1 and λ_2 are the eigenvalues of A. Since $\det A = -4$, λ_1 and λ_2 have opposite sign, and the graph of $z = \mathbf{x}^T A\mathbf{x}$ is a hyperbolic paraboloid.

To state the result of Example 8.3 even more succintly: The graph of $z = \mathbf{x}^T A\mathbf{x}$ is a hyperbolic paraboloid because $\det A < 0$. Applying this reasoning to cases where $\det A > 0$ and $\det A = 0$ as well we obtain Theorem 8.4.

Theorem 8.1. Let A be a symmetric nonzero 2×2 matrix. The graph of $z = \mathbf{x}^T A\mathbf{x}$ is

(a) an elliptic paraboloid if $\det A > 0$,

(b) a hyperbolic paraboloid if $\det A < 0$, and

(c) a parabolic cylinder if $\det A = 0$.

Here is a subtle point that can cause some confusion if not understood: If $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for $\mathbb{R}^2$ that diagonalizes the symmetric matrix A, then there are eight different choices for the diagonalizing matrix U. They are

$$\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_1 & -\mathbf{u}_2 \end{bmatrix}, \begin{bmatrix} -\mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}, \begin{bmatrix} -\mathbf{u}_1 & -\mathbf{u}_2 \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{u}_2 & \mathbf{u}_1 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_2 & -\mathbf{u}_1 \end{bmatrix}, \begin{bmatrix} -\mathbf{u}_2 & \mathbf{u}_1 \end{bmatrix}, \text{ and } \begin{bmatrix} -\mathbf{u}_2 & -\mathbf{u}_1 \end{bmatrix}.$$

These generate eight different coordinate systems that result from the same coordinate axes. The variations result from the eight different ways of choosing the x' and y' axes and which direction is positive in each case. Figure 8.3 illustrates. The x' and y' labels indicate the positive directions for the x' and y' axes respectively.

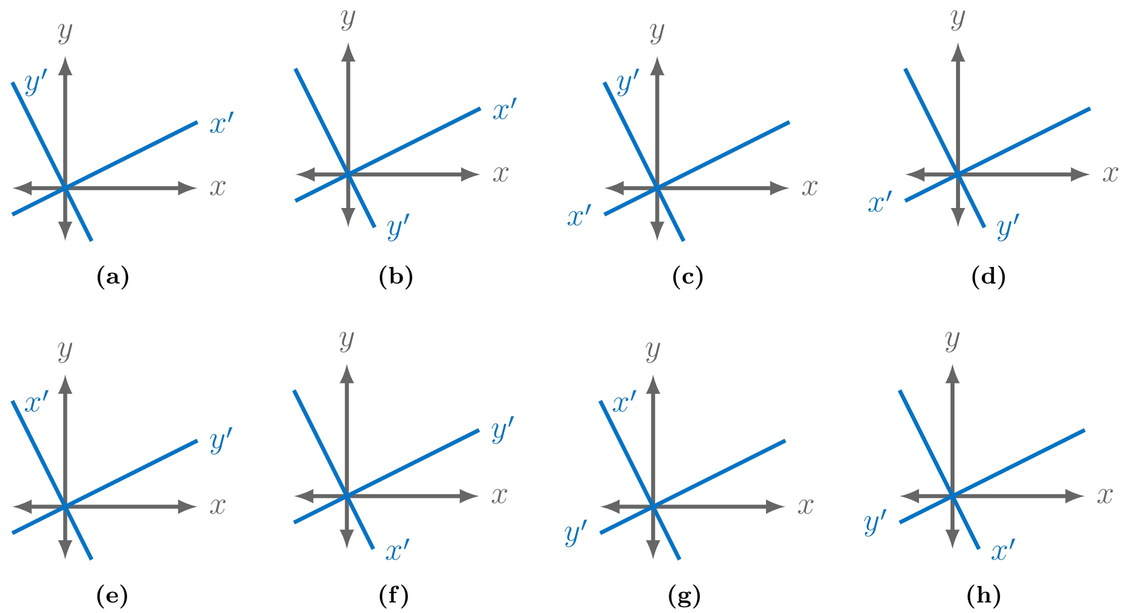


Figure 8.3

Four of the systems illustrated in Figure 8.3 represent rotations of the original coordinate axes, and four represent reflections. If $\det U = 1$ you have a rotation and if $\det U = -1$ you have a reflection.

Problem Set 8.1

1. Find the symmetric matrix A such that $Q(x,y) = \mathbf{x}^T A\mathbf{x}$.

   (a) $Q(x,y) = 5x^2 + 5y^2 + 2xy$

   (b) $Q(x,y) = 9x^2 + y^2 - 6xy$

   (c) $Q(x,y) = 2x^2 - 7y^2 + 12xy$

   (d) $Q(x,y) = x^2 + 2xy$

2. For (a) through (d) of exercise 1, describe the basic shape of the graph of $z = Q(x,y)$. (*Hint:* The basic shape in example 8.3 is hyperbolic paraboloid.)

3. For part (a) of exercise 1, write out all eight possible orthogonal matrices that orthogonally diagonalize the symmetric matrix A.

4. For parts (b) through (d) of exercise 1, find the orthogonal matrix that orthogonally diagonalizes A and is also a rotation matrix R_θ for an acute angle θ.

5. For parts (a) through (d) of exercise 2, rewrite z in terms of new coordinates x' and y' that eliminates the cross-product term. In each case there are two possible answers. Write both.

6. Rewrite w in terms of new coordinates x', y', and z' that eliminates all cross-product terms.

   (a) $w = 2x^2 + 11y^2 + 5z^2 + 4xy - 20xz + 16yz$

   (b) $w = 3x^2 + 5y^2 - 3z^2 - 8xz$

7. For both (a) and (b) from exercise 6, how many different correct answers are possible?

8. Suppose A is a 4×4 symmtric matrix and $y = \mathbf{x}^T A\mathbf{x}$ is a quadratic form in the variables x_1, x_2, x_3, and x_4. In how many different ways can y be rewritten in terms of a new coordinate system x_1', x_2', x_3', and x_4' if

   (a) A has four distince eigenvalues?

   (b) A has one eigenvalue with multiplicity 3 and one eigenvalue with multiplicity 1?

   (c) A has two distinct eigenvalues each with mulitplicity 2?

   (d) A has three distinct eigenvalues, one with multiplicity 2 and two of multiplicity 1?

9. Suppose A is an $n \times n$ symmetric matrix and $y = \mathbf{x}^T A\mathbf{x}$ is a quadratic form in the variables $x_1, \cdots, x_n$. In how many different ways can y be rewritten in terms of a new coordinate system $x_1', \cdots, x_n'$ if A has k distinct eigenvalues $\lambda_1, \cdots, \lambda_k$ with multiplicities $m_1, \cdots, m_k$ respectively? $(m_1 + \cdots + m_k = n)$

8.2 Constrained Optimization of Quadratic Forms

Let $Q(\mathbf{x})$ be a quadratic form in the variables $x_1, \cdots, x_n$ with $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. We wish to optimize (maximize and minimize) $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\| = c > 0$.

Since $Q(\mathbf{x})$ is a quadratic form, there is a symmetric matrix A such that $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. And since A is symmetric, it can be orthogonally diagonalized by an orthogonal matrix U. The columns of U form an orthnormal basis for $\mathbb{R}^n$ of eigenvectors of A in some order. Without loss of generality, suppose that $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ where each $\mathbf{u}_i$ is an eigenvector of A associated with the real eigenvalue λ_i such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

By convention, the eigenvalues of symmetric matrices, being all real numbers, are listed in decreasing order. Eigenvalues with multiplicity great than one are repeated in a listing in accordance to their multiplicity. We adopt this convention throughout the rest of this text.

Let $\mathbf{x} \in \mathbb{R}^n$ be such that $\|\mathbf{x}\| = c$ and let $\mathbf{x}' = U^T\mathbf{x} = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}$. Then $\mathbf{x} = U\mathbf{x}'$ and $\mathbf{x}'$ is the coordinate vector of $\mathbf{x}$ relative to the ordered orthonomal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$. Since $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonomal basis, $\|\mathbf{x}\| = \|\mathbf{x}'\| = \sqrt{x_1^2 + \cdots + x_n^2}$.

Diagonalizing A we get

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} \\ &= \left(U\mathbf{x}'\right)^T A\left(U\mathbf{x}'\right) \\ &= (\mathbf{x}')^T \left(U^T AU\right)\mathbf{x}' \\ &= (\mathbf{x}')^T D\mathbf{x}' \text{ where } D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \\ &= \lambda_1 {x'_1}^2 + \cdots + \lambda_n {x'_n}^2. \end{aligned}$$

Note that since λ_1 is the largest eigenvalue of A

$$\begin{aligned} Q(\mathbf{x}) &= \lambda_1 x_1'^2 + \cdots + \lambda_n x_n'^2 \\ &\leq \lambda_1 x_1'^2 + \cdots + \lambda_1 x_n'^2 \\ &= \lambda_1 (x_1'^2 + \cdots + x_n'^2) \\ &= \lambda_1 \|\mathbf{x}'\|^2 \\ &= \lambda_1 \|\mathbf{x}\|^2 \\ &= \lambda_1 c^2. \end{aligned}$$

Similarly, since λ_n is the smallest eigenvalue of A, $Q(\mathbf{x}) \geq \lambda_n c^2$ so for all $\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = c$ we have

$$\lambda_n c^2 \leq Q(\mathbf{x}) \leq \lambda_1 c^2.$$

Finally, note that if $\mathbf{x}$ is an eigenvector of A associated with the eigenvalue λ_i such that $\|\mathbf{x}\| = c$, then

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} \\ &= \mathbf{x}^T (A\mathbf{x}) \\ &= \mathbf{x}^T(\lambda_i \mathbf{x}) \\ &= \lambda_i(\mathbf{x}\cdot\mathbf{x}) \\ &= \lambda_i\|\mathbf{x}\|^2 \\ &= \lambda_i c^2. \end{aligned}$$

We put this all together in Theorem 8.2.

Theorem 8.2. Let $Q(\mathbf{x})$ be a quadratic form and A a symmetric matrix such that $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of A (repeats to match multiplicity) such that $\lambda_1 \geq \cdots \geq \lambda_n$. For all $\mathbf{x}$ such that $\|\mathbf{x}\| = c$,

$$\lambda_n c^2 \leq Q(\mathbf{x}) \leq \lambda_1 c^2.$$

Further, these upper and lower bounds for $Q(\mathbf{x})$ are constrained maximums and minimums for $Q(\mathbf{x})$. The function $Q(\mathbf{x})$ takes on these constrained maximum and minimum values when $\|\mathbf{x}\| = c$ and $\mathbf{x}$ is an eigenvector of A associated with the largest and smallest eigenvalues of A respectively.

Example 8.4

Maximize and minimize $Q(x,y) = 5x^2 + 5y^2 + 4xy$ subject to the constraint $x^2 + y^2 = 4$.

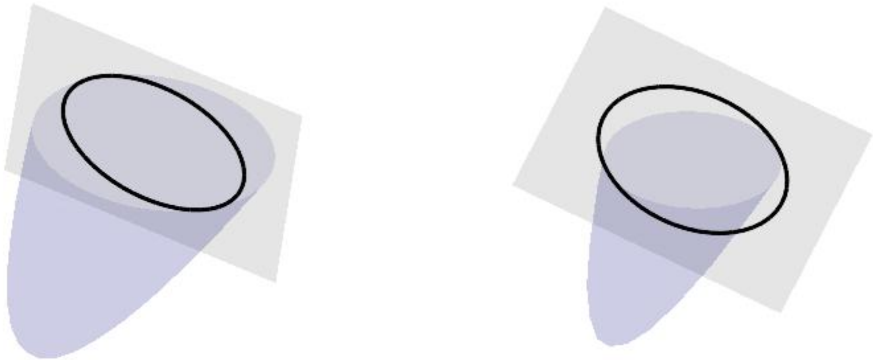
Solution Since

$$5x^2 + 5y^2 + 4xy = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

we let

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}.$$

It is easy to determine that the eigenvalues of A are $\lambda_1 = 7$ and $\lambda_2 = 3$ with associated eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ respectively. The constraint $x^2 + y^2 = 4$ is equivalent to saying $\|\mathbf{x}\| = 2$, so Q has a maximum of $7\cdot 2^2 = 28$ subject to this constraint. Q attains this maximum at the mulitples of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ that have a norm of 2. There are two vectors that satisfy these properties. They are $\begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$ and $\begin{bmatrix} -\sqrt{2} \\ -\sqrt{2} \end{bmatrix}$. Similarly, Q has a minimum of $3\cdot 2^2 = 12$ subject to this constraint when $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix}$ or $\begin{bmatrix} \sqrt{2} \\ -\sqrt{2} \end{bmatrix}$. See Figures 8.4a and 8.4b.



(a) The maximum of $Q(x,y)$ occurs at $z = 28$. The circle defined by intersecting the plane $z = 28$ and cylinder $x^2 + y^2 = 4$ just touches the paraboloid from the inside.

(b) The minimum of $Q(x,y)$ occurs at $z = 12$. The circle defined by intersecting the plane $z = 12$ and cylinder $x^2 + y^2 = 4$ just touches the paraboloid from the outside.

Figure 8.4 *The maximum and minimum values of* $Q(x,y) = 5x^2 + 5y^2 + 4xy$ *subject to* $x^2 + y^2 = 4$.

Example 8.5

Maximize and minimize $Q(x,y,z) = x^2 + 4y^2 + 4z^2 - 4xy - 4xz + 8yz$ subject to the constraint $x^2 + y^2 + z^2 = 1$.

Solution Rewriting,

$$Q(x,y,z) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

So, we let

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

The eigenspaces of A are

$$E_9 = span\left\{ \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix} \right\} \text{ and } E_0 = span\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Thus, Q takes on its constrained maximum of 9 at

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \pm\frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix}.$$

Those two points are the intersection of the eigenspace E_9 (a line) and the constraint $x^2 + y^2 + z^2 = 1$ (a sphere) in the domain of Q. Similarly, Q attains its constrained minimum of 0 where the eigenspace E_0 (a plane) intersects the constraint (a sphere) in the domain of Q. This gives us a circle of radius 1 centered at the origin and lying on the plane $x - 2y - 2z = 0$. To describe these vectors explicitly, note that any vector in the eigenspace E_0 can be written as $\begin{bmatrix} 2s+2t \\ s \\ t \end{bmatrix}$ for some s and t. To get a vector on the circle we simply normalize (s and t not both 0):

$$\frac{1}{\sqrt{(2s+2t)^2+s^2+t^2}}\begin{bmatrix} 2s+2t \\ s \\ t \end{bmatrix} = \frac{1}{\sqrt{5s^2+5t^2+8st}}\begin{bmatrix} 2s+2t \\ s \\ t \end{bmatrix}.$$

Problem Set 8.2

Maximize and minimize each of the following functions subject to the given constraint and find all points in the domain of each function where the constrained optimums occur.

1. $Q(x, y) = 3x^2 + 3y^2 + 2xy$ subject to $x^2 + y^2 = 9$.
2. $Q(x, y) = 3x^2 - 5y^2 + 6xy$ subject to $x^2 + y^2 = 2$.
3. $Q(x, y) = 9x^2 + 16y^2 + 24xy$ subject to $x^2 + y^2 = 1$.
4. $Q(x, y) = 3x^2 - 2xy$ subject to $x^2 + y^2 = 1$.
5. $Q(x, y, z) = -x^2 - y^2 + 2z^2 + 10xy + 4xz + 4yz$ subject to $\|\mathbf{x}\| = 2$.
6. $Q(x, y, z) = z^2 + 2xy$ subject to $x^2 + y^2 + z^2 = 3$.

8.3 Conic Sections

Definition 8.1. A **quadratic equation** in the two variables x and y is an equation that can be put in the form

$$ax^2 + by^2 + cxy + dx + ey = f$$

where a, b, c, d, e, and f are constants and we assume a, b, and c are not all 0.

Notice that the left-hand side of the quadratic equation is the sum of a quadratic form $ax^2 + by^2 + cxy$ and a **linear form** $dx + ey$, and the right-hand side is a constant. Using section 8.1, we generally want to rewrite the quadratic form

$$ax^2 + by^2 + cxy = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

and the linear form

$$dx + ey = \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

So, the quadratic equation will generally be rewritten to look like

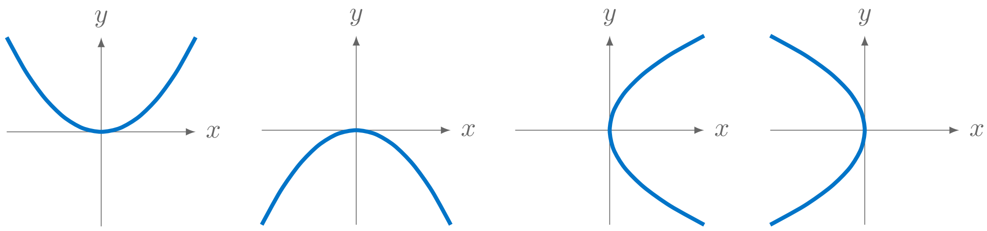
$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = f.$$

Our main goal is to find the graphs of these equations.

Quadratic equations in which the cross-product term cxy is zero are often studied in precalculus and/or calculus classes. Our main concern is when $c \neq 0$, but we mention what happens when $c = 0$ either as a review for you or as a brief introduction.

In this introduction, the roles of parameters like a, b, and c need not be the same as in the quadratic equation at the beginning of this section.

Equations of the form $y = ax^2$ $(a \neq 0)$ have graphs that look like those in Figures 8.5a and 8.5b depending on whether $a > 0$ or $a < 0$. The roles of x and y can be switched resulting in equations of the form $x = ay^2$ and graphs that look like those in Figures 8.5c and 8.5d. These graphs are called **parabolas** with vertex at the origin.



(a) $y = ax^2, a > 0$ **(b)** $y = ax^2, a < 0$ **(c)** $x = ay^2, a > 0$ **(d)** $x = ay^2, a < 0$

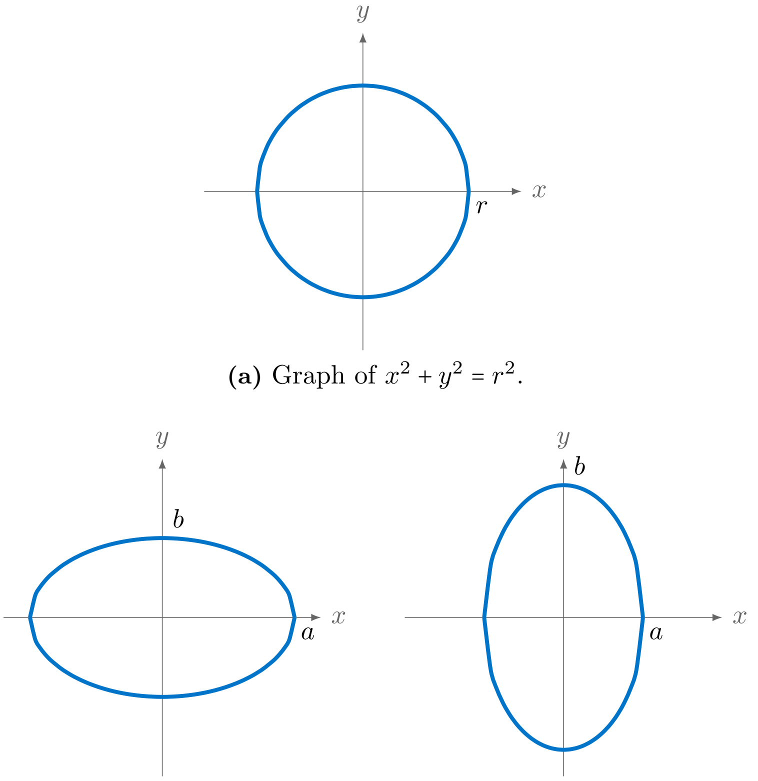
Figure 8.5

Equations of the form $x^2 + y^2 = r^2$ $(r > 0)$ have graphs that are **circles** centered at the origin with a radius of r. The graph of a circle is as in Figure 8.6a.

Equations of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ $(a, b > 0)$ have graphs that look like those found in Figures 8.6b and 8.6c depending on which of a and b is larger. These graphs are called **ellipses** centered at the origin. If $a = b$, the graph is a circle (a special case of an ellipse).

Equations of the form $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ or $\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$ have graphs that resemble those in Figures 8.7a and 8.7b. These are called **hyperbolas** centered at the origin.

All of these equations either are or can be placed into the form of the general quadratic equation with a cross-product term of 0 described at the beginning of this section.



(a) Graph of $x^2 + y^2 = r^2$.

(b) Graph of $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ $(a > b)$ (c) Graph of $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ $(a < b)$

Figure 8.6

By substituting $x - h$ for x and $y - k$ for y in each of these equations we shift the graphs so the vertex (in the case of a parabola) or its center is at the point (h, k) rather than the origin. Many more quadratic equations with a cross-product term of 0 can be placed into these substituted forms by completing squares.

Example 8.6

Describe the graph of $2x^2 - 4x - y = 3$.

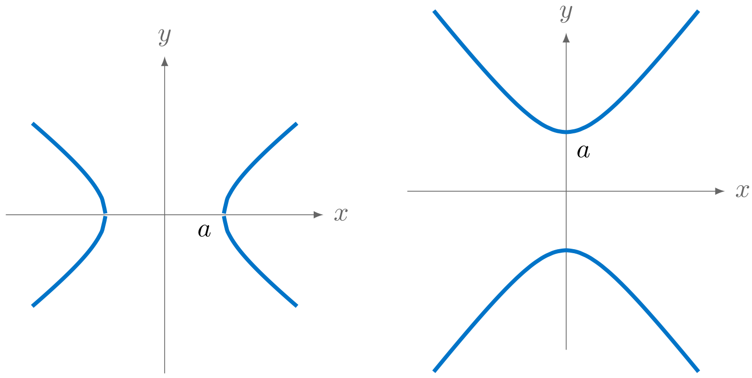
Solution

$$\begin{aligned} 2x^2 - 4x - y &= 3 \\ 2x^2 - 4x &= y + 3 \\ 2(x^2 - 2x + 1) &= y + 3 + 2 \\ 2(x-1)^2 &= y + 5 \end{aligned}$$

This is a parabola with a vertex at $(1, -5)$ that opens up (see Figure 8.8a).

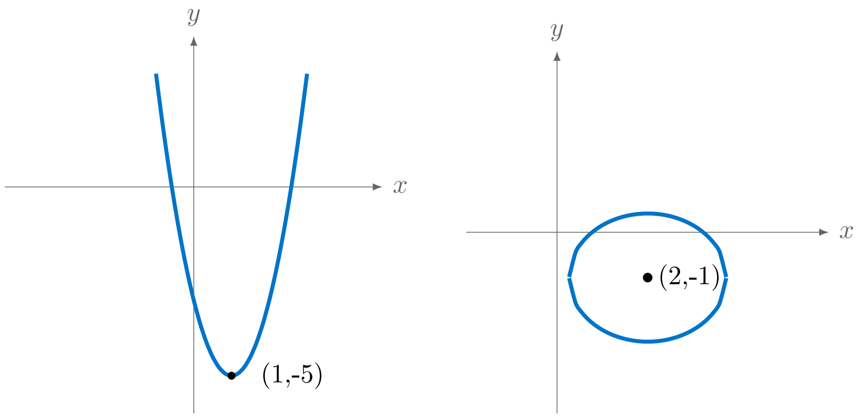
Example 8.7

Describe the graph of $2x^2 + 3y^2 - 8x + 6y = -5$.



(a) Graph of $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$. **(b)** Graph of $\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$.

Figure 8.7



(a) Graph of $2x^2 - 4x - y = 3$. **(b)** Graph of $2x^2 + 3y^2 - 8x + 6y = -5$.

Figure 8.8

Solution

$$\begin{aligned} 2x^2 + 3y^2 - 8x + 6y &= -5 \\ 2(x^2 - 4x + 4) + 3(y^2 + 2y + 1) &= -5 + 8 + 3 \\ 2(x-2)^2 + 3(y+1)^2 &= 6 \\ \frac{(x-2)^2}{3} + \frac{(y+1)^2}{2} &= 1 \end{aligned}$$

This can also be written as

$$\frac{(x-2)^2}{(\sqrt{3})^2} + \frac{(y+1)^2}{(\sqrt{2})^2} = 1.$$

This is an ellipse that opens wider horizontally ($\sqrt{3} > \sqrt{2}$) and is centered at $(2, -1)$ (see Figure 8.8b).

In order to exhaust all possibilities we must consider a few degenerate cases.

Equation	Graph
$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 0$	the single point (h,k)
$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = -1$	the solution set is empty; called an imaginary ellipse
$\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 0$	two lines that intersect at (h,k)
$\frac{(x-h)^2}{a^2} = 1$ or $\frac{(y-k)^2}{b^2} = 1$	two vertical or two horizontal lines
$(x-h)^2 = 0$ or $(y-k)^2 = 0$	a single vertical or a single horizontal line
$\frac{(x-h)^2}{a^2} = -1$ or $\frac{(y-k)^2}{b^2} = -1$	the solution set is empty

For brevity, we call the first two degenerate cases **degenerate ellipses**. The third we call a **degenerate hyperbola**, and the last three are called **degenerate parabolas**. Between the nondegenerate and the degenerate cases, this covers all possible quadratic equations with a cross-product term of 0.

In light of the last two sections the process we should follow when the cross-product term is not zero should be fairly clear. Write the quadratic equation in the form

$$\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = f$$

where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, A = \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} d \\ e \end{bmatrix}.$$

Do a change of basis (substitute $\mathbf{x} = U\mathbf{x}'$) where $\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix}$ and U is an orthogonal matrix that orthogonally diagonalizes A. This transforms the original quadratic equation into a new quadratic equation in the variables x' and y'

$$(\mathbf{x}')^T D\mathbf{x}' + (\mathbf{b}')^T\mathbf{x}' = f$$

where D is a diagonal matrix and $(\mathbf{b}')^T = \mathbf{b}^T U$.

As mentioned in section 8.1, you have eight choices for U. The form of the resulting quadratic equation can vary depending on your choice of U. Though any of the eight can be used, we tend to choose U to be a rotation matrix by an angle θ where $0 < \theta < \pi/2$. This occurs if the only negative entry of U is the top right entry.

Since D is a diagonal matrix, the resulting quadratic equation has a cross-product term of 0. You can then complete any squares necessary to determine the graph of the quadratic equation relative to the new variables x' and y'. Thus, the graphs of all quadratic equations are conic sections or their degenerates. Some are rotated, and some are not.

Example 8.8

Determine the shape of the graph of the quadratic equation $5x^2 + 2y^2 + 4xy = 6$.

Solution Rewriting this equation we get

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 6.$$

Let

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Then $\mathbf{x}^T A\mathbf{x} = 6$. The eigenvalues for A are $\lambda_1 = 6$ and $\lambda_2 = 1$ with associated eigenvectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ respectively. Let

$$U = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$

U is an orthogonal matrix that diagonalizes A. Let $\mathbf{x}' = U^T\mathbf{x}$ so $\mathbf{x} = U\mathbf{x}'$. Now substitute $\mathbf{x} = U\mathbf{x}'$ in the equation $\mathbf{x}^T A\mathbf{x} = 6$ and simplify:

$$\begin{aligned}
\left(U\mathbf{x}'\right)^T A(U\mathbf{x}') &= 6 \\
(\mathbf{x}')^T(U^T AU)\mathbf{x}' &= 6 \\
\begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} &= 6 \\
6(x')^2 + (y')^2 &= 6 \\
\frac{(x')^2}{1} + \frac{(y')^2}{(\sqrt{6})^2} &= 1
\end{aligned}$$

The graph is an ellipse centered at the origin (see Figure 8.9). The axis is rotated by $\theta = \tan^{-1}\left(\frac{1}{2}\right) \approx 26.6°$.

Example 8.9

Determine the shape of the graph of $x^2 + y^2 + 2xy + 3x + 2y = 0$.

Solution Rewriting this equation we get

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0.$$

Let

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

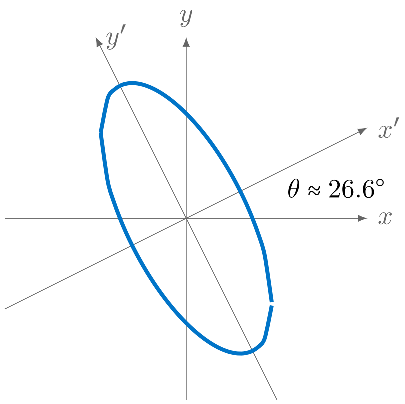


Figure 8.9 *Graph of* $5x^2 + 2y^2 + 4xy = 6$.

Then $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = 0$. The eigenvalues for A are $\lambda_1 = 2$ and $\lambda_2 = 0$ with associated eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ respectively. Let

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

U is an orthogonal matrix that diagonalizes A. Let $\mathbf{x}' = U^T\mathbf{x}$ so $\mathbf{x} = U\mathbf{x}'$. Now substitute $\mathbf{x} = U\mathbf{x}'$ into the equation $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = 0$ and simplify.

$$\begin{aligned}
(U\mathbf{x}')^T A(U\mathbf{x}') + \mathbf{b}^T(U\mathbf{x}') &= 0 \\
(\mathbf{x}')^T(U^T AU)(\mathbf{x}')^T + \mathbf{b}^T U\mathbf{x}' &= 0 \\
\begin{bmatrix} x' & y' \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} + \frac{1}{\sqrt{2}}\begin{bmatrix} 3 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} &= 0 \\
2(x')^2 + \frac{1}{\sqrt{2}}\begin{bmatrix} 5 & -1 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} &= 0 \\
2(x')^2 + \frac{5}{\sqrt{2}}x' - \frac{1}{\sqrt{2}}y' &= 0
\end{aligned}$$

Completing the square:

$$\begin{aligned}
2\left((x')^2 + \frac{5}{2\sqrt{2}}x' + \frac{25}{32}\right) &= \frac{1}{\sqrt{2}}y' + \frac{25}{16} \\
2\left(x' + \frac{5}{4\sqrt{2}}\right)^2 &= \frac{1}{\sqrt{2}}\left(y' + \frac{25\sqrt{2}}{16}\right) \\
2\sqrt{2}\left(x' + \frac{5\sqrt{2}}{8}\right)^2 &= y' + \frac{25\sqrt{2}}{16}
\end{aligned}$$

This is a parabola with its vertex at $\left(-\frac{5\sqrt{2}}{8}, -\frac{25\sqrt{2}}{16}\right)$ under the new coordinate system that is rotated by $\theta = \tan^{-1}(1/1) = 45°$ (see Figure 8.10). Since $\mathbf{x} = U\mathbf{x}'$, we can find the coordinates of this vertex under the original coordinate system:

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -5\sqrt{2}/8 \\ -25\sqrt{2}/16 \end{bmatrix} = \begin{bmatrix} 15/16 \\ -35/16 \end{bmatrix}.$$

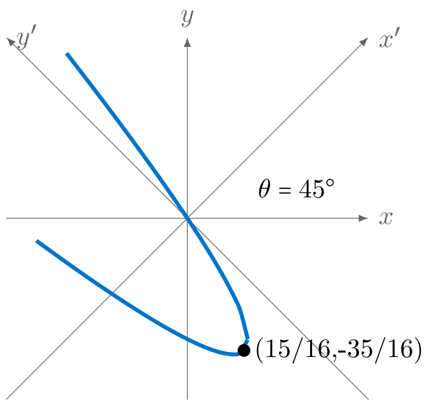


Figure 8.10 *Graph of* $x^2 + y^2 + 2xy + 3x + 2y = 0$.

Example 8.10

Determine the shape of the graph of $3x^2 - 3y^2 + 8xy - 30x + 10y = 0$.

Solution Rewriting this equation we get

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -30 & 10 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0.$$

Let

$$A = \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -30 \\ 10 \end{bmatrix}.$$

The eigenvalues for A are $\lambda_1 = 5$ and $\lambda_2 = -5$ with associated eigenvectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ respectively. Let

$$U = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$

Let $\mathbf{x}' = U^T\mathbf{x}$ so $\mathbf{x} = U\mathbf{x}'$. Now substitute $\mathbf{x} = U\mathbf{x}'$ into $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = 0$ and simplify:

$$\begin{aligned} (\mathbf{x}')^T(U^TAU)\mathbf{x}' + (\mathbf{b}^TU)\mathbf{x}' &= 0 \\ \begin{bmatrix} x' & y' \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & -5 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} + \frac{1}{\sqrt{5}}\begin{bmatrix} -50 & 50 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} &= 0 \\ 5(x')^2 - 5(y')^2 + 10\sqrt{5}(-x' + y') &= 0 \\ 5((x')^2 - 2\sqrt{5}x' + 5) - 5((y')^2 - 2\sqrt{5}y' + 5) &= 0 + 25 - 25 \\ 5(x' - \sqrt{5})^2 - 5(y' - \sqrt{5})^2 &= 0 \end{aligned}$$

The graph of this equation is a degenerate hyperbola - two intersecting lines:

$$\begin{aligned} (y' - \sqrt{5})^2 &= (x' - \sqrt{5})^2 \\ y' - \sqrt{5} = \pm(x' - \sqrt{5}) \end{aligned}$$

That is, either $y' = x'$ or $y' = -x' + 2\sqrt{5}$ under the new coordinates. To translate back into the old, substitute $\mathbf{x}' = U^T\mathbf{x}$. You obtain either

$$\begin{aligned}
x' - y' &= 0 \\
\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} &= 0 \\
\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= 0 \\
\begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= 0 \\
3x - y &= 0
\end{aligned}$$

or

$$\begin{aligned}
x' + y' &= 2\sqrt{5} \\
\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} &= 2\sqrt{5} \\
\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= 2\sqrt{5} \\
\begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= 10 \\
x + 3y &= 10
\end{aligned}$$

Figure 8.11 gives the graph.

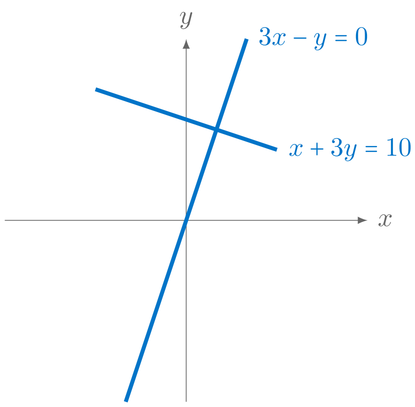


Figure 8.11 *Graph of* $3x^2 - 3y^2 + 8xy - 30x + 10y = 0$.

As noted earlier, the quadratic equation

$$ax^2 + by^2 + cxy + dx + ey = f$$

can be rewritten

$$\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = f$$

where

$$A = \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} d \\ e \end{bmatrix}.$$

We see that the basic shape of the graph of the quadratic equation depends on whether the signs of the eigenvalues of A are the same, are different, or one is 0. Since the determinant of A equals the product of its eigenvalues, $\det A = ab - \frac{1}{4}c^2$ can be used to quickly determine the basic shape of the graph, though it does not determine whether the graph is a degenerate.

Theorem 8.3. The shape of the graph of the quadratic equation $ax^2+by^2+cxy+dx+ey = f$ (a, b, c not all 0) or rewritten $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = f$ as shown above is determined by the following chart:

Shape	**Eigenvalues of A**	**det A**	**Coefficients of quadratic term**
ellipse or degenerate ellipse	same sign	positive	$4ab - c^2 > 0$
hyperbola or degenerate hyperbola	opposite signs	negative	$4ab - c^2 < 0$
parabola or degenerate parabola	an eigenvalue of 0	zero	$4ab - c^2 = 0$

Problem Set 8.3

1. Use Theorem 8.3 to identify the graphs of each of the following equations as an ellipse, hyperbola, parabola, or a degenerate one of those three. Do not distinguish between the degenerates and non-degenerates at this time.

   (a) $2xy = 1$

   (b) $6x^2 + 9y^2 + 4xy = 10$

   (c) $18x^2 + 2y^2 + 12xy + 13x + y = -5$

   (d) $16x^2 + y^2 - 8xy + 8x - 2y = 0$

2. For each equation (a) - (d) in exercise 1, rewrite the equation in the form $\mathbf{x}^T A\mathbf{x}+\mathbf{b}^T\mathbf{x} = f$ where A is 2×2 and symmetric, $\mathbf{b}^T$ is 1×2, f is constant, and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$.

3. For each equation (a) - (d) in exercise 2 there is an orthogonal matrix U that orthogonally diagonalizes A and rotates the coordinate axes by an angle θ where $0 < \theta < 90°$. Find U and θ.

4. For each equation in exercise 1, rewrite the equation in terms of a new coordinate system x' and y' so that the equation has no cross-product term.

5. For each equation (a) - (d) in exercise 1, find one of the following (i) - (iii) depending on the results of exercise 4.

   (i) If the graph is a parabola, find the coordinates of its vertex under the new coordinate system x' and y'.

   (ii) If the graph is an ellipse or hyperbola, find the coordinates of its center under the new coordinate system x' and y'.

   (iii) if the graph is a degenerate conic section, describe its point or line(s) in terms of the new coordinate system x' and y'.

6. For each answer (a) - (d) in exercise 5, describe that answer in terms of the original coordinate system x and y.

7. Sketch the graph of each equation (a) - (d) in exercise 1.

8.4 Quadric Surfaces

The quadric surfaces are the three-dimensional analogs of the conic sections. They are worthy of mention at this point because they are important geometric shapes and they can be analyzed by applying to 3×3 matrices the same techniques applied in section 8.3 to 2×2 matrices.

The quadric surfaces are graphs of quadratic equations in three variables

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz = j$$

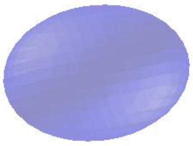
where not all a, b, c, d, e, and f are zero. In Figure 8.12 we list six basic quadric surfaces in standard positions with their equations.

By interchanging the roles of x, y and z we change the orientation of the figures. The surfaces matching these equations are either centered at the origin or have their vertices at the origin. As with the conic sections, the quadric surfaces can be translated by substituting $x - h$, $y - k$, and $z - l$ for x, y, and z respectively.

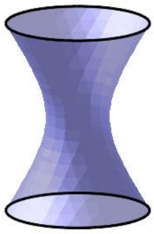
There are several other surfaces that are graphs of quadratic equations in three variables including the parabolic cylinder mentioned in section 8.1 and several direct three-dimensional analogs of the degenerate conic sections. Both the quadric surfaces and their degenerates can be thought of as level surfaces of quadratic forms in three variables just as the conic sections and their degenerates are the level curves of quadratic forms in two variables.

The quadratic equation in three variables

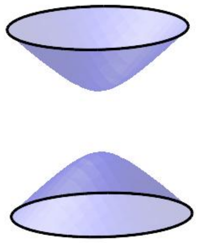
$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz = j$$



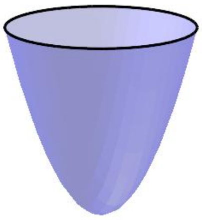
(a) Ellipsoid.
$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$



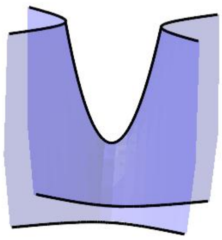
(b) Hyperboloid of one sheet.
$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$



(c) Hyperboloid of two sheets.
$\frac{z^2}{a^2} - \frac{x^2}{b^2} - \frac{y^2}{c^2} = 1$



(d) Elliptic paraboloid.
$z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$



(e) Hyperbolic paraboloid.
$z = \frac{y^2}{a^2} - \frac{x^2}{b^2}$

(f) Elliptic cone.
$z^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2}$

Figure 8.12

can be rewritten as $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} = j$ where

$$A = \begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} f \\ g \\ h \end{bmatrix}.$$

The symmetric matrix A can be orthogonally diagonalized as in section 8.3.

Example 8.11

Determine the shape of the surface $3x^2 - y^2 + 3z^2 - 2xz = 4$.

Solution Rewriting, we have

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 3 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 4.$$

Let

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

The characteristic polynomial of A is $p(\lambda) = (3-\lambda)(-1-\lambda)(3-\lambda) - (-1-\lambda) = -(\lambda -$

$4)(\lambda - 2)(\lambda + 1)$. So the eigenvalues are $\lambda = 4, 2, -1$ with associated eigenvalues of

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

respectively. Let

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \\ 1 & 1 & 0 \end{bmatrix}.$$

Then $\mathbf{x}^T A \mathbf{x} = 4$ simplifies to $4(x')^2 + 2(y')^2 - (z')^2 = 4$ under the substitution $\mathbf{x} = U\mathbf{x}'$ so its graph is a hyperboloid of one sheet.

Problem Set 8.4

For each equation 1-3 below:

(a) Rewrite the equation in terms of new variables x', y', and z' with no cross-product terms.

(b) Describe the graph of the equations as a quadric surface as in Figure 8.12.

1. $2x^2 + y^2 + 2z^2 - 2xz = 9$
2. $5x^2 + 2y^2 + 8z^2 - 8yz - 2\sqrt{5}y - \sqrt{5}z = 0$
3. $x^2 + y^2 + z^2 + 2xy + 2xz + 2yz = 1$

8.5 Positive Definite Matrices

As we have seen through the last five sections, symmetric matrices are very nice. Mostly, what makes them so nice is that all of their eigenvalues are real and they are orthogonally diagonalizable. We now focus our attention on a class of symmetric matrices that are particularly nice and useful in applications.

We start by defining five classes of symmetric matrices.

Definition 8.2. Let A be an $n \times n$ symmetric matrix.

- A is **positive definite** if all of its eigenvalues are positive.
- A is **positive semidefinite** if all of its eigenvalues are greater than or equal to zero.
- A is **negative definite** if all of its eigenvalues are negative.
- A is **negative semidefinite** if all of its eigenvalues are less than or equal to zero.
- A is **indefinite** if it has both positive and negative eigenvalues.

A few facts that are clear from the definition:

- All positive definite matrices are positive semidefinite and all negative definite matrices are negative semidefinite.
- A symmetric matrix is positive definite if and only if $-A$ is negative definite. That same relationship holds between positive semidefinite and negative semidefinite matrices.
- A positive semidefinite matrix is positive definite if and only if it is invertible. The same relationship holds between negative semidefinite and negative definite matrices.
- There is no overlap between the indefinite matrices and any of the other four classes of symmetric matrices.
- The only matrix that is both positive semidefinite and negative semidefinite is the zero matrix.

Definition 8.3. Let A be a symmetric matrix. A quadratic form $\mathbf{x}^T A\mathbf{x}$ is **positive definite** if A is a positive definite matrix. **Positive semidefinite**, **negative definite**, **negative semidefinite**, and **indefinite** quadratic forms are defined similarly.

Theorem 8.4 follows immediately from Theorem 8.2 in section 8.2.

Theorem 8.4. Let A be an $n \times n$ symmetric matrix.

- For all $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} > 0$ if and only if A is positive definite.
- For all $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} \geq 0$ if and only if A is positive semidefinite.
- For all $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} < 0$ if and only if A is negative definite.
- For all $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} \leq 0$ if and only if A is negative semidefinite.
- A is indefinite if and only if there exists $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ such that $\mathbf{x}_1^T A\mathbf{x}_1 > 0$ and $\mathbf{x}_2^T A\mathbf{x}_2 < 0$.

In the case where A is a 2×2 nonzero, symmetric matrix (see section 8.1) the graph of the function $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a(n):

- elliptical paraboloid that opens up if A is positive definite.
- elliptical paraboloid that opens down if A is negative definite.
- parabolic cylinder that opens up if A is positive semidefinite, but not positive definite.
- parabolic cylinder that opens down if A is negative semidefinite, but not negative definite.
- hyperbolic paraboloid if A is indefinite.

We focus mainly on positive definite and positive semidefinite matrices. One place where positive semidefinite matrices arise naturally is by starting with an $m \times n$ matrix A (note A need not be square) and then looking at the $n \times n$ matrix $A^T A$.

Note that the quadratic form $\mathbf{x}^T(A^T A)\mathbf{x}$ can be viewed as a dot product of a vector with itself:

$$\mathbf{x}^T(A^T A)\mathbf{x} = (\mathbf{x}^T A^T)(A\mathbf{x}) = (A\mathbf{x})^T(A\mathbf{x}) = (A\mathbf{x}) \cdot (A\mathbf{x}).$$

The observation above is the key to understand the following theorems and corollaries.

Theorem 8.5. Let A be an $m \times n$ matrix. The matrix $A^T A$ is positive semidefinite.

Proof Since $(A^T A)^T = A^T(A^T)^T = A^T A$, $A^T A$ is symmetric. From the key observation above, $\mathbf{x}^T(A^T A)\mathbf{x} = (A\mathbf{x}) \cdot (A\mathbf{x}) \geq 0$, so $A^T A$ is positive semidefinite by Theorem 8.4. ■

Theorem 8.6. Let A be an $m \times n$ matrix. The null space of A equals the eigenspace of $A^T A$ associated with the eigenvalue $\lambda = 0$.

Proof Let *null* A be the null space of A, and let E_0 be the eigenspace of $A^T A$ associated with the eigenvalue $\lambda = 0$. We show *null* $A = E_0$.

Let $\mathbf{v} \in$ *null* A. Then $A\mathbf{v} = \mathbf{0}$. Multiplying both sides on the left by A^T yields $A^T A\mathbf{v} = A^T\mathbf{0} = \mathbf{0} = 0\mathbf{v}$. Thus $\mathbf{v} \in E_0$ so *null* $A \subseteq E_0$.

Now let $\mathbf{v} \in E_0$. Then $A^T A\mathbf{v} = 0\mathbf{v} = \mathbf{0}$. Multiplying through on the left by $\mathbf{v}^T$ gives $\mathbf{v}^T A^T A\mathbf{v} = \mathbf{v}^T\mathbf{0} = 0$. Now, using the key observation above we have $(A\mathbf{v}) \cdot (A\mathbf{v}) = 0$ which implies $A\mathbf{v} = \mathbf{0}$, and so $\mathbf{v} \in$ *null* A. Thus $E_0 \subseteq$ *null* A. This, together with the paragraph above gives us *null* $A = E_0$. ■

Since every $m \times n$ matrix A has a null space, Theorem 8.6 might lead one to believe that $\lambda = 0$ is always an eigenvalue of $A^T A$. This is not the case. Precisely when the null space of A is the trivial subspace $\{\mathbf{0}\}$, Theorem 8.6 tells us that 0 is not an eigenvalue of $A^T A$ because by definition eigenspaces are never the trivial subspace $\{\mathbf{0}\}$. We can go a step further by noting that $null\ A = \{\mathbf{0}\}$ if and only if $rank\ A = n$. This proves Corollary 8.7.

Corollary 8.7. Let A be an $m \times n$ matrix. The matrix $A^T A$ is positive definite if and only if $rank\ A = n$.

Corollary 8.8 is helpful in section 8.6 and follows quickly from Theorems 8.5 and 8.6.

Corollary 8.8. Let A be an $m \times n$ matrix. The rank of A is equal to r if and only if $A^T A$ has exactly r positive eigenvalues (counting multiplicities).

Proof Since $A^T A$ is positive semidefinite, it has n nonnegative eigenvalues $\lambda_1, \cdots, \lambda_n$ (repeats match their multiplicity), and there is an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $A^T A$, $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$. Without loss of generality, suppose $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ and suppose $\mathbf{u}_i$ is an eigenvector associated with λ_i for $i = 1, \cdots, n$.

Theorem 8.6 tells us that the eigenvectors in $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ associated with the eigenvalue 0 form an orthonormal basis for the null space of A. And since

$$dim(null\ A) = nullity\ A = n - rank\ A$$

we see that the first r eigenvalues $\lambda_1, \cdots, \lambda_r$ are positive ($r = rank\ A$) because only the last $n - r$ eigenvalues $\lambda_{r+1}, \cdots, \lambda_n$ are 0 (this collection is empty if $r = n$). ■

Recall that we adopt the convention of listing eigenvalues of symmetric matrices in decreasing order. Thus, the first eigenvalue is always the biggest, and the last is always the smallest.

Not only is it true that $A^T A$ is positive semidefinite no matter what the matrix A is, in fact, every positive semidefinite matrix has such a factorization. Thus, by examining matrices of the form $A^T A$ we are studying all positive semidefinite matrices. We can find this factorization by orthogonally diagonalizing the positive semidefinite matrix.

Let B be a positive semidefinite matrix. Since B is symmetric, B can be orthogonally diagonalized, so there is a diagonal matrix D and an orthogonal matrix U such that

$$U^T B U = D$$

which, in turn, implies

$$B = U D U^T$$

where the diagonal entries of

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

are the eigenvalues $\lambda_1, \cdots, \lambda_n$ of B. Since B is positive semidefinite, for each i, $\lambda_i \geq 0$, so $\lambda_i = (\sqrt{\lambda_i})(\sqrt{\lambda_i})$. Let

$$S = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{bmatrix}.$$

It is clear that $S = S^T$ and that $D = SS^T$. Thus,

$$\begin{aligned} B &= UDU^T \\ &= U(SS^T)U^T \\ &= (US)(S^TU^T) \\ &= (US)(US)^T. \end{aligned}$$

Let $A = (US)^T$. Then $A^T = US$, so $B = A^TA$.

To this point, if we wish to determine whether a symmetric matrix A is positive definite, we simply calculate the eigenvalues of A and look to see whether they are all positive. That seems simple enough. By this time, we have calculated the eigenvalues of many matrices, but we recognize that this can be difficult, particularly if A is large. For many applications, it turns out to be important to know whether a matrix is positive definite but not so important to know exactly what the eigenvalues are. We won't be discussing those applications in this text, but we briefly discuss a standard method for determining whether a symmetric matrix is positive definite without going through all the work of actually calculating its eigenvalues. This can be very helpful in applications of linear algebra. We begin with 2×2 matrices.

Theorem 8.9. A 2×2 symmetric matrix

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

is positive definite if and only if $a > 0$ and $ac - b^2 > 0$.

Proof Let λ_1 and λ_2 be the eigenvalues of A. Suppose A is positive definite. Then $\lambda_1 > 0$ and $\lambda_2 > 0$. We show that $a > 0$ and $ac - b^2 > 0$. Since $\lambda_1 > 0$ and $\lambda_2 > 0$, we know $ac - b^2 = \det A = \lambda_1\lambda_2 > 0$, so $ac - b^2 > 0$. In addition, $ac - b^2 > 0$ implies $ac > 0$, which implies a and c have the same sign. But $a + c = \mathrm{tr} A = \lambda_1 + \lambda_2 > 0$ implies $a > 0$.

For the other direction, suppose $a > 0$ and $ac - b^2 > 0$. We show A is positive definite. Since $\lambda_1\lambda_2 = \det A = ac - b^2 > 0$, λ_1 and λ_2 have the same sign. In addition, $ac - b^2 > 0$ implies $ac > 0$ which implies a and c have the same sign. Since $a > 0$ we know $c > 0$ as well. Therefore $\lambda_1 + \lambda_2 = \mathrm{tr} A = a + c > 0$. Since λ_1 and λ_2 must have the same sign and their sum is positive, they are both positive and A is positive definite. ■

This makes it easy to make up your own or check on whether a 2×2 symmetric matrix

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

is positive definite. Both a and c must be positive and b^2 must be less than ac. So, for example, both

$$\begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

are positive definite, but

$$\begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 4 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} -5 & 3 \\ 3 & -2 \end{bmatrix}$$

are not.

The way to think about Theorem 8.9 that generalizes to larger symmetric matrices, is as follows. Let A be an $n \times n$ symmetric matrix. Let A_1 be the 1×1 submatrix of A taken from the upper left corner of A. Similarly let A_2 be the 2×2 submatrix of A taken from the upper left corner of A, and in general let A_i be the $i \times i$ submatrix of A taken from the upper left corner of A for $i = 1, 2, \cdots, n$. It should be clear that $A_1 = [a_{11}]$ and $A_n = A$.

Theorem 8.9 tells us that a 2×2 symmetric matrix A is positive definite if and only if $\det A_1 > 0$ and $\det A_2 > 0$. Though we do not prove it in this text, Theorem 8.9 generalizes to Theorem 8.10

Theorem 8.10. An $n \times n$ symmetric matrix A is positive definite if and only if $\det A_i > 0$ (as defined above) for $i = 1, \cdots, n$.

Example 8.12

For which values of x are the following symmetric matrices positive definite?

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 3 & 5 \\ 4 & 5 & x \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 2 & 5 \\ 4 & 5 & x \end{bmatrix}$$

Solution It is easy to see that $\det A_1 = 2$, $\det A_2 = 5$, and $\det A_3 = 5x - 58$, so A is positive definite if and only if $x > 58/5 = 11.6$.

Since $\det B_2 = -7$, B is not positive definite for any value of x.

Problem Set 8.5

1. Determine the definiteness of each of the following symmetric matrices.

   (a) $A = \begin{bmatrix} 3 & 4 \\ 4 & 5 \end{bmatrix}$ (b) $A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}$ (c) $A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

   (d) $A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ (e) $A = \begin{bmatrix} -8 & -2 & -2 \\ -2 & -5 & 4 \\ -2 & 4 & -5 \end{bmatrix}$

2. Determine the definiteness of each of the following quadratic forms.

   (a) $Q(x, y) = 6x^2 + 3y^2 + 4xy$
   (b) $Q(x, y) = x^2 + y^2 + 4xy$
   (c) $Q(x, y, z) = 3y^2 + 4xz$

3. (a) If A is a symmetric matrix, what can you say about the definiteness of A^2?
   (b) If A is a symmetric matrix, when is A^2 positive definite?

4. Recall that a real square matrix A is skew symmetric if $A^T = -A$.

   (a) If A is skew symmetric, is A^2 skew symmetric or symmetric?
   (b) If A is skew symmetric, what can you say about the definiteness of A^2?

5. If A is an invertible symmetric matrix, what is the relationship between the definiteness of A and A^{-1}?

6. A permutation matrix P is a square matrix with entries of 0 and 1 only and has exactly one 1 in each row and in each column.

   (a) Explain why permutation matrices are orthogonal matrices.
   (b) Permutation matrices rearrange the entries of a vector. That is, the entries of $P\mathbf{x}$ are the same as the entries of $\mathbf{x}$ except in a different order (unless $P = I$). Let $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$. Find the permutation matrix P such that $P\mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 5 \\ 2 \end{bmatrix}$.
   (c) Find the 5×5 permutation matrix that swaps the top two entries of a vector in $\mathbb{R}^5$ and leaves every other entry fixed. Also, find the 5×5 permutation matrix that swaps the first and third entry leaving the others fixed.
   (d) By looking at a few examples, describe how $P^T AP$ changes a square matrix A where P is a permutation matrix.
   (e) Use Definition 5.17, Corollary 6.7, Theorem 8.10, and what you have learned about permutation matrices above to explain why all the diagonal entries of a positive definite symmetric matrix must be positive.

7. For each positive semidefinite matrix B below, find a matrix A such that $B = A^T A$.

   (a) $B = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

(b) $B = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

8. In parts (a) - (e) below, let Q be a quadratic form in the variables x, y, and z and A the symmetric matrix such that $Q(x,y,z) = \mathbf{x}^T A\mathbf{x}$. In each case suppose Q and A have the stated properties. In each case describe the quadric surface with equation $Q(x,y,z) = 1$.

   (a) Q is positive definite.

   (b) Q is positive semidefinite and $rank\ A = 2$.

   (c) Q is positive semidefinite and $rank\ A = 1$.

   (d) Q is indefinite and $\det A > 0$.

   (e) Q is indefinite and $\det A < 0$.

8.6 Singular Value Decomposition

We know that there are several reasons why we may wish to diagonalize an $n \times n$ matrix A. We also know that the process of diagonalizing A involves finding a basis for $\mathbb{R}^n$ of vectors that get sent to multiples of themselves by the linear operator $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by the formula $T_A(\mathbf{x}) = A\mathbf{x}$. Since A is square, $\mathbb{R}^n$ serves as both the domain and codomain for T_A, and the basis of eigenvectors proves to be the right basis for $\mathbb{R}^n$ to help with the analysis. In the special case when A is symmetric we know that A is orthogonally diagonalizable.

The singular value decomposition (SVD) gives us a similar way to analyze matrices that need not be diagonalizable. These matrices need not be (and typically are not) square.

If A is $m \times n$ with $m \neq n$, then the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ could not possibly send a vector to a multiple of itself because its domain and codomain are completely different vector spaces. The first thing we need to do to understand SVD is to choose the right basis for the domain of T_A and then choose the right basis for the codomain.

Let A be $m \times n$. We have observed that $A^T A$ is a positive semidefinite symmetric $n \times n$ matrix. Thus, there is an orthonormal basis for $\mathbb{R}^n$ that diagonalizes $A^T A$ and all eigenvalues of $A^T A$ are nonnegative. It turns out that this orthonormal basis is exactly the basis for the domain, $\mathbb{R}^n$, we wish to choose for the SVD of A.

Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be an orthonormal basis of eigenvectors of $A^T A$ with $\mathbf{v}_i$ an eigenvector associated with the eigenvalue λ_i for $i = 1, \cdots, n$ with $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. (Eigenvalues are repeated to match their multiplicity.) Let's take a close look at $\mathcal{B}$.

Let r be the rank of A. By Corollary 8.8 in section 8.5, $\lambda_1, \cdots, \lambda_r$ are positive. If $r = n$, then all the eigenvalues are positive and if $r < n$, then $\lambda_{r+1}, \cdots, \lambda_n$ are all zero.

Since $\{\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$ is a basis for the eigenspace of $A^T A$ associated with the eigenvalue 0, $\{\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$ is an orthonormal basis for the null space of A by Theorem 8.6 of section 8.5 (the empty set ϕ is the basis for *null* A if $r = n$).

Since the vectors in $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ are orthogonal to the vectors in $\{\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$, $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\} \subseteq (null\ A)^\perp = row\ A$. Also $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is linearly independent and since $dim(row\ A) = rank\ A = r$, $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is an orthonormal basis for the row space of A.

In summary, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_r, \cdots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ with $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ an orthonormal basis for *row* A and $\{\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$ (or ϕ if $r = n$) an orthonormal basis for *null* A. Now let's turn to the codomain of T_A, $\mathbb{R}^m$.

The next lemma provides the key observation necessary to construct the right basis for $\mathbb{R}^m$, the codomain for T_A.

Lemma 8.11. Let A be an $m \times n$ matrix. If $\mathbf{v}$ is an eigenvector of $A^T A$ and $\mathbf{w}$ is orthogonal to $\mathbf{v}$ in $\mathbb{R}^n$, then $A\mathbf{w}$ is orthogonal to $A\mathbf{v}$ in $\mathbb{R}^m$.

Proof Suppose $\mathbf{v}$ is an eigenvalue of $A^T A$ associated with the eigenvalue λ and $\mathbf{w}$ is orthogonal to $\mathbf{v}$. Then, $(A\mathbf{w}) \cdot (A\mathbf{v}) = (A\mathbf{w})^T (A\mathbf{v}) = \mathbf{w}^T (A^T A)\mathbf{v} = \mathbf{w}^T (\lambda \mathbf{v}) = \lambda(\mathbf{w}^T \mathbf{v}) = \lambda(\mathbf{w} \cdot \mathbf{v}) = 0$. Therefore $A\mathbf{w}$ is orthogonal to $A\mathbf{v}$ in $\mathbb{R}^m$. ∎

By Lemma 8.11, $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_r\}$ is an orthogonal set of vectors in $\mathbb{R}^m$. The vectors in that set are all nonzero because none of the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_r$ are in the null space of A. That tells us that $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_r\}$ is linearly independent in the column space of A. Finally, since the dimension of *col* A = *rank* $A = r$, $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_r\}$ is an orthogonal basis for *col* A in $\mathbb{R}^m$. To get an orthonormal basis for *col* A, we normalize each of these vectors, so we let $\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i$ for $i = 1, \cdots, r$. Then $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ is an orthonormal basis for *col* A, and $A\mathbf{v}_i = \|A\mathbf{v}_i\| \mathbf{u}_i$ for $i = 1, \cdots, r$. Recall that the column space of A is the range of T_A in $\mathbb{R}^m$.

To expand $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ to an orthonormal basis for $\mathbb{R}^m$ (it already is one if $r = m$) we note that *col* A = *row* A^T and just like *null* $A = (row\ A)^\perp$ we have *null* $A^T = (row\ A^T)^\perp = (col\ A)^\perp$.

Back in chapter 4, we learned how to find a basis for the null space of a matrix, and, in chapter 7 we learned how to use the Gram-Schmidt process to construct an orthonormal basis from that. Since the dimension of *null* A^T is $m - r$, we let $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m\}$ be that basis. (The basis for *null* A^T is ϕ if $r = m$.)

Putting these together we get $\mathcal{C} = \{\mathbf{u}_1, \cdots, \mathbf{u}_r, \cdots, \mathbf{u}_m\}$ is an orthonormal basis for $\mathbb{R}^m$ with $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ an orthonormal basis for *col* A and $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m\}$ an orthonormal basis for *null* A^T.

Observe that for $i = 1, \cdots, n$,

$$\|A\mathbf{v}_i\|^2 = (A\mathbf{v}_i) \cdot (A\mathbf{v}_i) = \mathbf{v}_i^T (A^T A)\mathbf{v}_i = \lambda_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \lambda_i(1) = \lambda_i \geq 0$$

so $\|A\mathbf{v}_i\| = \sqrt{\lambda_i}$.

Definition 8.4. Let A be an $m \times n$ matrix and let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of $A^T A$ with $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ (repeats match multiplicity). For $i = 1, \cdots, n$, let $\sigma_i = \sqrt{\lambda_i}$. The real numbers $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ are called the **singular values** of A.

Note that for $i = 1, \cdots, r$, $A\mathbf{v}_i = \|A\mathbf{v}_i\|\mathbf{u}_i = \sigma_i \mathbf{u}_i \neq \mathbf{0}$ and for $i = r+1, \cdots, n$, $A\mathbf{v}_i = \mathbf{0}$.

Let Σ be the matrix representation of T_A relative to the orthonormal bases $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ and let

$$V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \text{ and } U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}.$$

Then V and U are orthogonal change-of-basis matrices so

$$V^T\mathbf{x} = [\mathbf{x}]_{\mathcal{B}}, V[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}, U^T\mathbf{y} = [\mathbf{y}]_{\mathcal{C}}, \text{ and } U[\mathbf{y}]_{\mathcal{C}} = \mathbf{y}.$$

To find the matrix representation of T_A relative to $\mathcal{B}$ and $\mathcal{C}$ follow the diagram in Figure 8.13.

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "A"] \arrow[d, shift right, "V"'] & \mathbb{R}^m \arrow[d, shift left, "U"] \\
\mathbb{R}^n \arrow[r, "\Sigma"'] \arrow[u, shift right, "V^T"'] & \mathbb{R}^m \arrow[u, shift left, "U^T"]
\end{tikzcd}$$

Figure 8.13 *The matrix representation of T_A relative to $\mathcal{B}$ and $\mathcal{C}$ is $\Sigma = U^T AV$.*

To find what Σ looks like, let $T_\Sigma : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ given by $T_\Sigma(\mathbf{x}) = \Sigma\mathbf{x}$. From chapter 5, recall

$$\Sigma = \begin{bmatrix} T_\Sigma(\mathbf{e}_1) & \cdots & T_\Sigma(\mathbf{e}_n) \end{bmatrix}.$$

For $i = 1, \cdots, r$,

$$\begin{aligned}
T_\Sigma(\mathbf{e}_i) &= \Sigma\mathbf{e}_i \\
&= U^T AV\mathbf{e}_i \\
&= U^T A\mathbf{v}_i \text{ since } [\mathbf{v}_i]_{\mathcal{B}} = \mathbf{e}_i \\
&= U^T(\sigma_i\mathbf{u}_i) \\
&= \sigma_i(U^T\mathbf{u}_i) \\
&= \sigma_i\mathbf{e}_i \text{ since } [\mathbf{u}_i]_{\mathcal{C}} = \mathbf{e}_i
\end{aligned}$$

and for $i = r+1, \cdots, n$,

$$\begin{aligned} T_{\Sigma} &= \Sigma \mathbf{e}_i \\ &= U^T A V \mathbf{e}_i \\ &= U^T A \mathbf{v}_i \\ &= U^T \mathbf{0} \\ &= \mathbf{0}. \end{aligned}$$

Thus,

$$\Sigma = \begin{array}{cc} & \begin{array}{cccc} 1 & & r & m \end{array} \\ \begin{array}{c} 1 \\ \\ r \\ \\ m \end{array} & \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \end{array}$$

Of course, solving $\Sigma = U^T A V$ for A yields

$$A = U \Sigma V^T.$$

This process is summarized in the Definition 8.5.

Definition 8.5. Let A be an $m \times n$ matrix and let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be an ordered orthonormal basis for $\mathbb{R}^n$ of eigenvectors for $A^T A$ with $\mathbf{v}_i$ associated with eigenvalue λ_i such that $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Let $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \cdots, n$. Let r be the rank of A. For $i = 1, \cdots, r$, let $\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i$. If $r < m$, let $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m\}$ be an orthonormal basis for the null space of A^T (if $r = m$, $\mathbf{u}_1, \cdots, \mathbf{u}_m$ are defined above) and let $\mathcal{C} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$. Let $V = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ and $U = [\mathbf{u}_1 \cdots \mathbf{u}_m]$. Finally, let Σ be the $m \times n$ matrix defined by

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right].$$

The factorization $A = U \Sigma V^T$ is called the **singular value decomposition** of A.

Example 8.13

Let

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

Find the singular value decomposition of A.

Solution First,

$$A^T A = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 2 \\ 2 & 2 & 4 \end{bmatrix}.$$

Familiar calculations show that the characteristic polynomial for $A^T A$ is $p(\lambda) = -\lambda^3 + 8\lambda^2 - 12\lambda = \lambda(\lambda-2)(\lambda-6)$, so the eigenvalues for $A^T A$ are $\lambda_1 = 6$, $\lambda_2 = 2$, and $\lambda_3 = 0$ (in decreasing order). Equally familiar calculations show that

$$\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

are associated eigenvectors for $A^T A$ respectively. Thus, the singular values of A are $\sigma_1 = \sqrt{6}$, $\sigma_2 = \sqrt{2}$, and $\sigma_3 = 0$ so

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

By normalizing the three orthogonal eigenvectors above we get vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \text{ and } \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix},$$

the orthonormal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, and the orthogonal matrix

$$V = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix}.$$

Turning to the codomain of A, we get

$$A\mathbf{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} \text{ and } A\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

Normalizing, we get

$$\mathbf{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \text{ and } \mathbf{u}_2 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

To complete the orthonormal basis for $\mathbb{R}^4$, the codomain of T_A, we find a basis for the null space of A^T. Row reducing A^T yields

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

Let $x_2 = s$ and $x_4 = t$. Then

$$\begin{aligned} x_1 &= -s \\ x_2 &= s \\ x_3 &= \quad -t \\ x_4 &= \quad t \end{aligned}$$

yielding a basis for the null space of A^T consisting of

$$\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Since these vectors are already orthogonal, we need not perform Gram-Schmidt. We just normalize to get

$$\mathbf{u}_3 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{bmatrix}, \mathbf{u}_4 = \begin{bmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix},$$

the orthonormal basis $\mathcal{C} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$, and the orthogonal matrix

$$U = \begin{bmatrix} 1/2 & -1/2 & -1/\sqrt{2} & 0 \\ 1/2 & -1/2 & 1/\sqrt{2} & 0 \\ 1/2 & 1/2 & 0 & -1/\sqrt{2} \\ 1/2 & 1/2 & 0 & 1/\sqrt{2} \end{bmatrix}.$$

The singular value decomposition of A is

$$A = U\Sigma V^T$$

or

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & -1/\sqrt{2} & 0 \\ 1/2 & -1/2 & 1/\sqrt{2} & 0 \\ 1/2 & 1/2 & 0 & -1/\sqrt{2} \\ 1/2 & 1/2 & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{3} & -1\sqrt{3} & 1/\sqrt{3} \end{bmatrix}.$$

The example above was chosen because it is relatively simple but illustrates all the aspects of calculating the singular value decomposition except the Gram-Schmidt process.

Though simple compared to other matrices, you probably agree that even this simple example is quite involved. It is important to calculate a small number of realtively "simple" singular value decompositions so you have a clear understanding of how the SVD fits together.

The SVD has many applications. You see two in the next section. As you shall see, it is not always necessary to calculate all of the SVD. We use the ideas made clear by understanding the SVD and minimize the calculations. When the actual SVD is necessary, we can use a software package like *Maple* for the calculations.

Problem Set 8.6

For each $m \times n$ matrix A in exercises 1 through 4:

(a) Find the singular values $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ of A and the $m \times n$ matrix Σ described in Definition 8.5.

(b) Find the $n \times n$ orthogonal matrix V described in Definition 8.5.

(c) Find the $m \times m$ orthogonal matrix U described in Definition 8.5.

(d) Verify that $U\Sigma V^T = A$ for the matrices Σ, V, and U you calculated.

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

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

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

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

5. Suppose A is an $n \times n$ orthogonal matrix. Find a SVD of A by using the same process described in this section and practiced in the previous four exercises. (*Hint:* The standard basis $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is a basis of eigenvectors for a diagonal matrix.) What are the singular values?

6. Suppose A is an $n \times n$ symmetric matrix. What is the relation between the eigenvalues of A and the singular values of A? (**Hint:** A symmetric implies $A^TA = A^2$. Use exercise 14 in section 6.1.)

7. Use the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ to show that the answer for exercise 6 need not hold if A is not symmetric.

8. (a) Show how any square matrix A can be written as

$$A = QS$$

where Q is orthogonal and S is symmetric positive semidefinite. This is called the **polar decomposition** of A. (*Hint:* $A = U\Sigma V^T = UV^TV\Sigma V^T$.)

(b) Is it possible to write $A = S_1Q_1$ where S_1 is symmetric and Q_1 is orthogonal?

9. Find the polar decomposition of $A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$ and A^T.

10. Let A be an $n \times n$ matrix. Prove that A^TA and AA^T are similar.

8.7 Matrix Norms

In a general intuitive way, a vector norm is a means for measuring the length of a vector. As mentioned earlier, this can be done in many ways depending on the vector space, but probably the most natural vector norm and certainly the vector norm that we have focused on in this text is the 2-norm or the Euclidean norm defined for vectors in $\mathbb{R}^n$.

If $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then the Euclidean norm of $\mathbf{x}$ is defined by

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$

We have yet to define what we mean by a matrix norm. Like vector norms, matrix norms can be defined in a variety of ways. Again, we focus on the most common and natural matrix norm and one that is compatible with the Euclidean vector norm.

The idea behind a matrix norm is to measure the amount that the matrix can stretch or shrink a nonzero vector. For example, let $A = \begin{bmatrix} 3 & -1 \\ 4 & 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Then

$$A\mathbf{v} = \begin{bmatrix} 3 & -1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 12 \end{bmatrix}.$$

We have $\|\mathbf{v}\| = \sqrt{9+16} = 5$ and $\|A\mathbf{v}\| = \sqrt{25+144} = 13$, so the matrix A stretches this vector $\mathbf{v}$ by a factor of

$$\frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|} = \frac{13}{5} = 2.6.$$

Unfortunately, if we choose another vector like $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and look at how much A stretches it we get a different result.

Since

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & -1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix},$$

we get

$$\frac{\|A\mathbf{e}_1\|}{\|\mathbf{e}_1\|} = \frac{5}{1} = 5.$$

The norm of a matrix A is defined to be the maximum that ratio can be over all nonzero vectors.

Definition 8.6. Let A be an $m \times n$ matrix. The **matrix norm of** A, denoted $\|A\|$, is defined by

$$\|A\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

In the example above, we know that $\|A\|$ is at least 5 because it stretches at least one vector by a factor of 5, but indeed there may be another vector that it stretches even more.

In a text that describes different kinds of matrix norms, Definition 8.6 might be called the **spectral norm** or the **matrix norm induced by the Euclidean vector norm**. Since this is the only matrix norm we study, as long as there is no confusion, we refer to it as the **matrix norm**. If there could be confusion we call it the spectral norm.

Lemma 8.12 provides some useful observations that follow quickly from the definition of matrix norm.

Lemma 8.12. Let A be an $m \times n$ matrix, $\mathbf{x}$ in Euclidean n-space, and $c \in \mathbb{R}$.

(a) $\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$

(b) For all $\mathbf{x} \in \mathbb{R}^n$, $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$.

(c) $\|A\| \geq 0$ and $\|A\| = 0$ if and only if A is a zero matrix.

(d) $\|cA\| = |c|\|A\|$

(e) For B an $m \times n$ matrix, $\|A + B\| \leq \|A\| + \|B\|$. (triangle inequality)

(f) For B an $n \times p$ matrix, $\|AB\| \leq \|A\|\|B\|$.

Proof

(a) For $\mathbf{x} \neq \mathbf{0}$, $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ is the unit vector in the direction of $\mathbf{x}$, so

$$\begin{aligned} \|A\| &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{1}{\|\mathbf{x}\|} \|A\mathbf{x}\| \\ &= \max_{\mathbf{x} \neq \mathbf{0}} \left\| \frac{1}{\|\mathbf{x}\|} A\mathbf{x} \right\| \\ &= \max_{\mathbf{x} \neq \mathbf{0}} \left\| A\left(\frac{1}{\|\mathbf{x}\|} \mathbf{x} \right) \right\| \\ &= \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|. \end{aligned}$$

(b) Note that the inequality holds if $\mathbf{x} = \mathbf{0}$, and if $\mathbf{x} \neq \mathbf{0}$,

$$\begin{aligned} \|A\mathbf{x}\| &= \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}\|\mathbf{x}\| \\ &\leq \left(\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}\right)\|\mathbf{x}\| \\ &= \|A\|\|\mathbf{x}\|. \end{aligned}$$

(c) Since $\|A\mathbf{x}\| \geq 0$ for all $\mathbf{x}$, $\|A\| = \max_{\|\mathbf{x}\|=1}\|A\mathbf{x}\| \geq 0$. If $A = 0$, then $\|A\mathbf{x}\| = \|\mathbf{0}\| = 0$ for all $\mathbf{x}$, so $\|A\| = \max_{\|\mathbf{x}\|=1} 0 = 0$. If $A \neq 0$, then A has at least one nonzero entry. Suppose $a_{ij} \neq 0$. Then $A\mathbf{e}_j \neq \mathbf{0}$, so $\|A\mathbf{e}_j\| > 0$ which implies $\|A\| = \max_{\|\mathbf{x}\|=1}\|A\mathbf{x}\| > 0$.

(d) Since $(cA)\mathbf{x} = c(A\mathbf{x})$, $\|(cA)\mathbf{x}\| = \|c(A\mathbf{x})\| = |c|\|A\mathbf{x}\|$ by Theorem 7.4(c) in section 7.2, so

$$\begin{aligned} \|cA\| &= \max_{\|\mathbf{x}\|=1}\|(cA)\mathbf{x}\| \\ &= \max_{\|\mathbf{x}\|=1}|c|\|A\mathbf{x}\| \\ &= |c|\cdot\max_{\|\mathbf{x}\|=1}\|A\mathbf{x}\| \\ &= |c|\|A\|. \end{aligned}$$

(e) $\|(A+B)\mathbf{x}\| = \|A\mathbf{x}+B\mathbf{x}\| \leq \|A\mathbf{x}\| + \|B\mathbf{x}\|$ by Theorem 7.4(d). So,

$$\begin{aligned} \|A+B\| &= \max_{\|\mathbf{x}\|=1}\|(A+B)\mathbf{x}\| \\ &\leq \max_{\|\mathbf{x}\|=1}(\|A\mathbf{x}\|+\|B\mathbf{x}\|) \\ &\leq \max_{\|\mathbf{x}\|=1}\|A\mathbf{x}\| + \max_{\|\mathbf{x}\|=1}\|B\mathbf{x}\| \\ &= \|A\|+\|B\|. \end{aligned}$$

(f) By repeated use of property (b),

$$\begin{aligned} \|AB\mathbf{x}\| &\leq \|A\|\|B\mathbf{x}\| \\ &\leq \|A\|\|B\|\|\mathbf{x}\| \end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^p$. So if in addition $\|\mathbf{x}\| = 1$,

$$\|AB\mathbf{x}\| \leq \|A\|\|B\|.$$

Thus,

$$\|AB\| = \max_{\|\mathbf{x}\|=1}\|AB\mathbf{x}\| \leq \|A\|\|B\|.$$

■

In Lemma 8.12, part (a) is an alternate definition of this matrix norm. Part (b) is a "compatibility" property. We say that this matrix norm is compatible with the Euclidean vector norm.

Parts (c), (d), and (e) taken together show that this matrix norm qualifies as a vector space norm (see Definition 7.6 in section 7.1) for the vector space, $\mathcal{M}_{mn}$, of all $m \times n$ matrices (see Definition 4.1 and Example 4.2 of part 11). And since a vector space norm gives us a vector space distance function (Definition 7.7 of section 7.1 and Theorem 7.5 of section 7.2) we now have a way of measuring the distance between two matrices A and B:

$$d(A, B) = \|A - B\|.$$

This opens a whole new way of looking at matrices!

Neither the definition nor Lemma 8.12 tell us how to actually calculate a matrix norm. The singular value decomposition tells us that.

Theorem 8.13. Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$. The norm of A equals the largest singular value of A. That is, $\|A\| = \sigma_1$.

Proof Note that $\|A\mathbf{x}\|^2 = (A\mathbf{x}) \cdot (A\mathbf{x}) = \mathbf{x}^T(A^TA)\mathbf{x}$, so $\max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2 = \lambda_1$ where λ_1 is the largest eigenvalue of the symmetric matrix A^TA by Theorem 8.2 of section 8.2. Thus,

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \sqrt{\lambda_1} = \sigma_1,$$

the largest singular value of A by Definition 8.4 of section 8.6. ■

Example 8.14

From Example 8.13 of section 8.6, if

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

then $\|A\| = \sqrt{6}$. Further, the eigenvector $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ of A^TA associated with the eigenvalue $\lambda_1 = 6$ gets that maximal stretch by A. We can check that:

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix},$$

so

$$\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\sqrt{36}}{\sqrt{6}} = \sqrt{6}.$$

Example 8.15

We saw earlier in this section that if $A = \begin{bmatrix} 3 & -1 \\ 4 & 0 \end{bmatrix}$, then $\|A\| \geq 5$ but we don't know the exact value of $\|A\|$. We can calculate it now. Since

$$A^T A = \begin{bmatrix} 3 & 4 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} 25 & -3 \\ -3 & 1 \end{bmatrix},$$

the characteristic polynomial of $A^T A$ is $p(\lambda) = \lambda^2 - 26\lambda + 16$. Thus, the largest eigenvalue of $A^T A$ is $\lambda_1 = \frac{26+\sqrt{26^2-4\cdot 16}}{2} = 13 + 3\sqrt{17}$. So

$$\|A\| = \sigma_1 = \sqrt{\lambda_1} = \sqrt{13 + 3\sqrt{17}} \approx 5.04.$$

Matrix norms are used for a variety of purposes. We illustrate by using matrix norms to prove that a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is continuous on its entire domain $\mathbb{R}^n$.

Recall from calculus that a function f being continuous at x_0 means $\lim_{x \to x_0} f(x) = f(x_0)$. Intuitively, this says "we can force $f(x)$ to be arbitrarily close to $f(x_0)$ by taking x close enough to x_0."

Let $d(x, y)$ represent the distance between x and y. The definition of continuity is made rigorous by using ϵ and δ.

Definition 8.7. Let f be a function. We say f is **continuous** at x_0 if for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$d(x, x_0) < \delta \implies d(f(x), f(x_0)) < \epsilon.$$

For a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, the domain and codomain are vector spaces and the distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined to be $\|\mathbf{u} - \mathbf{v}\|$. Further, from chapter 5 we learned that if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear transformation, then there is an $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

To show, therefore, that a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$ is continuous at $\mathbf{x}_0$ we must show:

For all $\epsilon > 0$, *there exists* $\delta > 0$ *such that* $\|\mathbf{x} - \mathbf{x}_0\| < \delta \implies \|A\mathbf{x} - A\mathbf{x}_0\| < \epsilon$.

Though understanding the intracacies of ϵ, δ proofs is not easy, the use of matrix norms makes the proof that such a linear transformation is continuous as easy as the easiest ϵ, δ proofs convered in calculus.

Theorem 8.14. All linear transformations $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are continuous at every point in their domain $\mathbb{R}^n$.

Proof Let A be the $m \times n$ matrix such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{x}_0 \in \mathbb{R}^n$. We show T is continuous at $\mathbf{x}_0$.

Suppose A is not a zero matrix. Then $\|A\| > 0$. Fix $\epsilon > 0$. Choose $\delta = \epsilon/\|A\|$. Note that

$$\|A\mathbf{x} - A\mathbf{x}_0\| = \|A(\mathbf{x} - \mathbf{x}_0)\| \le \|A\|\|\mathbf{x} - \mathbf{x}_0\|.$$

So

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}_0\| < \delta \quad &\Longrightarrow \quad \|\mathbf{x} - \mathbf{x}_0\| < \epsilon/\|A\| \\ &\Longrightarrow \quad \|A\|\|\mathbf{x} - \mathbf{x}_0\| < \epsilon \\ &\Longrightarrow \quad \|A\mathbf{x} - A\mathbf{x}_0\| < \epsilon \end{aligned}$$

from the note above. Therefore T is continuous at $\mathbf{x}_0$ if A is not a zero matrix.

If A is a zero matrix then

$$\|A\mathbf{x} - A\mathbf{x}_0\| = \|\mathbf{0} - \mathbf{0}\| = 0 < \epsilon$$

for all $\mathbf{x}$. Thus, any value of $\delta > 0$ would work for any given $\epsilon > 0$. Therefore, T is continuous at $\mathbf{x}_0$ in this case too. ■

Problem Set 8.7

1. Find the norms of the following matrices.

   (a) $\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$ (c) $\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

   (f) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ (g) $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$

2. Let U be an orthogonal $n \times n$ matrix. Prove that $\|U\| = 1$.

3. Let A be an $m \times n$ matrix and U an $m \times m$ orthogonal matrix. Prove that $\|UA\| = \|A\|$. (*Hint:* Use Lemma 8.12(a) and Theorem 7.28 from section 7.6).

4. Use exercise 6 of section 8.6 to show that if A is symmetric with eigenvalues $\lambda_1, \cdots, \lambda_n$, then $\|A\| = \max\{|\lambda_1|, \cdots, |\lambda_n|\}$.

5. Use exercise 7 of section 8.6 to show that $\|A\| = \max\{|\lambda_1|, \cdots, |\lambda_n|\}$ need not be true if A is not symmetric.

6. Show that the spectral norm for 2×2 matrices does not satisfy the parallelogram law (exercise 11 in section 7.1). (*Hint:* Keep things simple. Construct a counterexample using matrices A and B that are easy to calculate with entries of just 0 and 1.)

7. Since the parallelogram law holds for all inner product spaces (exercise 11 in section 7.1), what does the counterexample in exercise 6 tell us about the normed vector space of all 2×2 matrices under the spectral norm?

8. Show that the spectral norm on the space of 2×1 matrices is the same as the Euclidean norm on $\mathbb{R}^2$.

9. Use the singular value decomposition to show that for any singular $n\times n$ matrix A and any $\epsilon > 0$, there is an invertible matrix B such that $\|B - A\| = \epsilon$. This tells us we can find invertible matrices as close to any given singular matrix as we wish. (*Hint:* If $A = U\Sigma V^T$ is the SVD of A with

$$\Sigma = \left[\begin{array}{c|c} \begin{matrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_r\end{matrix} & \mathbf{0} \\ \hline \mathbf{0} & \begin{matrix}0 & & \\ & \ddots & \\ & & 0\end{matrix}\end{array}\right]$$

let $B = U\Gamma V^T$ with

$$\Gamma = \left[\begin{array}{c|c} \begin{matrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_r\end{matrix} & \mathbf{0} \\ \hline \mathbf{0} & \begin{matrix}\epsilon & & \\ & \ddots & \\ & & \epsilon\end{matrix}\end{array}\right]$$

10. Let A be a singular $n\times n$ matrix and I_n the $n\times n$ identity matrix. Prove that $\|I_n - A\| \geq 1$. Problem 9 tells us that all singular matrices have invertible matrices that are arbitrarily close, but the roles of singular and invertible cannot be reversed in this respect. The invertible identity matrix has no singular matrices close to it.

8.8 The Pseudoinverse

In this section we develop the pseudoinverse or Moore-Penrose inverse of a matrix. All matrices have a unique pseudoinverse. If a square matrix is invertible, its pseudoinverse is its inverse. If a matrix has one-sided inverses, its pseudoinverse is one of its one-sided inverses. But even matrices that have neither (two-sided) inverses nor one-sided inverses have unique pseudoinverses.

One big use of matrices is to describe linear transformations. As you recall, an $m \times n$ matrix A is used to describe a linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $T_A(\mathbf{x}) = A\mathbf{x}$. Recall that the function T_A has a unique inverse if and only if A is square and invertible. In that case, $T_A^{-1} = T_{A^{-1}}$. For functions in general (not just linear transformations) it is often helpful to construct partial inverses for those functions that do not have inverses. This is a good way to think of the pseudoinverse. If A is not invertible and B is its pseudoinverse, then T_B is a partial inverse of T_A.

Suppose A and B are two sets. Recall that functions $f : A \longrightarrow B$ and $g : B \longrightarrow A$ are inverses means $g \circ f(a) = a$ for all $a \in A$ and $f \circ g(b) = b$ for all $b \in B$.

Recall too that a function $f : A \longrightarrow B$ has a unique inverse, denoted f^{-1}, if and only if f is both one to one and onto.

Example 8.16

The function $f : \mathbb{R} \longrightarrow \mathbb{R}$ given by $f(x) = x^3$ is a good example of a function that has a unique inverse $f^{-1} : \mathbb{R} \longrightarrow \mathbb{R}$. It is $f^{-1}(x) = \sqrt[3]{x}$.

Example 8.17

Another well-known invertible function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is the exponential function $f(x) = e^x$. Even this function has a little problem in that we must restrict the codomain of f to just $(0, \infty)$ rather than all of $\mathbb{R}$ because f is not onto $\mathbb{R}$. That is, $f : \mathbb{R} \longrightarrow (0, \infty)$. Then $f^{-1} : (0, \infty) \longrightarrow \mathbb{R}$ is $f^{-1}(x) = \ln x$.

Any function that is one to one but not onto will have a unique inverse by simply restricting the codomain of the function to be the range of the function.

Functions that are not one to one, of course, do not have true inverses. However, we can construct partial inverses by restricting the domain of the function to a subset of the entire domain on which the function is one to one.

Example 8.18

The sine function $f(x) = \sin x$ is a periodic function and so, of course, is not one to one. However, if we define a restricted sine function $f : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \longrightarrow [-1, 1]$ by $f(x) = \sin x$, then this resricted sine function does have an inverse. We call that inverse the arcsine function $f^{-1}(x) = \sin^{-1} x$. We say that the arcsine function is a partial inverse of the sine function. It is not a true inverse of the sine function because $\sin^{-1}(\sin \pi) = \sin^{-1} 0 = 0$ and $0 \neq \pi$. But for $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $\sin^{-1}(\sin x) = x$.

In the same way, the square root function is a partial inverse of the squaring function.

Let A be an $m \times n$ matrix. Since the range of T_A is the column space of A, we start by restricting the codomain from $\mathbb{R}^m$ to *col* A. If T_A is an onto function then *col* $A = \mathbb{R}^m$ and the codomain of T_A need not be restricted at all. We recognize that this occurs if and only if *rank* $A = m$.

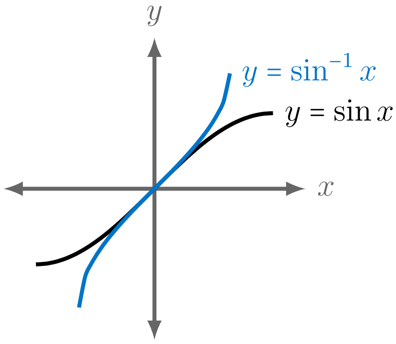


Figure 8.14 *The restricted sine function has an inverse.*

To determine whether the domain needs to be restricted, we again look at the rank of A. Recall that T_A is one to one if and only if the null space of A is $\{\mathbf{0}\}$. So T_A is one to one if and only if the nullity of A is 0. Since $\textit{nullity } A = n - \textit{rank } A$, T_A is one to one if and only if $\textit{rank } A = n$.

If $\textit{rank } A < n$, then T_A is not one to one and the domain of T_A needs to be restricted to a subset of $\mathbb{R}^n$ on which T_A is one to one. When $\textit{rank } A < n$, this can be done in several ways. This situation is similar to the sine function. We restrict the domain of the sine function to $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ because the sine function is one to one on that interval. This allows us to define the arcsine function as a partial inverse of the sine function. If a different interval like $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ were chosen, the result would be a different partial inverse of the sine function. The interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ was chosen long ago because it seemed to be the most natural choice.

If $\textit{rank } A < n$, we wish to make a similar natural choice for the restricted domain on which T_A is one to one. Lemma 8.15 shows that the row space of A is a good choice because T_A is both one to one and onto when the domain of T_A is restricted to the row space of A and the codomain of T_A is restricted to the column space of A. It seems to be a natural choice because the row space of A is one of the fundamental subspaces associated with A.

Lemma 8.15. Let A be an $m \times n$ matrix, let $\textit{row } A$ be the row space of A, and let $\textit{col } A$ be the column space of A. The function T_A given by $T_A(\mathbf{x}) = A\mathbf{x}$ but with domain restricted to $\textit{row } A$, is one to one and onto $\textit{col } A$.

Proof To show that this restricted function is onto $\textit{col } A$, let $\mathbf{y} \in \textit{col } A$. Since $\textit{col } A$ is the range of the unrestricted T_A, there is a vector $\mathbf{w} \in \mathbb{R}^n$ such that $T_A(\mathbf{w}) = \mathbf{y}$. Since $\mathbb{R}^n = \textit{row } A \oplus \textit{null } A$, there exists $\mathbf{u} \in \textit{row } A$ and $\mathbf{v} \in \textit{null } A$ such that $\mathbf{w} = \mathbf{u} + \mathbf{v}$. So

$$T_A(\mathbf{u}) = T_A(\mathbf{u}) + \mathbf{0} = T_A(\mathbf{u}) + T_A(\mathbf{v}) = T_A(\mathbf{u} + \mathbf{v}) = T_A(\mathbf{w}) = \mathbf{y}.$$

Therefore the restricted T_A function is onto $\textit{col } A$.

To show that the restricted T_A is one to one, suppose $\mathbf{u}, \mathbf{v} \in \textit{row } A$ and $T_A(\mathbf{u}) = T_A(\mathbf{v})$. We show $\mathbf{u} = \mathbf{v}$. Since $T_A(\mathbf{u}) = T_A(\mathbf{v})$, $T_A(\mathbf{u} - \mathbf{v}) = T_A(\mathbf{u}) - T_A(\mathbf{v}) = \mathbf{0}$. Thus, $\mathbf{u} - \mathbf{v} \in$

null A. But since *row* A is a subspace of $\mathbb{R}^n$, $\mathbf{u}-\mathbf{v} \in row\ A$ too. But $row\ A \cap null\ A = \{\mathbf{0}\}$, so $\mathbf{u}-\mathbf{v} = \mathbf{0}$. Thus, $\mathbf{u} = \mathbf{v}$. Therefore, the restricted T_A is one to one. ■

Thus for any $m \times n$ matrix A, the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ provides a one-to-one onto function when the domain of T_A is restricted to the row space of A and the codomain is restricted to the column space of A.

Next, we wish to construct an $n \times m$ matrix A^+ with the property that $T_{A^+} : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is a particular partial inverse of T_A in that $T_{A^+} \circ T_A(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in row\ A$ and $T_A \circ T_{A^+}(\mathbf{y}) = \mathbf{y}$ for all $\mathbf{y} \in col\ A$. The singular value decomposition of A can help.

As we know from section 8.6,

$$A = U\Sigma V^T$$

with

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

where $r = rank\ A$ and $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the positive singular values of A. The matrix

$$V = [\mathbf{v}_1 \cdots \mathbf{v}_r \cdots \mathbf{v}_n]$$

is an orthogonal matrix with $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ an orthonormal basis for the row space of A and $\{\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$ an orthonormal basis for the null space of A (this basis is ϕ if $r = n$). The matrix

$$U = [\mathbf{u}_1 \cdots \mathbf{u}_r \cdots \mathbf{u}_m]$$

is an $m \times m$ orthogonal matrix with $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ an orthonormal basis for the column space of A and $T_A(\mathbf{v}_i) = A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ for $i = 1, \cdots, r$ and $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m\}$ is an orthonormal basis for *null* A^T (this basis is ϕ if $r = m$).

Since $T_A(\mathbf{v}_i) = A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ for $i = 1, \cdots, r$, we want to construct A^+ so that $T_{A^+}(\mathbf{u}_i) = A^+\mathbf{u}_i = (1/\sigma_i)\mathbf{v}_i$ for $i = 1, \cdots, r$. That is, in order for T_{A^+} to act as an inverse of T_A between *col* A and *row* A, we want $\mathbf{u}_i \mapsto \frac{1}{\sigma_i}\mathbf{v}_i$ by T_{A^+} since $\mathbf{v}_i \mapsto \sigma_i \mathbf{u}_i$ by T_A.

Definition 8.8. Let A be an $m \times n$ matrix with singular value decomposition

$$A = U\Sigma V^T.$$

Define A^+ as the $n \times m$ (note the dimensions) matrix

$$A^+ = V\Sigma^+ U^T$$

where U and V come from the SVD of A and the $n \times m$ matrix

$$\Sigma^+ = \left[\begin{array}{ccc|c} 1/\sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1/\sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

where $\sigma_1, \cdots, \sigma_r$ are the positive singular values of A. The matrix A^+ is called the **pseudoinverse** or the **Moore-Penrose inverse** of A.

Note that since U and V are orthogonal matrices and

$$\Sigma^+\Sigma = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$$

we have that

$$A^+A = V\left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]V^T.$$

Thus, $T_{A^+} \circ T_A$ acts as an identity on *row* A, and

$$AA^+ = U\left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]U^T$$

so $T_A \circ T_{A^+}$ acts as an identity on *col* A.

Example 8.19

From Example 8.13 of section 8.6, the SVD of A is given by

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & -1/\sqrt{2} & 0 \\ 1/2 & -1/2 & 1/\sqrt{2} & 0 \\ 1/2 & 1/2 & 0 & -1/\sqrt{2} \\ 1/2 & 1/2 & 0 & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}.$$

So the pseudoinverse of A is

$$A^+ = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 & 1/2 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$= \frac{1}{6}\begin{bmatrix} 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

In the exercises you are asked to verify that $T_{A^+} \circ T_A(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in row\ A$ and $T_A \circ T_{A^+}(\mathbf{y}) = \mathbf{y}$ for all $\mathbf{y} \in col\ A$, for this example. So T_{A^+} is a partial inverse of T_A. Also $T_{A^+} \circ T_A(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in null\ A$ and $T_A \circ T_{A^+}(\mathbf{y}) = \mathbf{0}$ for all $\mathbf{y} \in null\ A^T$ for this example.

Once the SVD of a matrix is calculated, probably the easiest way to calculate its pseudoinverse is to use the SVD, and in terms of theory the SVD is an easy way to explain the pseudoinverse. It is not necessary, however, to calculate the SVD in order to find the pseudoinverse. It may, in fact, be easier to calculate the pseudoinverse without going through the SVD because of all the work involved in calculating those orthogonal matrices V and U. Next, we recalculate the pseudoinverse of A from Example 8.19 without the use of the SVD.

Example 8.19 (redone)

We start by finding bases for *row* A and *null* A in the standard way.

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

So $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for *row* A and $\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$ is a basis for *null* A. Name these three vectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ respectively. Since $\mathbb{R}^3 = row\ A \oplus null\ A$,

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^3$, though not an orthonormal basis, and

$$P = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{S}_3$.

Since T_A restricted to *row* A is one to one and onto *col* A, the vectors $\mathbf{u}_1 = A\mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = A\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}$ form a basis for *col* A. We also need a basis *null* A^T. Using the standard procedure,

$$A^T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

So, $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}$ is a basis for *null* A^T. Label these vectors $\mathbf{u}_3$ and $\mathbf{u}_4$ respectively.

Since *row* A^T = *col* A and $\mathbb{R}^4$ = *row* $A^T \oplus$ *null* A^T, the set

$$\mathcal{C} = \left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^4$ and

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

is the change-of-basis matrix from $\mathcal{C}$ to $\mathcal{S}_4$. Let

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

By the actions of A on $\mathcal{B}$ and the diagram in Figure 8.15 it is clear that $S = Q^{-1}AP$ (verify this).

$$\begin{tikzcd} \mathbb{R}^3 \arrow[r, "A"] \arrow[d, shift right, "P^{-1}"'] & \mathbb{R}^4 \arrow[d, shift left, "Q^{-1}"] \\ \mathbb{R}^3 \arrow[r, "S"'] \arrow[u, shift right, "P"'] & \mathbb{R}^4 \arrow[u, shift left, "Q"] \end{tikzcd}$$

Figure 8.15 $S = Q^{-1}AP$

Let A^+ be the pseudoinverse of A. It is completely determined by its action on the basis $\mathcal{C}$ of $\mathbb{R}^4$ and we need $A^+\mathbf{u}_1 = \mathbf{v}_1$, $A^+\mathbf{u}_2 = \mathbf{v}_2$, $A^+\mathbf{u}_3 = \mathbf{0}$, and $A^+\mathbf{u}_4 = \mathbf{0}$. By completing the diagram in Figure 8.15, we see $A^+ = PS^TQ^{-1}$ (see Figure 8.16).

$$\begin{tikzcd}
\mathbb{R}^3 \arrow[r, shift left, "A"] \arrow[d, shift right, "P^{-1}"'] & \mathbb{R}^4 \arrow[l, shift left, "A^+"] \arrow[d, shift left, "Q^{-1}"] \\
\mathbb{R}^3 \arrow[r, shift left, "S"] \arrow[u, shift right, "P"'] & \mathbb{R}^4 \arrow[l, shift left, "S^T"] \arrow[u, shift left, "Q"]
\end{tikzcd}$$

Figure 8.16 $A^+ = PS^TQ^{-1}$

Thus,

$$\begin{aligned}
A^+ &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \left(\frac{1}{6}\right) \begin{bmatrix} 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 3 & 3 \end{bmatrix} \\
&= \frac{1}{6}\begin{bmatrix} 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}.
\end{aligned}$$

In general, if A is an $m\times n$ matrix, you need bases $\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$ for *row* A and $\{\mathbf{v}_{r+1},\cdots,\mathbf{v}_n\}$ for *null* A put together to form

$$\mathcal{B} = \{\mathbf{v}_1,\cdots,\mathbf{v}_r,\cdots,\mathbf{v}_n\},$$

a basis for $\mathbb{R}^n$. Then,

$$P = [\mathbf{v}_1\cdots\mathbf{v}_n]$$

is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{S}_n$.

We use $\{\mathbf{u}_1,\cdots,\mathbf{u}_r\}$ where $\mathbf{u}_i = A\mathbf{v}_i$ for $i = 1,\cdots,r$ for a basis for *col* A = *row* A^T. Then find a basis $\{\mathbf{u}_{r+1},\cdots,\mathbf{u}_m\}$ for *null* A^T. Put these together to form a basis

$$\mathcal{C} = \{\mathbf{u}_1,\cdots,\mathbf{u}_r,\cdots,\mathbf{u}_m\}$$

for $\mathbb{R}^m$ and

$$Q = [\mathbf{u}_1\cdots\mathbf{u}_m]$$

the change-of-basis matrix from $\mathcal{C}$ to $\mathcal{S}_m$.

Next, consider the diagram in Figure 8.17. The matrix representation of T_A with respect to $\mathcal{B}$ and $\mathcal{C}$ is the $m\times n$ matrix

$$S = Q^{-1}AP = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right].$$

$$
\begin{tikzcd}[row sep=large, column sep=huge]
\mathbb{R}^n \arrow[r, "A"] \arrow[d, shift right=1ex, "P^{-1}"'] & \mathbb{R}^m \arrow[d, shift left=1ex, "Q^{-1}"] \\
\mathbb{R}^n \arrow[r, "S"'] \arrow[u, shift right=1ex, "P"'] & \mathbb{R}^m \arrow[u, shift left=1ex, "Q"]
\end{tikzcd}
$$

Figure 8.17 $S = Q^{-1}AP$

Let A^+ be the pseudoinverse of A. Then A^+ is to undo the action of A on *col* A = *row* A^T and send everything in *null* A^T to $\mathbf{0}$, so the matrix representation of T_A with respect to $\mathcal{C}$ and $\mathcal{B}$ is just S^T. Completing the diagram in Figure 8.17, gives the diagram in Figure 8.18. From this, we see that $A^+ = PS^TQ^{-1}$.

$$
\begin{tikzcd}[row sep=large, column sep=huge]
\mathbb{R}^n \arrow[r, shift left=1ex, "A"] \arrow[d, shift right=1ex, "P^{-1}"'] & \mathbb{R}^m \arrow[l, shift left=1ex, "A^+"] \arrow[d, shift left=1ex, "Q^{-1}"] \\
\mathbb{R}^n \arrow[r, shift left=1ex, "S"] \arrow[u, shift right=1ex, "P"'] & \mathbb{R}^m \arrow[l, shift left=1ex, "S^T"] \arrow[u, shift left=1ex, "Q"]
\end{tikzcd}
$$

Figure 8.18 $A^+ = PS^TQ^{-1}$

Example 8.20

Find the pseudoinverse of

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

Solution Row reduction gives

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

So $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for *row* A and $\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for *null* A. Name these four vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and $\mathbf{v}_4$ respectively. Let

$$P = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Let

$$\mathbf{u}_1 = A\mathbf{v}_1 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ and } \mathbf{u}_1 = A\mathbf{v}_1 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}.$$

Since $rank\ A^T = rank\ A = 2$, we know $col\ A = \mathbb{R}^2$, so $null\ A^T = \{\mathbf{0}\}$. Let

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

Since $rank\ A^T = rank\ A = 2$, $col\ A = \mathbb{R}^2$ and $null\ A^T = \{\mathbf{0}\}$. Let $Q = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$. Then,

$$\begin{aligned} A^+ &= PS^TQ^{-1} \\ &= \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\left(\frac{1}{8-3}\right)\begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix} \\ &= \frac{1}{5}\begin{bmatrix} 4 & -1 \\ -3 & 2 \\ 1 & 1 \\ -3 & 2 \end{bmatrix}. \end{aligned}$$

Least Squares and the Pseudoinverse

Suppose A is an $m \times n$ matrix of rank r and let $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be the linear transformation defined by $T_A(\mathbf{x}) = A\mathbf{x}$. We know that both the row space of A in the domain of T_A (i.e. $\mathbb{R}^n$), and the column space of A in the codomain of T_A, (i.e. $\mathbb{R}^m$) have dimension r and that if the domain of T_A is restricted to $row\ A$ and the codomain of T_A is resricted to $col\ A$, then the restricted T_A is a bijection from $row\ A$ to $col\ A$. We also know that T_{A^+}, where A^+ is the pseudoinverse of A, is the partial inverse of T_A that undoes what T_A does between $row\ A$ and $col\ A$. Thus, for all $\mathbf{u} \in row\ A$ and $\mathbf{w} \in col\ A$,

$$A\mathbf{u} = \mathbf{w} \text{ if and only if } A^+\mathbf{w} = \mathbf{u}.$$

We wish to investigate how A acts on other vectors in $\mathbb{R}^n$ and how A^+ acts on other vectors in $\mathbb{R}^m$.

To begin, recall that the null space of A is the orthogonal complement of the row space of A ($null\ A = (row\ A)^\perp$), and by the definition of null space if $\mathbf{v} \in null\ A$, then $A\mathbf{v} = \mathbf{0}$. Also, by design, $null\ A^+ = null\ A^T = (col\ A)^\perp$, so if $\mathbf{z} \in null\ A^T$, then $A^+\mathbf{z} = \mathbf{0}$.

Beyond that, since $\mathbb{R}^n = row\ A \oplus null\ A$, if $\mathbf{x} \in \mathbb{R}^n$, there exists unique $\mathbf{u} \in row\ A$ and $\mathbf{v} \in null\ A$ such that $\mathbf{x} = \mathbf{u} + \mathbf{v}$, and $A\mathbf{x} = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = A\mathbf{u} + \mathbf{0} = A\mathbf{u}$. See Figure 8.19.

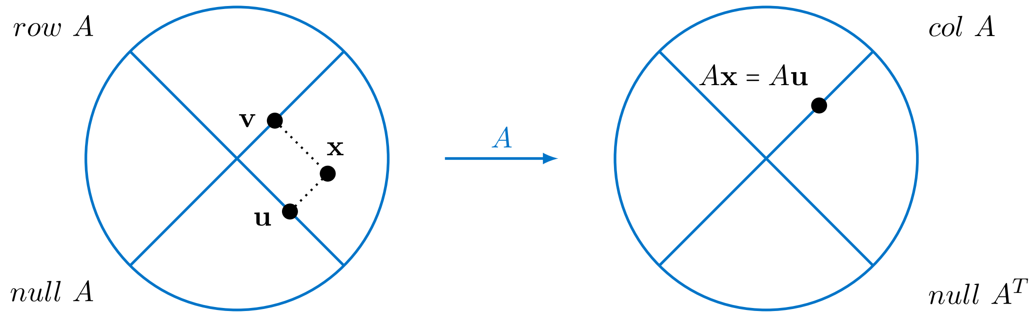


Figure 8.19 $A\mathbf{x} = A\mathbf{u}$

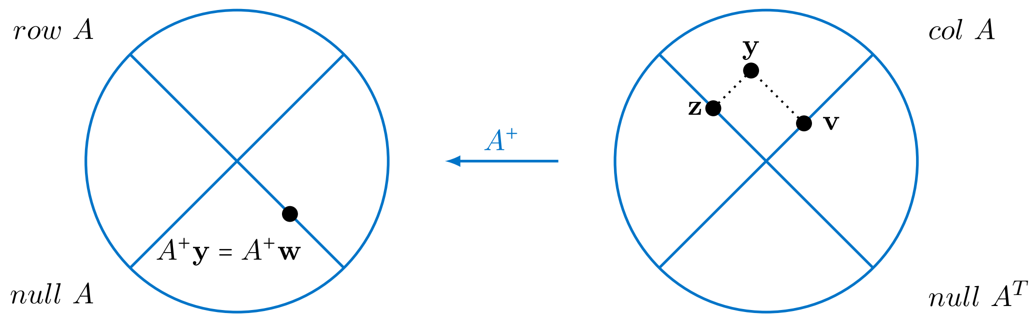


Figure 8.20 $A^+\mathbf{y} = A^+\mathbf{w}$

Similarly, if $\mathbf{y} \in \mathbb{R}^m$, then there exists unique $\mathbf{w} \in col\ A$ and $\mathbf{z} \in null\ A^T$ such that $\mathbf{y} = \mathbf{w} + \mathbf{z}$ and

$$A^+\mathbf{y} = A^+(\mathbf{w} + \mathbf{z}) = A^+\mathbf{w} + A^+\mathbf{z} = A^+\mathbf{w} + \mathbf{0} = A^+\mathbf{w}.$$

See Figure 8.20.

We know that if $\mathbf{b} \in col\ A$, then the system $A\mathbf{x} = \mathbf{b}$ has at least one solution, and because of the bijection discussed above, $A^+\mathbf{b}$ is one of them. That is the only solution if $rank\ A = n$ (i.e. $null\ A = \{\mathbf{0}\}$), but if $rank\ A < n$, there are other solutions. See Figure 8.21.

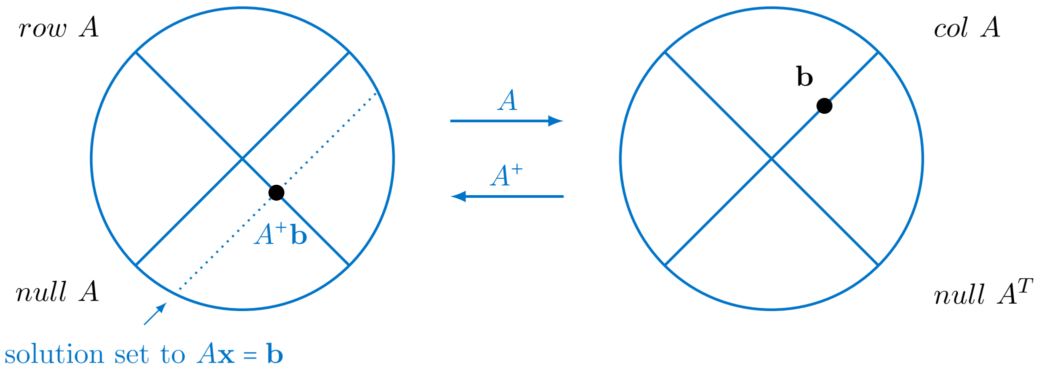


Figure 8.21 $A^+\mathbf{b}$ *is one solution to* $A\mathbf{x} = \mathbf{b}$.

If $\mathbf{b} \notin col\ A$, then $A\mathbf{x} = \mathbf{b}$ has no solution, but it does have a least-squares solution. The least-squares solution to $A\mathbf{x} = \mathbf{b}$ is exactly the same as the solution to $A\mathbf{x} = \widehat{\mathbf{b}}$ where $\widehat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto $col\ A$. As noted in Figure 8.22, $A^+\mathbf{b} = A^+\widehat{\mathbf{b}}$, so $A^+\mathbf{b}$ is a least-squares solution to $A\mathbf{x} = \mathbf{b}$. That is the only least-squares solution if $rank\ A = n$, but if $rank\ A < n$, then it is the only least-squares solution that is in $row\ A$. By the Pythagorean theorem, the other solutions have a larger norm. See Figure 8.22.

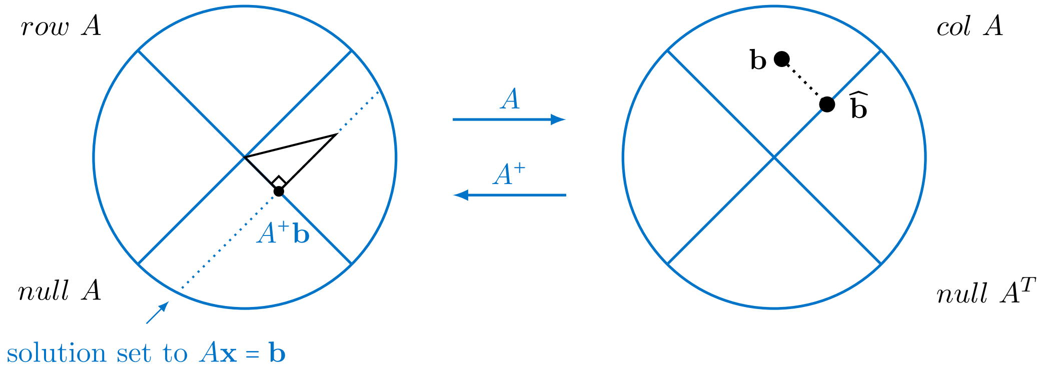


Figure 8.22 $A^+\mathbf{b}$ *is a least-squares solution to* $A\mathbf{x} = \mathbf{b}$.

We summarize with the following theorem that is proved in the exercises.

Theorem 8.16. Let A be an $m \times n$ matrix, A^+ its pseudoinverse, and $\mathbf{b} \in \mathbb{R}^m$. The vector $A^+\mathbf{b}$ is the only least-squares solution to the system $A\mathbf{x} = \mathbf{b}$ that lies in the row space of A. The system has more than one least-squares solution if and only if $rank\ A < n$ in which case $A^+\mathbf{b}$ is the unique least-squares solution with minimum Euclidean norm.

Example 8.21

Find the least-squares solutions to the system

$$\begin{aligned} x + y &= 2 \\ x + 2y &= 1 \\ x - y &= 1 \end{aligned}$$

using the method of section 7.5 and by calculating the pseudoinverse. Compare the results.

Solution The system in matrix form is

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & -1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$. The normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Using Cramer's rule to solve the normal system, we get

$$x = \frac{\begin{vmatrix} 4 & 2 \\ 3 & 6 \end{vmatrix}}{\begin{vmatrix} 3 & 2 \\ 2 & 6 \end{vmatrix}} = \frac{18}{14} = \frac{9}{7}$$

and

$$y = \frac{\begin{vmatrix} 3 & 4 \\ 2 & 3 \end{vmatrix}}{\begin{vmatrix} 3 & 2 \\ 2 & 6 \end{vmatrix}} = \frac{1}{14},$$

so

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 18 \\ 1 \end{bmatrix}$$

is the least-squares solution. Now the pseudoinverse.

Reduce A to find bases for *row* A and *null* A:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & -1 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

so $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ form a basis for *row* A. The null space of A is $\{\mathbf{0}\}$ since *rank* $A = 2$. Let $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$. The vectors $\mathbf{u}_1 = A\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = A\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ form a basis for *col* A.

Reduce A^T to find a basis for *null* A^T:

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

so $\mathbf{u}_3 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$ forms a basis for *null* A^T. Let $Q = \begin{bmatrix} 1 & 1 & -3 \\ 1 & 2 & 2 \\ 1 & -1 & 1 \end{bmatrix}$. Then the pseudoinverse of A is

$$A^+ = PS^TQ^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\frac{1}{14}\begin{bmatrix} 4 & 2 & 8 \\ 1 & 4 & -5 \\ -3 & 2 & 1 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 4 & 2 & 8 \\ 1 & 4 & -5 \end{bmatrix}.$$

Finally,

$$A^+\mathbf{b} = \frac{1}{14}\begin{bmatrix} 4 & 2 & 8 \\ 1 & 4 & -5 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 18 \\ 1 \end{bmatrix},$$

the same as above.

Problem Set 8.8

1. For each matrix A below, find its pseudoinverse A^+.

   (a) $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

   (b) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$

   (e) $\begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

2. For A the 4×3 matrix in Example 8.19, show the following:

   (a) $T_{A^+} \circ T_A(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in$ *row* A.
   (*Hint:* By Theorem 5.2 of section 5.1, you need only show that the composition acts as the identity on a basis for the subspace to know it acts as the identity on the whole subspace. The basis in Example 8.19 (redone) is easier to work with.)

   (b) $T_A \circ T_{A^+}(\mathbf{y}) = \mathbf{y}$ for all $\mathbf{y} \in$ *col* A.
   (*Hint:* Similar to part (a) as are the last two parts.)

   (c) $T_{A^+} \circ T_A(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in$ *null* A.

   (d) $T_A \circ T_{A^+}(\mathbf{y}) = \mathbf{0}$ for all $\mathbf{y} \in$ *null* A^T.

3. Let A be an $m \times n$ matrix and $\mathbf{b}$ be from the column space of A.

   (a) Use Lemma 8.15 to prove that there exists a unique vector $\mathbf{v}$ from the row space of A that is a solution to the system $A\mathbf{x} = \mathbf{b}$.

   (b) Prove that $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$ if and only if there is a vector $\mathbf{z}$ from the null space of A such that $\mathbf{w} = \mathbf{v} + \mathbf{z}$ where $\mathbf{v}$ is the solution to $A\mathbf{x} = \mathbf{b}$ from *row* A (see part (a)).

   (c) Use parts (a) and (b) and the Pythagorean theorem to prove that the solution to $A\mathbf{x} = \mathbf{b}$ from the row space of A (the vector $\mathbf{v}$ in part (a)) is the unique solution to $A\mathbf{x} = \mathbf{b}$ with a minimum norm.
   (*Hint:* Recall that *null* $A = ($*row* $A)^\perp$.)

4. Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. (Note that in exercise 3, $\mathbf{b}$ came from the column space of A. This is different.) Prove that $A^+\mathbf{b}$ is the least-squares solution to $A\mathbf{x} = \mathbf{b}$ with the smallest norm.

5. Prove that A^+A is the orthogonal projection matrix from $\mathbb{R}^n$ to *row* A.

6. Prove that AA^+ is the orthogonal projection from $\mathbb{R}^m$ to *col* A.

7. Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 2 & 1 & 3 \end{bmatrix}$. Use exercises 5 and 6 to find the projection matrix:

   (a) From $\mathbb{R}^3$ onto the row space of A.

   (b) From $\mathbb{R}^3$ onto the column space of A.

8. Let A be an $m \times n$ matrix and A^+ its pseudoinverse. Prove that $(A^T)^+ = (A^+)^T$. (*Hint:* Start with the SVD of A and form both $(A^T)^+$ and $(A^+)^T$ from it. Note that $(\Sigma^+)^T = (\Sigma^T)^+$.)

Selected Answers

Problem Set 1.1

1. **(a)** $3x_1 - 13x_2 - 4x_3 = -11$

(c) not linear

(e) $2x + 3y = \ln 5$

2. **(a)** 2×2, yes, no

(c) 2×4, yes, yes, yes, yes, no

3. **(a)** $(2/7, 16/7)$

(c) no solution

4. **(a)** $\begin{bmatrix} 4 & -15 & -8 \\ -2 & 1 & 0 \\ -1 & 7 & 3 \end{bmatrix}, \begin{bmatrix} 4 & -15 & -8 & -3 \\ -2 & 1 & 0 & 3 \\ -1 & 7 & 3 & 0 \end{bmatrix}$

5. **(a)** $\begin{array}{rcrcr} 4x & - & 15y & = & -6 \\ -2x & + & y & = & 0 \\ -x & + & 7y & = & 3 \end{array}$

6. **(a)** $y = -\frac{5}{42}x^2 + \frac{1}{2}x + \frac{73}{21}$

(b) $x^2 + y^2 - 4x + 2y = 20$

Problem Set 1.2

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

2. **(a)** $\begin{bmatrix} 3 & -1 & 1 & 6 \\ 2 & 1 & 5 & 8 \\ 1 & -1 & -2 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

(c) $\begin{array}{rcr} x & = & 4 \\ y & = & 5 \\ z & = & -1 \end{array}$

(d) $(4, 5, -1)$

(e) yes

3. If allowed to scale by 0, the row operation $r_i \to 0r_i$ results in the underlying equation becoming $0 = 0$. It is easy to construct examples of linear systems in which such a row operation would expand the solution set since the equation $0 = 0$ is always true and thus is satisfied by any ordered n-tuple. Since $r_i + 0r_j = r_i$, this elementary row operation does not change the row at all, thus it changes neither the underlying linear equation nor the solution set. Since it does not change the solution set it can be allowed for solving linear systems, but since it does not change the underlying system, it is not helpful in solving the system.

Problem Set 1.3

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

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

(e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

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

(i) $\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

2. **(a)** 8 **(c)** 2^n

(d) Each $(n-1)\times(n-1)$ category extends to two $n \times n$ categories bases on whether the last column of the $n \times n$ matrix is a pivot column.

Problem Set 1.4

1. $\begin{aligned} x &= 3 \\ y &= 2 \\ z &= 4 \end{aligned}$ or $(3,2,4)$

3. $\begin{aligned} x &= 5 - 3t \\ y &= t \\ z &= 2 \end{aligned}$

5. inconsistent

7. $\begin{aligned} x &= 4 \\ y &= 2 \end{aligned}$ or $(4,2)$

9. $\begin{aligned} x &= 5 \\ y &= 4 \end{aligned}$ or $(5,4)$

11. $\begin{aligned} x &= 1 - 2t \\ y &= 3 + t \\ z &= t \end{aligned}$

13. $\begin{aligned} x_1 &= -2 \\ x_2 &= 3 \\ x_3 &= 1 \\ x_4 &= -4 \end{aligned}$ or $(-2,3,1,-4)$

15. $\begin{aligned} x &= 3t \\ y &= 0 \\ z &= t \end{aligned}$

17. $\begin{aligned} x &= -2 \\ y &= 1 \\ z &= 1 \end{aligned}$ or $(-2, 1, 1)$

19. $y = x^3 - 2x^2 + 3x - 4$

21. If $b = 2a$, $\begin{aligned} x &= a - 2t \\ y &= t \end{aligned}$ If $b \neq 2a$, inconsistent

Problem Set 1.5

1. **(a)** $\begin{bmatrix} 13 \\ -4 \\ 31 \end{bmatrix}$ **(b)** $\begin{bmatrix} 13 \\ -4 \\ 31 \end{bmatrix}$, equivalence of linear combinations and matrix multiplication

3. **(a)** $\begin{bmatrix} 5 & 2 \\ 1 & -1 \end{bmatrix}$ **(b)** $\begin{bmatrix} 1 & 2 \\ 5 & 3 \end{bmatrix}$, matrix multiplication is not commutative

5. **(a)** $\begin{bmatrix} 5 & 1 & 6 \\ -3 & -1 & 2 \end{bmatrix}$ **(b)** undefined, matrix multiplication is not commutative

7. $x\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + y\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} + z\begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$, $\begin{bmatrix} 1 & -1 & 3 \\ 2 & -1 & 6 \\ 3 & -1 & 9 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$,

solution: $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$

9. $\begin{aligned} x_1 + x_2 + 3x_3 + 4x_4 &= 3 \\ 2x_1 + 3x_2 + 8x_3 + 11x_4 &= 7 \\ -x_1 + x_2 + x_3 + 2x_4 &= 2 \end{aligned}$

$x_1\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + x_2\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 3 \\ 8 \\ 1 \end{bmatrix} + x_4\begin{bmatrix} 4 \\ 11 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix}$, inconsistent

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

Problem Set 1.6

1. $\begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}$

3. singular

5. $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$

7. $\begin{bmatrix} -6 & 3 & 5 \\ 5 & -3 & -4 \\ -1 & 1 & 1 \end{bmatrix}$

9. $\begin{bmatrix} -6/67 & 18/67 & -11/67 \\ 1/67 & -3/67 & 13/67 \\ 26/67 & -11/67 & 3/67 \end{bmatrix}$

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

13. $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (2b_1 + 7b_2)/15 \\ (-b_1 + 4b_2)/15 \end{bmatrix}$

14. **(a)** $E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 1 & -2 \end{bmatrix}$

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

15. **(a)** $\begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -3 \\ 2 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -3 \\ 0 & 7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 7 \end{bmatrix}\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 3/7 & 1/7 \\ 1/7 & -2/7 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1/7 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Problem Set 2.1

1. **(a)** $\begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix}$

(b) $\sqrt{38}$

(c) $\begin{bmatrix} -2 \\ -5 \\ 3 \end{bmatrix}$

3. $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ doesn't, but $\begin{bmatrix} -1 \\ 4 \\ 5 \end{bmatrix}$ does.

4. **(a)** $\begin{bmatrix} 3 \\ -2 \\ 4 \end{bmatrix}$

(c) $\begin{bmatrix} -t \\ r \\ s \end{bmatrix}$

5. **(a)** $5\mathbf{e_1} + 3\mathbf{e_2} - 7\mathbf{e_3}$

(c) $x\mathbf{e_1} + y\mathbf{e_2} + z\mathbf{e_3}$

6. **(a)** 3

(c) $5|t|$

7. **(a)** $\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$ and $\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}$

8. **(a)** $\begin{bmatrix} 5/13 \\ -12/13 \end{bmatrix}$

(c) $\begin{bmatrix} 2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$ if $t > 0$, $\begin{bmatrix} -2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$ if $t < 0$, and undefined if $t = 0$

10. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. Since $\mathbf{u} - \mathbf{v} = \begin{bmatrix} u_1 - v_1 \\ u_2 - v_2 \end{bmatrix}$, by the definition of norm

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}.$$

But, by the definition of distance between two vectors,

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}.$$

So,

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Problem Set 2.2

1. **(a)** -4

(c) 15/2

2. **(a)** 45°

(c) $\cos^{-1} 3/5 \approx 127°$

3. **(a)** $\begin{bmatrix} 28/13 \\ 42/13 \end{bmatrix}$

(c) $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

5. $\cos\alpha = 3/\sqrt{14}$, $\cos\beta = -1/\sqrt{14}$, $\cos\gamma = 2/\sqrt{14}$, $\alpha = \cos^{-1}(3/\sqrt{14}) \approx 37°$, $\beta = \cos^{-1}(-1/\sqrt{14}) \approx 106°$, $\gamma = \cos^{-1}(2/\sqrt{14}) \approx 58°$.

7. **(a)** $\cos^{-1}(1/\sqrt{3}) \approx 55°$

8. $\angle QPR = \cos^{-1}(25/\sqrt{45}\sqrt{19}) \approx 31°$, $\angle PQR = \cos^{-1}(20/\sqrt{45}\sqrt{14}) \approx 37°$, $\angle PRQ = \cos^{-1}(-6/\sqrt{19}\sqrt{14}) \approx 112°$

9. **(a)** 3/2

(c) 1/5 (-5 is an extraneous root)

Problem Set 2.3

1. **(a)** $\begin{bmatrix} -7 \\ 8 \\ -3 \end{bmatrix}$

3. $\sqrt{66}$

5. 35

7. $2(\mathbf{v} \times \mathbf{u})$

Problem Set 2.4

1. **(a)** i. $\mathbf{x}(t) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + t \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$.

ii. $\begin{aligned} x &= 1 + 4t \\ y &= 2 + 5t \\ z &= 3 + 6t \end{aligned}$

iii. $\frac{x-1}{4} = \frac{y-2}{5} = \frac{z-3}{6}$

(c) i. $\mathbf{x}(t) = \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix}$.

ii. $\begin{aligned} x &= 3 \\ y &= 4 + t \\ z &= -1 + 5t \end{aligned}$

iii. $\frac{y-4}{1} = \frac{z+1}{5}$; $x = 3$

2. **(a)** intersect at the point $\begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix}$

(c) parallel, no intersection

3. **(a)** $\sqrt{66}/3$

(c) 0

4. **(a)** $\mathbf{x}(t) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

(b) $\begin{aligned} x &= 1 \\ y &= 2 + t \\ z &= 3 \end{aligned}$

(c) $x = 1$; $z = 3$

Problem Set 2.5

1. **(a)** $2x + 5y - 4z = -13$

(c) $2x + 5y - 11z = 42$

(e) $2x + 3y - z = -4$

(g) $3x - 5y - 7z = -14$

2. **(b)** $\begin{bmatrix} 2 \\ 5 \\ -4 \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} \right) = 0$

(d) $z - 0 = \frac{1}{2}(x - 1) + \frac{5}{4}(y + 3)$

(f) $\mathbf{x}(s,t) = \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} + s \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$

3. $\mathbf{x}(t) = \begin{bmatrix} -5 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 17 \\ -5 \\ 1 \end{bmatrix}$

5. $(-14, 4, 3)$

7. $23\sqrt{29}/29$

8. **(a)** $2\sqrt{30}$

9. **(a)** A plane and a line are parallel if they do not intersect. One way to show they are parallel is to substitute the values of x, y, and z in terms of t from the parametric equations for the line into the equation for the plane and show that the equation has no solution. Another way is to start by showing that the direction vector for the line is orthogonal to the normal vector of the plane. That implies that the line and plane are parallel or the line lies on the plane. Then pick any fixed point on the line and show that it does not satisfy the equation of the plane.

10. **(a)** Two lines are parallel if they have parallel direction vectors and they are not coincident. Show the direction vectors are multiples of each other, and show that one fixed point from one line is not on the other line.

11. **(a)** It is a horizontal plane, that is, a plane perpendicular to the z axis.

(c) It is a vertical plane, that is, a plane that is parallel to or contains the z axis.

12. **(a)** A (vertical) plane with a normal vector $\begin{bmatrix} b \\ -a \\ 0 \end{bmatrix}$ that passes through the point (x_0, y_0, z_0). The second is a plane with a normal vector of $\begin{bmatrix} 0 \\ c \\ -b \end{bmatrix}$ that also passes through the point (x_0, y_0, z_0).

Problem Set 3.1

1. **(a)** $A_{1,1} = \begin{bmatrix} 0 & 4 \\ 6 & 3 \end{bmatrix}$, $A_{1,2} = \begin{bmatrix} 7 & 4 \\ 2 & 3 \end{bmatrix}$, $A_{1,3} = \begin{bmatrix} 7 & 0 \\ 2 & 6 \end{bmatrix}$,

$A_{2,1} = \begin{bmatrix} 8 & 9 \\ 6 & 3 \end{bmatrix}$, $A_{2,2} = \begin{bmatrix} 1 & 9 \\ 2 & 3 \end{bmatrix}$, $A_{2,3} = \begin{bmatrix} 1 & 8 \\ 2 & 6 \end{bmatrix}$,

$A_{3,1} = \begin{bmatrix} 8 & 9 \\ 0 & 4 \end{bmatrix}$, $A_{3,2} = \begin{bmatrix} 1 & 9 \\ 7 & 4 \end{bmatrix}$, $A_{3,3} = \begin{bmatrix} 1 & 8 \\ 7 & 0 \end{bmatrix}$.

(c) $C_{1,1} = -24$, $C_{1,2} = -13$, $C_{1,3} = 42$,
$C_{2,1} = 30$, $C_{2,2} = -15$, $C_{2,3} = 10$,
$C_{3,1} = 32$, $C_{3,2} = 59$, $C_{3,3} = -56$.

2. **(a)** -7

(c) 40

3. **(a)** -15

5. **(a)** 0

6. **(a)** 80

(c) 40

(e) 0

(g) $-6! = -720$

Problem Set 3.2

1. **(a)** 0

2. **(a)** 13 **(c)** 5 **(e)** 11/15

3. **(a)** 7 **(c)** 14

Problem Set 3.3

1. **(a)** invertible **(c)** invertible

2. **(a)** 20 **(c)** 125 **(e)** 1/4

3. **(a)** $x = 1, 4$

4. **(a)** $\begin{vmatrix} a & b & c \\ e & f & g \\ i & j & k \end{vmatrix} + \begin{vmatrix} a & b & d \\ e & f & h \\ i & j & l \end{vmatrix}$

(c) $\begin{vmatrix} a & c & e \\ g & i & k \\ m & p & r \end{vmatrix} + \begin{vmatrix} a & c & f \\ g & i & l \\ m & p & s \end{vmatrix} + \begin{vmatrix} a & d & e \\ g & j & k \\ m & q & r \end{vmatrix} + \begin{vmatrix} a & d & f \\ g & j & l \\ m & q & s \end{vmatrix} + \begin{vmatrix} b & c & e \\ h & i & k \\ n & p & r \end{vmatrix} + \begin{vmatrix} b & c & f \\ h & i & l \\ n & p & s \end{vmatrix} + \begin{vmatrix} b & d & e \\ h & j & k \\ n & q & r \end{vmatrix} + \begin{vmatrix} b & d & f \\ h & j & l \\ n & q & s \end{vmatrix}$

Problem Set 3.4

1. **(a)** $\begin{bmatrix} 51/23 \\ -2/23 \end{bmatrix}$

3. $x = \dfrac{ed - bf}{ad - bc}$, $y = \dfrac{af - ec}{ad - bc}$

4. **(a)** 13

5. 3

7. 12

Problem Set 4.1

1. Yes

3. **(a)** done

(c) A4, A5, A3, A10, A8, property of $\mathbb{R}$, A10, A5

(e) A4, A5, A3, A10, A8, property of $\mathbb{R}$, Thm. 2c, A4

4. **(a)** yes **(c)** yes

5. **(a)** yes **(c)** no

6. **(a)** no **(c)** yes

7. **(a)** yes **(c)** yes

9. **(a)** yes **(c)** yes

Problem Set 4.2

1. $2x + y - 5z = 0$

3. $\dfrac{x}{4} = \dfrac{y}{3} = \dfrac{z}{2}$

5. $\operatorname{span}\left\{\begin{bmatrix} 2 \\ -3 \end{bmatrix}\right\}$

7. **(a)** i. implicit: $3x - 2y - z = 0$, explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}\right\}$

ii. implicit: $x - 3y + z = 0$, explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 & 2 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 8 \end{bmatrix}\right\}$

iii. implicit: $\dfrac{x}{1} = \dfrac{y}{-3} = \dfrac{z}{1}$, explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}\right\}$

(c) i. $\mathbb{R}^3$

ii. $\mathbb{R}^3$

iii. $\{\mathbf{0}\}$

(e) i. $\mathbb{R}^2$

ii. implicit: $x - 2y + z = 0$, explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 5 \end{bmatrix}\right\}$

iii. implicit: $\dfrac{x}{1} = \dfrac{y}{-2} = \dfrac{z}{1}$, explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right\}$

(g) i. implicit: $x_1 + x_2 + x_3 - x_4 = 0$, explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \\ 6 \end{bmatrix}\right\}$

ii. implicit: $-2x_1 + 3x_2 - 2x_3 + x_4 = 0$,
explicit: $\operatorname{span}\left\{\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 & 2 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 2 & 2 \end{bmatrix}\right\}$

iii. implicit: $\dfrac{x_1}{-2} = \dfrac{x_2}{3} = \dfrac{x_3}{-2} = \dfrac{x_4}{1}$, explicit: $\operatorname{span}\left\{\begin{bmatrix} -2 \\ 3 \\ -2 \\ 1 \end{bmatrix}\right\}$

Problem Set 4.3

1. **(a)** dependent, $\begin{bmatrix} 2 \\ 6 \\ 0 \end{bmatrix} = -4\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix}$

(c) independent

2. **(a)** independent

(d) dependent, $\begin{bmatrix} 1 \\ 3 \\ 2 \\ 0 \end{bmatrix} = -1\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 2 \\ 5 \\ 3 \\ 1 \end{bmatrix}$

3. **(a)** independent
(b) dependent

4. **(a)** independent
(c) dependent

Problem Set 4.4

1. **(a)** 2 **(c)** 1 **(e)** 0

2. **(a)** yes **(d)** no

3. **(a)** $\left\{\begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}\right\}, 2$ **(c)** $\left\{\begin{bmatrix} 2 \\ 5 \\ 7 \end{bmatrix}\right\}, 1$ **(e)** $\left\{\begin{bmatrix} -3 \\ 7 \\ 5 \end{bmatrix}\right\}, 1$

4. **(a)** i. $\left\{\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right\}, 1$ ii. $\left\{\begin{bmatrix} 1 & 2 \end{bmatrix}\right\}, 1$ iii. $\left\{\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right\}, 1$

(c) i. $\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right\}, 2$

ii. $\left\{\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 3 \end{bmatrix}\right\}, 2$

iii. $\left\{\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}\right\}, 1$

5. **(a)** $\left\{\begin{bmatrix} 4 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ -1 \end{bmatrix}\right\}$

(c) $\left\{\begin{bmatrix} 4 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right\}$

(e) $3t - 4t - 5t + 6t = 0$

6. **(a)** yes **(c)** no

7. **(a)** 5 **(c)** 7 **(e)** 7 **(g)** 5

8. **(a)** $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -5 \\ -8 \\ 8 \end{bmatrix} + \begin{bmatrix} 6 \\ 9 \\ -6 \end{bmatrix}$

(c) no

Problem Set 5.1

1. **(a)** yes, $\begin{bmatrix} 2 & 3 \\ 1 & -4 \end{bmatrix}$

(c) no, $T(0\mathbf{0}) = T(\mathbf{0}) = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ but $0T(\mathbf{0}) = 0\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

(e) no, $T(-1\mathbf{i}) = \|(-1)\mathbf{i}\| = |-1|\|\mathbf{i}\| = (1)(1) = 1$ but $(-1)T(\mathbf{i}) = (-1)\|\mathbf{i}\| = (-1)(1) = -1$

2. **(a)** $\det\begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} = 1 \neq 0$ **(c)** $\begin{bmatrix} 6 & -2 \\ -37 & 16 \end{bmatrix}$

3. **(a)** $\begin{bmatrix} 3 & 2 & 2 \\ 3 & 6 & 0 \\ 3 & 2 & 5 \end{bmatrix}$ **(c)** $\begin{bmatrix} 6 & 3 & -3 \\ 3 & 9 & 3 \\ 9 & 0 & 12 \end{bmatrix}$ **(e)** $\begin{bmatrix} 12 & 4 & 12 \\ 4 & 11 & -3 \\ 5 & 6 & 6 \end{bmatrix}$

4. **(a)** $\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$

(c) $R_\beta \circ F_\alpha = F_{\frac{2\alpha + \beta}{2}}$

(e) $R_\beta \circ F_\alpha = F_{\frac{2\alpha - \beta}{2}}$

Problem Set 5.2

1. **(a)** The kernel of T is the trivial subspace $\{(0, 0)\}$. The preimages of $\{\mathbf{u}\}$, $\{\mathbf{v}\}$, and $\{\mathbf{w}\}$ are the single-element sets $\{(-4, 3)\}$, $\{(7, -4)\}$, and $\{(1, 0)\}$ respectively.

(c) The kernel of T is all of $\mathbb{R}^2$. All three preimages are the empty set.

2. **(a)** The range of T is all of $\mathbb{R}^2$. T is both injective and surjective. T has an inverse.

(c) The range of T is the trivial subspace $\{(0, 0)\}$. T is neither injective nor surjective. T does not have an inverse.

3. **(a)** Basis for *range* $T = \left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}$, Basis for *ker* $T = \left\{\begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix}\right\}$.

(c) Basis for *range* $T = \left\{\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}\right\}$, Basis for *ker* $T = \varnothing$.

4. **(a)** neither injective nor surjective, no inverse.

(c) injective but not surjective, no inverse.

5. **(a)** yes, 1+2=3 **(c)** yes, 0+2=2

6. **(a)** $x = -t$, $y = -t$, $z = t$

7. **(a)** $\mathbb{R}^7$ **(c)** 3 **(e)** no **(g)** no

9. The kernel is the plane $x + 2y + 3z = 0$.

Problem Set 5.3

1. **(a)** $P_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ **(b)** $P_{\mathcal{B}}^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ **(c)** $[\mathbf{u}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

3. Impossible because $\mathcal{B}$ is not a basis for $\mathbb{R}^3$.

4. **(a)** $\begin{bmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -4 \\ 1 \end{bmatrix}$, $p(t) = t^3 - 4t^2 + 5t + 1$.

(c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ 6 \\ 1 \end{bmatrix}$, $r(t) = t(t-1)(t-2) + 6t(t-1) + 7t + 1$.

5. **(a)** $P_{\mathcal{B}} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$, $P_{\mathcal{C}} = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}$, $M = P_{\mathcal{C}}^{-1}P_{\mathcal{B}} = \begin{bmatrix} -3 & -4 \\ 5 & 7 \end{bmatrix}$, $[\mathbf{u}]_{\mathcal{C}} = \begin{bmatrix} -25 \\ 43 \end{bmatrix}$

(b) $[P_{\mathcal{C}}|P_{\mathcal{B}}] = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 0 & -3 & -4 \\ 0 & 1 & 5 & 7 \end{bmatrix}$, $M = \begin{bmatrix} -3 & -4 \\ 5 & 7 \end{bmatrix}$

Problem Set 5.4

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

3. **(a)** $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

5. $Q = \begin{bmatrix} 1 & e \\ -e & 1 \end{bmatrix}$, $A = \begin{bmatrix} e & 1 \\ 1 & e^{-1} \end{bmatrix}$, $M = \begin{bmatrix} 0 & 0 \\ 0 & e + e^{-1} \end{bmatrix}$

Problem Set 6.1

1. **(a)** $\lambda = 2, \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$

(b) no

(c) $\lambda = -1, \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$

3. **(a)** no

(b) $\lambda = 1, \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$

(c) $\lambda = -1, \left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

(d) $\lambda = 0, \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$

5. **(a)** $\lambda = -2, \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$

(b) no

(c) $\lambda = 1, \left\{ \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}$

7. **(a)** for $\lambda = 4 \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$, for $\lambda = 3 \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$

(c) for $\lambda = -1 \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$, for $\lambda = 2 \left\{ \begin{bmatrix} 4 \\ 3 \\ 9 \end{bmatrix} \right\}$

8. **(a)** false (true if nonzero vector) **(c)** true

(e) false (converse is true) **(g)** true

9. eigenvector $\mathbf{v} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

11. **(a)** $\lambda = 1, -1$

(b) E_1 is the line $y = 2x$, E_{-1} is the line $y = -\frac{1}{2}x$

13. **(a)** $\lambda = 0, 1$

(b) E_1 is the plane $x + 2y + 3z = 0$, E_0 is the line $\mathbf{x}(t) = t \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

15. $A\mathbf{v} = \lambda\mathbf{v} \Rightarrow A^{-1}A\mathbf{v} = A^{-1}(\lambda\mathbf{v}) \Rightarrow \mathbf{v} = \lambda A^{-1}\mathbf{v} \Rightarrow A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}$.

Problem Set 6.2

1. **(a)** $p(\lambda) = \lambda^2 - 2\lambda - 3$, $\lambda = 3, -1$

(c) $p(\lambda) = \lambda^2 - 4\lambda + 4$, $\lambda = 2$

(e) $p(\lambda) = \lambda^2 + 1$, no real eigenvalues $\lambda = \pm i$

(g) $p(\lambda) = -\lambda^3 - \lambda^2 + 2\lambda = -\lambda(\lambda+2)(\lambda-1)$ (factor out the common monomial factor $-\lambda$), $\lambda = 0, 1, -2$

(i) $p(\lambda) = -(\lambda-1)^2(\lambda+3)$, $\lambda = 1, -3$

2. For each of the following, let $\mathcal{B}_\lambda$ be a basis for the eigenspace associated with λ.

(a) $\mathcal{B}_3 = \left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$, $\mathcal{B}_{-1} = \left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$

(c) $\mathcal{B}_2 = \left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$

(e) no eigenvectors in $\mathbb{R}^2$, but in $\mathbb{C}^2$ $\mathcal{B}_i = \left\{\begin{bmatrix} i \\ 1 \end{bmatrix}\right\}$, $\mathcal{B}_{-i} = \left\{\begin{bmatrix} -i \\ 1 \end{bmatrix}\right\}$

(g) $\mathcal{B}_0 = \left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right\}$, $\mathcal{B}_1 = \left\{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right\}$, $\mathcal{B}_{-2} = \left\{\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}\right\}$

(i) $\mathcal{B}_1 = \left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$, $\mathcal{B}_{-3} = \left\{\begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}\right\}$

3. Let $p(\lambda)$ and $q(\lambda)$ be the characteristic polynomials of A and A^T respectively.
$q(\lambda) = det(A^T - \lambda I) = det(A^T - \lambda I^T) = det(A - \lambda I)^T = det(A - \lambda I) = p(\lambda)$.

5. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and let $p(\lambda)$ be the characteristic polynomial of A.
$p(\lambda) = \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a-d)\lambda + (ad-bc) = \lambda^2 - tr(A)\lambda + det(A)$.

8. **(a)** False, -3 is an eigenvalue of A.

(c) False. Counterexample, $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = B$. The characteristic polynomial of A is $p(\lambda) = \lambda^2 - 2\lambda + 1$, but the characteristic polynomial of B is $q(\lambda) = \lambda^2 - 3\lambda + 1$.

(e) True. Since the characteristic polynomials are the same (exercise 3), the eigenvalues are the same.

9. Let A be an $n \times n$ matrix with column sums all equal to s, then A^T has all row sums equal to s. By exercise 9 of section 6.1, A^T has an eigenvalue of s, so $\lambda - s$ is a factor of the characteristic polynomial of A^T. By exercise 3 of this section, the characteristic polynomials of A and A^T are the same, so $\lambda - s$ is a factor of the characteristic polynomial of A. Therefore, s is an eigenvalue of A.

Problem Set 6.3

1. **(a)** $\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$

(c) $\begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$

(e) not diagonalizable

(g) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

3. **(a)** $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

(c) $P = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}$

(e) not diagonalizable

(g) $P = \begin{bmatrix} -1 & 2 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

5. **(a)** $D_1D_2 = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mu_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n\mu_n \end{bmatrix}$

(c) $D_1^{-1} = \begin{bmatrix} \lambda_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{-1} \end{bmatrix}$

6. **(a)** $D^k = DD\ldots DD = (P^{-1}AP)(P^{-1}AP)\ldots(P^{-1}AP)(P^{-1}AP) = P^{-1}A(PP^{-1})A\ldots A(PP^{-1})AP = P^{-1}A^kP$.

7. $\begin{bmatrix} 8 & -18 \\ 3 & -7 \end{bmatrix}^k = \begin{bmatrix} 3\cdot 2^k - 2(-1)^k & -6\cdot 2^k + 6(-1)^k \\ 2^k - (-1)^k & -2\cdot 2^k + 3(-1)^k \end{bmatrix}$

8. $A = \begin{bmatrix} -2 & 20 \\ -1 & 7 \end{bmatrix}$

10. **(a)** T **(c)** T **(e)** T **(g)** F **(i)** F

13. A diagonalizable $\Rightarrow$ there exist an invertible P and a diagonal D such that $P^{-1}AP = D$. B similar to $A \Rightarrow$ there exist an invertible Q such that $A = Q^{-1}BQ$. So, $D = P^{-1}AP = P^{-1}(Q^{-1}BQ)P = (QP)^{-1}B(QP)$. Therefore, B is diagonalizable.

14. **(a)** $rank(A) = 1$, $nullity(A) = n - 1$.

(c) Yes, $dim(E_0) = nullity(A) = n - 1$.

(e) $\lambda = \mathbf{v}\cdot\mathbf{u}$.

(g) $p(\lambda) = (-1)^n\lambda^{n-1}(\lambda - \mathbf{v}\cdot\mathbf{u})$.

15. **(a)** $rank(A) = 1$, $nullity(A) = n - 1$.

(c) Yes, $dim(E_0) = nullity(A) = n - 1$.

(e) $\lambda = 0$.

(g) $p(\lambda) = (-1)^n \lambda^n$.

Problem Set 6.4

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

(ii) $\lambda = 0$, $\mathcal{B}_0 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$

(iii) $\lambda = 0$, $\mathcal{C}_0 = \{1\}$

(iv) no

(c) (i) $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

(ii) $\lambda = 0, 1, 2, 3$, $\mathcal{B}_0 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$, $\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$, $\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\}$, $\mathcal{B}_3 = \left\{ \begin{bmatrix} 1 \\ 3 \\ 3 \\ 1 \end{bmatrix} \right\}$

(iii) $\lambda = 0, 1, 2, 3$, $\mathcal{C}_0 = \{1\}$, $\mathcal{C}_1 = \{t + 1\}$, $\mathcal{C}_2 = \{t^2 + 2t + 1\}$, $\mathcal{C}_3 = \{t^3 + 3t^2 + 3t + 1\}$

(iv) yes

2. **(b)** (i) $A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & -5 \\ 0 & 1 & 4 \end{bmatrix}$

(ii) $\lambda = 1, 2$, $\mathcal{B}_1 = \left\{ \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \right\}$, $\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$

(iii) $\lambda = 1, 2$, $\mathcal{C}_1 = \left\{ \begin{bmatrix} -3 & -9 & 9 \\ 0 & 0 & 0 \\ -1 & -3 & 3 \end{bmatrix} \right\}$, $\mathcal{C}_2 = \left\{ \begin{bmatrix} 0 & 0 & 0 \\ -1 & -2 & 3 \\ -1 & -2 & 3 \end{bmatrix} \right\}$

(iv) no

Problem Set 7.1

1. **(a)** (i) -1, (ii) 0, (iii) $\sqrt{5}$, (iv) $\sqrt{17}$

2. **(b)** (i) -6, (ii) 0, (iii) $\sqrt{43}$, (iv) $\sqrt{68}$

3. **(a)** 0 **(c)** $\sqrt{2\pi}$

4. $y = -3x + 5$ and $y = -\frac{31}{8}x + \frac{47}{8}$

5. $A = \begin{bmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{w_n} \end{bmatrix}$

Problem Set 7.2

1. **(a)** yes **(c)** no

2. **(a)** $\dfrac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(c) $\dfrac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

3. **(a)** $\pi/3$ **(c)** $\pi/2$

4. **(a)** yes **(c)** no

5. $\pm\dfrac{1}{\sqrt{39}}\begin{bmatrix} 5 \\ -2 \\ -3 \\ 1 \end{bmatrix}$

6. **(a)** $\left\{\begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix}\right\}$

7. **(a)** $dim\ W = 3$, $dim\ W^{\perp} = 1$

9. $3x + 4y + 5z = 0$

10. **(a)** $\left\{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}$

11. $\sqrt{13}$

Problem Set 7.3

1. **(a)** (i) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -4 \\ 3 \end{bmatrix} = -1 - 8 + 9 = 0$, (ii) $\left\{\dfrac{1}{\sqrt{14}}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \dfrac{1}{\sqrt{26}}\begin{bmatrix} -1 \\ -4 \\ 3 \end{bmatrix}\right\}$, (iii) no

(c) (i) $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\cdot\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}=0$, $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}=0$, $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}=0$, $\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}=0$, $\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}=0$, $\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}=0$, (ii) $\left\{\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}\right\}$,
(iii)yes

2. **(a)** $\begin{bmatrix}7/9\\2/9\\-1/9\end{bmatrix}$

(c) $\begin{bmatrix}8/5\\-1/5\end{bmatrix}$

3. **(a)** $\begin{bmatrix}7/4\\7/4\\7/4\\7/4\end{bmatrix}, \begin{bmatrix}-3/4\\1/4\\5/4\\-3/4\end{bmatrix}$ **(c)** $\begin{bmatrix}3/2\\3/2\\2\\2\end{bmatrix}, \begin{bmatrix}-1/2\\1/2\\1\\-1\end{bmatrix}$

4. **(a)** $P\mathbf{x} = \left(\left(\frac{1}{\mathbf{w}^T\mathbf{w}}\right)\mathbf{w}\mathbf{w}^T\right)\mathbf{x} = \left(\left(\frac{1}{\mathbf{w}^T\mathbf{w}}\right)\mathbf{w}\right)(\mathbf{w}^T\mathbf{x}) = \left(\frac{\mathbf{w}\cdot\mathbf{x}}{\mathbf{w}\cdot\mathbf{w}}\right)\mathbf{w} = proj_{\mathbf{w}}\mathbf{x}$.

(c) $P = \frac{1}{14}\begin{bmatrix}1&2&3\\2&4&6\\3&6&9\end{bmatrix}$

5. **(a)** Since $W^\perp$ is a subspace of $\mathbb{R}^n$, there exists $\hat{\mathbf{y}}$ and $\hat{\hat{\mathbf{y}}}$ such that $proj_{W^\perp}\mathbf{x} = \hat{\mathbf{y}}$ and the component of $\mathbf{x}$ orthogonal to $W^\perp$ is $\hat{\hat{\mathbf{y}}}$, so $\mathbf{x} = \hat{\mathbf{y}} + \hat{\hat{\mathbf{y}}}$ with $\hat{\mathbf{y}} \in W^\perp$ and $\hat{\hat{\mathbf{y}}} \in (W^\perp)^\perp = W$. Since $\mathbb{R}^n = W \oplus W^\perp$, $\hat{\mathbf{y}} = \hat{\hat{\mathbf{x}}}$ and $\hat{\hat{\mathbf{y}}} = \hat{\mathbf{x}}$. Thus, $ot_{W^\perp}\mathbf{x} = \hat{\mathbf{y}} - \hat{\hat{\mathbf{y}}} = \hat{\hat{\mathbf{x}}} - \hat{\mathbf{x}} = -ot_W\mathbf{x}$.

(c) $(2P - I_n)\mathbf{x} = 2P\mathbf{x} - I_n\mathbf{x} = 2proj_W\mathbf{x} - \mathbf{x} = ot_W\mathbf{x}$ by part (b). Thus, $R = 2P - I_n$.

(e) $R = \frac{1}{9}\begin{bmatrix}1&8&-4\\8&1&4\\-4&4&7\end{bmatrix}$

7. **(a)** F **(b)** T **(c)** F **(d)** F
(e) T **(f)** T **(g)** T **(h)** T

Problem Set 7.4

1. **(a)** $\left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\-1\end{bmatrix}\right\}$ **(c)** $\left\{\begin{bmatrix}4\\1\end{bmatrix}, \begin{bmatrix}-1\\4\end{bmatrix}\right\}$

3. $\left\{\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\-1\\2\\0\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\\3\end{bmatrix}\right\}$

4. **(a)** $\{x - 1, 4x^2 + 13x - 5\}$

5. **(a)** T **(c)** F **(e)** T

Problem Set 7.5

1. **(a)** $\begin{bmatrix} 21 & 7 \\ 7 & 10 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$

(c) $\begin{bmatrix} 7 & 4 \\ 4 & 7 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$

2. **(a)** $\begin{bmatrix} 44/161 \\ -9/23 \end{bmatrix}$

(c) $\begin{bmatrix} 41/33 \\ -14/33 \end{bmatrix}$

3. $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7/17 \\ 3/17 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$

4. **(a)** $y = \frac{3}{2}x + \frac{2}{3}$

5. $z = -\frac{7}{3}x - \frac{3}{2}y + \frac{7}{2}$

7. $P = \frac{1}{6}\begin{bmatrix} 5 & 1 & -2 \\ 1 & 5 & 2 \\ -2 & 2 & 2 \end{bmatrix}$

9. **(a)** F **(c)** T **(e)** F **(g)** T

Problem Set 7.6

1. **(a)** $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$;
$\frac{1}{5}\begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix}\frac{1}{5}\begin{bmatrix} 4 & 3 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$;
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(c) $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

(e) $\frac{1}{5}\begin{bmatrix} 4 & 3 \\ -3 & 4 \end{bmatrix}$

(g) $\frac{1}{5\sqrt{2}}\begin{bmatrix} 7 & 1 \\ -1 & 7 \end{bmatrix}$

2. **(a)** $\frac{1}{15}\begin{bmatrix} -5 & 2 & 14 \\ 10 & 11 & 2 \\ 10 & -10 & 5 \end{bmatrix} \frac{1}{15}\begin{bmatrix} -5 & 10 & 10 \\ 2 & 11 & -10 \\ 14 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$;

$\frac{1}{25}\begin{bmatrix} 15 & -12 & -16 \\ 20 & 9 & 12 \\ 0 & 20 & -15 \end{bmatrix} \frac{1}{25}\begin{bmatrix} 15 & 20 & 0 \\ -12 & 9 & 20 \\ -16 & 12 & -15 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$;

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(c) $\frac{1}{15}\begin{bmatrix} -5 & 10 & 10 \\ 2 & 11 & -10 \\ 14 & 2 & 5 \end{bmatrix}$

(e) $\frac{1}{25}\begin{bmatrix} 15 & 20 & 0 \\ -12 & 9 & 20 \\ -16 & 12 & -15 \end{bmatrix}$

(g) $\frac{1}{15}\begin{bmatrix} 5 & 14 & 2 \\ 10 & -5 & 10 \\ 10 & -2 & -11 \end{bmatrix}$

3. **(a)** left-handed **(c)** right-handed

4. **(a)** $Q^T = (I - \frac{2}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}\mathbf{n}^T)^T = I^T - \frac{2}{\mathbf{n}\cdot\mathbf{n}}(\mathbf{n}\mathbf{n}^T)^T = I - \frac{2}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}\mathbf{n}^T = Q$

(c) $Q\mathbf{n} = (I - \frac{2}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}\mathbf{n}^T)\mathbf{n} = I\mathbf{n} - \frac{2}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}(\mathbf{n}^T\mathbf{n}) = \mathbf{n} - 2\mathbf{n} = -\mathbf{n}$

(e) $\lambda = \pm 1$

(g) 2,2

(i) $\lambda = -1$ with algebraic & geometric multiplicity 1, $\lambda = 1$ with algebraic & geometric multiplicity 2.

(k) $\frac{1}{7}\begin{bmatrix} 6 & -2 & -3 \\ -2 & 3 & -6 \\ -3 & -6 & -2 \end{bmatrix}$

5. **(a)** $R_\theta R_\theta^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(c) $AA^T = (UR_\theta U^T)(UR_\theta U^T)^T = UR_\theta(U^TU)R_\theta^T U^T = U(R_\theta R_\theta^T)U^T = UU^T = I$

(e) $R_{\frac{2\pi}{3}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/2 & -\sqrt{3}/2 \\ 0 & \sqrt{3}/2 & -1/2 \end{bmatrix}$

(g) The positive x axis rotates to the positive y axis. The positive y axis rotates to the positive z axis. The positive z axis rotates to the positive x axis.

Problem Set 7.7

1. **(a)** $D = \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix}$, $U = \frac{1}{\sqrt{5}}\begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}$

(c) $D = \begin{bmatrix} 5 & 0 \\ 0 & -5 \end{bmatrix}$, $U = \dfrac{1}{\sqrt{10}} \begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}$

(e) $D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{bmatrix}$, $U = \dfrac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$

(g) $D = \begin{bmatrix} 18 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$, $U = \dfrac{1}{3} \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$

(i) $D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $U = \begin{bmatrix} 1/2 & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} \\ 1/2 & -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} \\ 1/2 & 0 & -2/\sqrt{6} & 1/\sqrt{12} \\ 1/2 & 0 & 0 & -3/\sqrt{12} \end{bmatrix}$

3. **(a)** $P\mathbf{x} = (\mathbf{u}\mathbf{u}^T)\mathbf{x} = \mathbf{u}(\mathbf{u}^T\mathbf{x}) = \mathbf{u}(\mathbf{u}\cdot\mathbf{x}) = \frac{\mathbf{u}\cdot\mathbf{x}}{1}\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{x}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = proj_{\mathbf{u}}\mathbf{x}$

(c) $P^2 = (\mathbf{u}\mathbf{u}^T)(\mathbf{u}\mathbf{u}^T) = \mathbf{u}(\mathbf{u}^T\mathbf{u})\mathbf{u}^T = \mathbf{u}(\mathbf{u}\cdot\mathbf{u})\mathbf{u}^T = \mathbf{u}(1)\mathbf{u}^T = \mathbf{u}\mathbf{u}^T = P$

(e) $\lambda = 0$, the eigenspace is the solution set of $\mathbf{u}\cdot\mathbf{x} = 0$, its dimension is $n-1$.

5. **(a)** T **(c)** T **(e)** F **(g)** T **(i)** T

Problem Set 8.1

1. **(a)** $A = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$ **(c)** $A = \begin{bmatrix} 2 & 6 \\ 6 & -7 \end{bmatrix}$

2. **(a)** elliptic paraboloid **(c)** hyperbolic paraboloid

3. $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$

4. **(c)** $\frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$

5. **(a)** $z = 6(x')^2 + 4(y')^2; z = 4(x')^2 + 6(y')^2$

(c) $z = 5(x')^2 - 10(y')^2; z = 5(y')^2 - 10(x')^2$

6. **(a)** $w = 18(x')^2 + 9(y')^2 - 9(z')^2$

7. **(a)** 6

8. **(a)** $4! = 24$ **(c)** 6

9. $\dfrac{n!}{m_1!\cdots m_k!}$

Problem Set 8.2

1. max: 36 at $\pm\left(\frac{3}{\sqrt{2}}, \frac{3}{\sqrt{2}}\right)$; min: 18 at $\pm\left(\frac{3}{\sqrt{2}}, -\frac{3}{\sqrt{2}}\right)$

3. max: 25 at $\pm\left(\frac{3}{5}, \frac{4}{5}\right)$; min: 0 at $\pm\left(-\frac{4}{5}, \frac{3}{5}\right)$

5. max: 24 at $\pm\frac{2}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$; min: -24 at $\pm\sqrt{2}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

Problem Set 8.3

1. **(a)** Hyperbola or degenerate **(c)** Parabola or degenerate

2. **(a)** $\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 1$

(c) $\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 18 & 6 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 13 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = -5$

3. **(a)** $U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \theta = 45°$

(c) $U = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}, \theta = \tan^{-1}\left(\frac{1}{3}\right) \approx 18.4°$

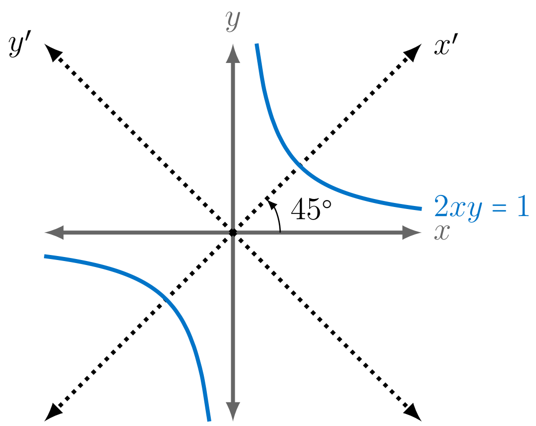
4. **(a)** $(x')^2 - (y')^2 = 1$

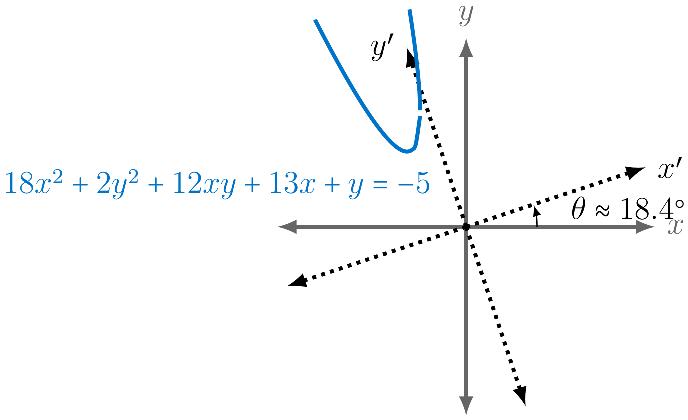
(c) $2\sqrt{10}\left(x' + \frac{1}{\sqrt{10}}\right)^2 = y' - \frac{3}{\sqrt{10}}$

5. **(a)** $(0,0)$ **(c)** $\left(-\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$

6. **(a)** $(0,0)$ **(c)** $\left(-\frac{3}{5}, \frac{4}{5}\right)$

7. **(a)**





(c)

Problem Set 8.4

1. **(a)** $\dfrac{(x')^2}{\sqrt{3}^2} + \dfrac{(y')^2}{3^2} + \dfrac{(z')^2}{3^2} = 1$

(b) Ellipsoid

3. **(a)** $3(x')^2 = 1$

(b) Degenerate quadric surface (two parallel planes: $x + y + z = 1$, $x + y + z = -1$)

Problem Set 8.5

1. **(a)** indefinite **(c)** positive semidefinite **(e)** negative semidefinite

2. **(a)** positive definite **(c)** negative definite

3. **(a)** A^2 is positive semidefinite.

4. **(a)** A^2 is symmetric.

5. They are the same.

6. **(a)** Since each column has all zeros except one 1, each column has a norm of 1. Since each row has only one 1, the ones in two different columns don't match up, so the dot product of any two different columns is 0.

(b) $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

(e) If A is a symmetric matrix with a negative entry in position (i, i), let P be the permutation matrix that swaps row and column 1 with row and column i.

Then P^TAP has a negative entry in position $(1,1)$ and is, therefore, not positive definite by Theorem 8.10. But by Corollary 6.7, P^TAP and A have exactly the same eigenvalues, so A is not positive definite either. Therefore, all diagonal entries of a positive definite matrix must be positive.

7. (a) $B = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix}$

8. (a) Ellipsoid

(c) Two parallel planes

(e) Hyperboloid of one sheet

Problem Set 8.6

1. $\sigma_1 = 2\sqrt{2}$, $\sigma_2 = \sqrt{2}$, $\Sigma = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$, $V = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

3. $\sigma_1 = \sqrt{6}$, $\sigma_2 = 1$, $\sigma_3 = 0$, $\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $V = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & -\frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}$, $U = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$

5. $A = AI_nI_n$, $U = A$, $\Sigma = I_n$, and $V = I_n$. The singular values are all 1.

6. The singular values equal the absolute value of the eigenvalues.

8. (a) From the hint, $A = \left(UV^T\right)\left(V\Sigma V^T\right)$. Let $Q = UV^T$ and $S = V\Sigma V^T$. Q is the product of two orthogonal matrices, hence orthogonal. Since Σ is diagonal with nonnegative entries and V is orthogonal, S is symmetric positive semidefinite.

9. $A = \left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\right)\left(\frac{1}{2}\begin{bmatrix} 3\sqrt{2} & \sqrt{2} \\ \sqrt{2} & 3\sqrt{2} \end{bmatrix}\right)$ and $A^T = \left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\right)\left(\begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}\right)$

Problem Set 8.7

1. (a) 5

(c) 5

(e) $\sqrt{\dfrac{3+\sqrt{5}}{2}}$

(g) $\sqrt{3+\sqrt{6}}$

3. By Theorem 7.28, for $\mathbf{x} \in \mathbb{R}^n$, $\|UA\mathbf{x}\| = \|U(A\mathbf{x})\| = \|A\mathbf{x}\|$, so $\|UA\| = \max\limits_{\|\mathbf{x}\|=1} \|AU\mathbf{x}\| = \max\limits_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \|A\|$ by Lemma 8.12(a).

4. By exercise 6 of section 8.6, since A is symmetric, the singular values of A are $|\lambda_1|, \cdots, |\lambda_n|$. By Theorem 8.13, $\|A\|$ equals the largest singular value of A, so $\|A\| = \max\{|\lambda_1|, \cdots, |\lambda_n|\}$.

6. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Then $A + B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $A - B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. So $\|A\| = 1$, $\|B\| = 1$, $\|A+B\| = 2$, and $\|A-B\| = 1$. But $\|A+B\|^2 + \|A-B\|^2 = 4+1 = 5$ and $2\|A\|^2 + 2\|B\|^2 = 2+2 = 4$ and $5 \neq 4$. Other examples may be used.

8. Let $A = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{x} \in \mathbb{R}^1$ be such that $\|\mathbf{x}\| = 1$, thus $\mathbf{x} = \pm 1$. Then,

$$A\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix} \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix} \text{ or } \begin{bmatrix} -a \\ -b \end{bmatrix},$$

so

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \sqrt{a^2 + b^2} = \text{Euclidean norm of } \begin{bmatrix} a \\ b \end{bmatrix}.$$

Thus, there is an inner product (namely, the dot product) on the vector space of 2×1 matrices that generates the spectral norm.

9. $$B - A = U\Gamma V^T - U\Sigma V^T = U(\Gamma - \Sigma)V^T = U \left[\begin{array}{ccc|ccc} 0 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & 0 & & & \\ \hline & & & \epsilon & & \\ & 0 & & & \ddots & \\ & & & & & \epsilon \end{array}\right] V^T.$$

So the singular values of $B - A$ are 0 and ϵ and thus $\|B - A\| = \epsilon$.

Problem Set 8.8

1. **(a)** $\frac{1}{50}\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$

(c) $\frac{1}{3}\begin{bmatrix} 1 & 1 \\ -1 & 2 \\ 2 & -1 \end{bmatrix}$

(e) $\frac{1}{6}\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & -2 & 2 & 1 \\ 2 & -1 & 1 & 2 \end{bmatrix}$

2. **(a)** Note that

$$\begin{aligned} T_{A^+} \circ T_A(\mathbf{v}_1) &= \frac{1}{6}\begin{bmatrix} 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{6}\begin{bmatrix} 4 & -2 & 2 \\ -2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \mathbf{v}_1 \end{aligned}$$

and

$$\begin{aligned} T_{A^+} \circ T_A(\mathbf{v}_2) &= \frac{1}{6}\begin{bmatrix} 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{6}\begin{bmatrix} 4 & -2 & 2 \\ -2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \mathbf{v}_2. \end{aligned}$$

Since $T_{A^+} \circ T_A$ acts as the identity on the basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for *row* A, it acts as the identity on all of *row* A.

(c) $T_{A^+} \circ T_A(\mathbf{v}_3) = \frac{1}{6}\begin{bmatrix} 4 & -2 & 2 \\ -2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}$.

Since $T_A \circ T_{A^+}$ acts as the zero transformation on the basis $\{\mathbf{v}_3\}$ for *null* A, it acts as the zero transformation on all of *null* A.

3. **(a)** Since T_A restricted to *row* A is onto *col* A, there is a unique $\mathbf{v} \in$ *row* A such that $T_A(\mathbf{v}) = A\mathbf{v} = \mathbf{b}$. Thus $A\mathbf{x} = \mathbf{b}$ has exactly one solution $\mathbf{v}$ that comes from *row* A.

(b) Suppose $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$. Then both $A\mathbf{v} = \mathbf{b}$ and $A\mathbf{w} = \mathbf{b}$ (from part (a)), so

$$\begin{aligned} A\mathbf{w} = A\mathbf{v} &\implies A\mathbf{w} - A\mathbf{v} = \mathbf{0} \\ &\implies A(\mathbf{w} - \mathbf{v}) = \mathbf{0} \\ &\implies \mathbf{w} - \mathbf{v} \in \textit{null } A. \end{aligned}$$

Let $\mathbf{z} = \mathbf{w} - \mathbf{v}$. Then $\mathbf{w} = \mathbf{v} + \mathbf{z}$.

Suppose $\mathbf{w} = \mathbf{v} + \mathbf{z}$ where $\mathbf{z} \in \mathit{null}\ A$. Then

$$\begin{aligned} A\mathbf{w} &= A(\mathbf{v}+\mathbf{z}) \\ &= A\mathbf{v} + A\mathbf{z} \\ &= \mathbf{b} + \mathbf{0} \text{ since } \mathbf{v} \text{ is a solution to } A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{z} \in \mathit{null}\ A \\ &= \mathbf{b} \end{aligned}$$

Thus $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$.

4. The least-squares solutions to $A\mathbf{x} = \mathbf{b}$ are the solutions to $A\mathbf{x} = \widehat{\mathbf{b}}$, where $\widehat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto *col* A by Definition 7.16. Since $\widehat{\mathbf{b}} \in \mathit{col}\ A$, exercise 3 implies that the solution to $A\mathbf{x} = \widehat{\mathbf{b}}$ with the smallest norm is the unique solution $\mathbf{v}$ to $A\mathbf{x} = \widehat{\mathbf{b}}$ that lies in *row* A. Thus $T_A(\mathbf{v}) = \mathbf{b}$. But T_A and T_{A^+} undo each other between *row* A and *col* A. So

$$T_A(\mathbf{v}) = \mathbf{b} \implies T_{A^+}(\mathbf{b}) = \mathbf{v} \implies A^+\mathbf{b} = \mathbf{v}.$$

Therefore $A^+\mathbf{b}$ is a solution to $A\mathbf{x} = \widehat{\mathbf{b}}$ with the smallest norm. Thus $A^+\mathbf{b}$ is the least-squares solution to $A\mathbf{x} = \mathbf{b}$ with the smallest norm.

5. Let $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{b} = A\mathbf{w}$. Then $\mathbf{b} \in \mathit{col}\ A$ and $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$. By exercise 3(b) there exists unique vectors $\mathbf{v} \in \mathit{row}\ A$ and $\mathbf{z} \in \mathit{null}\ A$ such that $A\mathbf{v} = \mathbf{b}$ and $\mathbf{w} = \mathbf{v} + \mathbf{z}$. Since $\mathit{null}\ A = (\mathit{row}\ A)^\perp$, $\mathbf{v}$ is the orthogonal projection of $\mathbf{w}$ onto *row* A. We show $A^+A\mathbf{w} = \mathbf{v}$. But

$$A^+A\mathbf{w} = A^+(A\mathbf{w}) = A^+\mathbf{b} = T_{A^+}(\mathbf{b}).$$

Also, T_A and T_{A^+} undo each other on *row* A and *col* A. Since $A\mathbf{v} = \mathbf{b}$ we know $A^+\mathbf{b} = \mathbf{v}$. Therefore $A^+A\mathbf{w} = \mathbf{v}$.

7. **(a)** $\frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

Index